True and Useful: On the Structure of a Two Level Normative Theory

FRED FELDMAN

University of Massachusetts at Amherst

Act-utilitarianism and other theories in normative ethics confront the implementability problem: normal human agents, with normal human epistemic abilities, lack the information needed to use those theories directly for the selection of actions. Two Level Theories have been offered in reply. The theoretical level component states alleged necessary and sufficient conditions for moral rightness. That component is supposed to be true, but is not intended for practical use. It gives an account of objective obligation. The practical level component is offered as an implementable system for the choice of actions by agents lacking some relevant information. It gives an account of subjective obligation. Several different ways of developing Two Levelism are explained and criticized. Five conditions that must be satisfied if the practical level principle is to be a good match for a given theoretical level principle are stated. A better form of Two Levelism is presented.

I. A PUZZLE IN ETHICS

Philosophers engaged in normative ethics seek a principle stating interesting necessary and sufficient conditions for the moral rightness of actions. Utilitarians, Kantians, Rossians, virtue ethicists and others defend different principles. But no matter which of these views a philosopher wants to defend, there is a certain difficulty that must be confronted. The problem is that these principles are 'unimplementable'. A morally conscientious person who believes in one of these principles would naturally want to select actions that conform to the requirements of that principle. Yet no one has the kind of detailed information that would be required to implement any of these principles. Thus, the principles are unhelpful, or seemingly pointless: we cannot use any of them to determine in any helpful way what we are supposed to do.

This problem has been discussed extensively in connection with utilitarianism. But it is important to see that other theories in normative ethics confront epistemic problems just as intractable as those faced by utilitarianism. On Ross's theory we are required to perform an act that maximizes the balance of prima facie rightness over prima facie wrongness. Yet, as Ross himself made clear, we often cannot determine which of our alternatives would have this feature.[1] The

[1] 'There is therefore much truth in the description of the right act as a fortunate act. If we cannot be certain that it is right, it is our good fortune if the act we do is the right act' (W. D. Ross, *The Right and the Good* (Oxford, 1930), p. 31).

© Cambridge University Press 2012
Utilitas Vol. 24, No. 2, June 2012
doi:10.1017/S0953820812000155

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

same is true of various forms of virtue ethics and all the other popular theories in normative ethics. The implementability problem confronts anyone who advocates any of the popular theories in normative ethics.

II. TWO LEVEL THEORIES IN GENERAL

One possible solution to this problem involves a move to a 'Two Levels Theory'. We can say that a complete theory in normative ethics consists of two distinct components. The first component is the familiar normative principle – the part that states alleged necessary and sufficient conditions for the moral rightness of actions. That principle – whatever it may be – will probably turn out to be unimplementable. The second part of the theory will be something intended to be immediately helpful as the agent tries to choose actions. It will have to be implementable. If each of the components does its job successfully, and they are suitably mated, the resulting package is both (a) plausible as an account of what makes right acts right, and (b) useful as a guide to the selection of actions in real life.

For convenience in discussion, I will say that the actual principle of moral rightness is the 'theoretical level principle' and I will say that the other item – the decision procedure or whatever it turns out to be – is the 'practical level principle'.[2]

In any plausible Two Level view, the two components must be properly connected. For any selected theoretical level principle, there must be a certain practical level principle suitable for use by those who have accepted that theoretical level principle. The combination of that theoretical level principle and that practical level principle will make a coherent Two Level Theory in normative ethics.

III. CRITERIA OF ADEQUACY FOR PRACTICAL LEVEL PRINCIPLES

What features would make a practical level principle the appropriate match for a given theoretical level principle? I think the most intuitive way to proceed is to introduce a sample theoretical level principle, and then to describe the conditions that must be satisfied by a practical level principle if it is to be the appropriate partner for the selected theoretical level principle. Because it's so familiar, I will use a generic form of act-utilitarianism as my sample theoretical level principle. The claims I make about the features of the associated practical level principle in this case will carry over pretty directly in other cases in which we start

[2] Others have used other terminology here. Some have described the first thing as the principle of 'objective obligation' and the second thing as the principle of 'subjective obligation'. Hare used the terms 'critical level principle' and 'intuitive level principle' in a closely related way. I have no objection to any terminology here. After all, they are just names.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

with a different theoretical level principle. In the final section of the article, extensions to Rossianism and a form of virtue ethics are discussed.

According to act-utilitarianism (AU), an act is morally right if and only if it maximizes utility. Classically, the utility of an act was understood to be the total amount of pleasure the act would cause, minus the total amount of pain it would cause. More plausible modern forms of act-utilitarianism make use of more sophisticated assumptions about value.[3] But no matter what the details, it will turn out that no ordinary human being has the information he would need in order actually to use AU when in the real world trying to figure out what to do.[4]

We seek one main feature in a theoretical level principle: it should state what we take to be the actual necessary and sufficient conditions for the absolute, objective moral rightness of actions. We do not insist that the principle be implementable. Implementability is to be sought in the associated practical level principle. Can we say more about the features we seek in a practical level principle? I will describe five conditions that must be satisfied by a practical level principle if it is to be a suitable partner for AU as the theoretical level principle in a Two Level Theory.[5]

The first condition is:

(a) Implementability: Suppose AU is my theoretical level principle. In some cases I don't know, and realize that I cannot figure out in any helpful way, which of my alternatives will lead to the best outcome.[6] So I need some practical level principle that will offer guidance in this condition of irremediable ignorance. Obviously, in this situation, it would be pointless to turn to another principle if that other principle were just as hard to implement as AU. Thus, the associated practical level principle must be implementable.

[3] I have defended a variant in which right acts are said to maximize 'desert adjusted' utility. For details, see Fred Feldman, 'Adjusting Utility for Justice', *Philosophy and Phenomenological Research* 55 (1995), pp. 567–85.

[4] The classic discussion of this point can be found in Moore's *Principia Ethica* (Cambridge, 1903, rev. edn., 1993), pp. 211–14 (sect. 99). See also Robert Frazier, 'Act Utilitarianism and Decision Procedures', *Utilitas* 6 (1994), pp. 43–53. R. E. Bales, 'Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?', *American Philosophical Quarterly* 8 (1971), pp. 257–65, contains a nice discussion of how the implementability problem arises for AU. Section 21 of Derek Parfit's *On What Matters* (Oxford, 2011) is devoted to a discussion of some closely related points.

[5] Holly Smith states six 'criteria of adequacy' for an account of subjective rightness in her 'Subjective Rightness', *Social Philosophy and Policy* 27 (2010), pp. 64–110, at 72–3. The interested reader is encouraged to study and compare her list with the one given here.

[6] Of course I *can* easily figure out, in an unhelpful way, what I should do. I can always say that I should maximize utility, or 'do whatever would be best'. But these recommendations are practically unhelpful; I cannot directly make use of the recommendation. If given such a recommendation, I would of course agree, but then I would have to ask for further help: 'But which of my alternatives in fact is the one that would maximize utility?'

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

To clarify this condition, let us consider some ways in which a practical level principle could fail to satisfy it. Consider this:

PLP1: When you cannot identify the act that is required according to the theoretical level principle that you accept, then perform the act that you would perform if you believed in that same principle *and were omniscient*.

The actions recommended by PLP1 in many ordinary cases would be good ones from the perspective of the agent's favoured theoretical level principle. But PLP1 picks them out in an unhelpful way. If the agent does not know what maximizes utility, but thinks that the correct theoretical level principle directs him to do what maximizes utility, then it is not helpful to tell him (as PLP1 does) that he should do what an omniscient utilitarian would do. How is he supposed to know what an omniscient utilitarian would do? That act description is just as opaque as the original description – 'the act that maximizes utility'.

Here's another practical level principle that fails this first test:

PLP2: When you cannot identify the act that is required according to the theoretical level principle that you accept, then you should perform the act that *you believe to be required by the theoretical level principle that you accept*.

Although some philosophers[7] have suggested principles like PLP2, it fails to be helpful in a wide variety of cases. Suppose an agent accepts AU but knows that he lacks the information he needs in order to determine what is required by AU. Suppose he is intellectually cautious; he does not allow himself to believe, with respect to anything that he takes to be one of his alternatives, that it is the one required by AU. He withholds belief, since he knows that his evidence is insufficient to justify any belief. In this case, there is no act such that he believes that it is the one required by AU. Thus, it would be pointless to tell him that he ought to perform the act that he believes to be required by his favoured theoretical level principle – there is no such act.[8]

Another popular answer to our question is suggested by this:

PLP3: When you cannot identify the act that is required according to the theoretical level principle that you accept, then determine which act, of those

[7] There are several passages in E. Mason, 'Consequentialism and the "Ought Implies Can" Principle', *American Philosophical Quarterly* 40 (2003), pp. 319–31, in which Mason suggests that she means to defend an answer along these lines. In one place she says that when you don't know what you should do, you should try to maximize utility. She goes on to say: 'An agent counts as trying to maximize utility when she does what she believes will maximize utility' (p. 324).

[8] Suppose we say that the agent's alternatives are 'doing what will be best' and 'doing something that will be less than the best'. Then there is a pointless way in which he does know what he should do; he should do what will be best. But that act description is unhelpful; the agent does not know in practical terms what he is supposed to do in order to do what would be best.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

alternatives available to you, has the highest probability on your evidence of being the one that is required by your theoretical level principle; then perform that act.[9]

PLP3 confronts a whole range of objections. The most obvious is this: in many cases an agent may have incomplete and internally conflicting evidence; he may not have information about the likelihoods of various possible outcomes for different alternatives; he may even lack clear information about what alternatives are available. In such a situation, he will be unable to identify the alternative that has the highest probability on his evidence of being the one that is required by his theoretical level principle. So he will be unable to implement PLP3.

An even deeper problem with PLP3 is that in some cases we may know for sure that some act is *not* permitted by our favoured theoretical level principle, and yet in light of our ignorance, we may think that this is precisely the one that should be selected by our practical level principle. A good example of this is provided by Frank Jackson's case involving Dr Jill and her delightful patient, John.[10] John has a minor skin ailment. Three drugs – A, B and C – are available for treatment. Dr Jill knows that A will be good enough but with some minor adverse side effects. Dr Jill also knows that one of B and C will yield a perfect cure; the other will kill John. But she does not know, and knows that she cannot figure out, which is the perfect cure and which is the killer drug. In this case Dr Jill – a utilitarian – recognizes that she does not know which drug AU requires her to give – it is either B or C, but she can't tell which.

In this sort of case (depending upon how the details are spelled out) it could be reasonable to seek a practical level principle that will direct Dr Jill to prescribe A – the 'good enough' second best drug. But giving A is not permitted by AU; it is stipulated that Dr Jill knows that giving A will be less good than giving the perfect cure drug, which is B or C. So the probability that giving A is required by Dr Jill's theoretical level principle is zero. This shows that PLP3 is wrong. We sometimes think that morality requires us to fall back to a second-best action even though we know for sure that it is not permitted by our favoured theoretical level principle.

[9] There are passages in Mason, 'Consequentialism', in which she seems to endorse this approach, too. See, for example, p. 323 where she says: 'whenever we are given an instruction like [maximize the good], we ought to figure out which course of action is most likely to fulfill the instruction, and pursue that course of action'.

[10] This example was introduced in Frank Jackson, 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection', *Ethics* 101 (1991), pp. 461–82, at 462–3. Donald Regan described a case that illustrates the same point in D. Regan, *Utilitarianism and Co-operation* (Oxford, 1980), pp. 264–5. Parfit's example involving the trapped miners in Parfit, *On What Matters*, sect. 21, seems to illustrate the same point.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

Holly Smith has proposed another idea.[11] Let's say that the 'success rate' of a practical level principle is the percentage of cases in which the act recommended by that principle is the same as the act recommended by the agent's theoretical level principle. A 'usable' principle is one that is implementable. Smith's idea (roughly) is this:

PLP4: When you cannot identify the act that is required according to the theoretical level principle that you accept, then (a) determine which usable practical level principle has the highest success rate; (b) figure out what act that practical level principle requires; and (c) perform that act.

PLP4 may seem to evade the implementability problem; after all, it advises the agent to abide by a usable principle. But before the agent can abide by such a principle, he has to identify it. When he tries to identify the most successful usable principle, he will confront all the same implementability problems that created the difficulty in the first place. If he cannot tell which acts are required by the theoretical level principle that he accepts, he is in no position to determine how many such acts are also required by any proposed back-up principle. In other words, given that he has not solved the implementability problem, he will not be able to determine the success rate of any proposed back-up principle. If he cannot determine the success rates of these principles, he cannot identify the one he should try to follow.

I have now described several ways in which a proposed practical level principle could fail to be implementable. This gives some content to the first condition that must be satisfied by a proposed practical level principle: in order to be acceptable, a practical level principle must be implementable.

(b) Non-repugnance. I'm imagining that I think my actual moral obligation is always to do what maximizes utility. Sometimes I cannot figure out what that is, so I need a practical level principle to which I can turn in my ignorance. Surely it would be absurd for me to turn to a practical level principle that directed me to do something that I would find hopelessly morally repugnant. Even if it were easy to follow this practical level principle, I would be disgusted with myself if I allowed myself to be guided by it.

On the other hand, we cannot demand that the practical level principle should, in every case, direct me to do precisely the same thing that my theoretical level principle directs me to do. For we have stipulated that the practical level principle comes into play only when, because of epistemic deficiencies, I cannot identify the act required by the theoretical level principle that I accept. Surely it would be a miracle

[11] Smith, 'Subjective Rightness', pp. 100–6.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

if we could find an easy-to-use principle that would manage to pick out precisely this action in every case.[12]

So I need a practical level principle that will be easy to use, and that will direct me to do something that will be at least morally defensible from my perspective. If I act in accordance with such a principle, I may have to face up to the fact that what I did was not right according to my own theoretical level principle. But I will understand why I did it. I lacked some essential information. I knew that I would not be able to get that information. So I can console myself with the thought that I did the best I could in light of my epistemic deficiency.

The second condition that we place on proposed practical level principles is this: in order to be acceptable for a given agent, a practical level principle must not direct that agent to do something that will be morally repugnant from the perspective of the theoretical level principle that he or she accepts. It has to recommend a course of action that will be at least 'defensible' from the perspective of that principle.

(c) Morality. In order to satisfy the morality condition, a proposed practical level principle must give *moral* guidance. It's not as if, when we can't figure out what morality requires of us, we are encouraged to forget about morality and proceed to do instead something that etiquette, or prudence, or law, or sheer rationality requires. We are still looking for *moral* guidance, even though it will be guidance possibly different from the guidance given by our fundamental theoretical level principle.

This may seem paradoxical. If we think that our favoured theoretical level principle gives necessary and sufficient conditions for moral rightness, and we acknowledge that our practical level principle may sometimes direct us to do something different from what is required by our theoretical level principle, it may seem that the recommendation given by the practical level principle cannot be a *moral* recommendation. How can we have a moral obligation to do something different from what is required by the correct theory of moral obligation?

This is where the distinction between two levels of obligation may come in.[13] I will say that the theoretical level principle is intended to provide information about *moral obligation in the first instance* and that the practical level principle is intended to provide information

[12] This feature may be what motivated Smith to seek practical level principles that have outstandingly good success rates, rather than to seek practical level principles that have *perfect* success rates. It's also important to note that in the Dr Jill case just discussed, we expect the practical level principle to recommend a course of action (giving A) that is not recommended by the theoretical level principle. So we can't expect perfect extensional equivalence.

[13] It would be interesting to compare this distinction to Hare's distinction between critical level thinking and intuitive level thinking in his formulation of a Two Level view in R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford, 1981).

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

about *moral obligation in the second instance*. I abbreviate these as 'obligation1' and 'obligation2'.

Moral obligation1 is your obligation as determined by the correct theory of absolute moral obligation and the facts about how things stand objectively in the world; it is your moral obligation in the first instance. It is natural to think of this as a form of objective obligation. *Moral obligation2* is your moral obligation in the second instance, or your 'fall-back' obligation. If you are having trouble identifying the right action according to what you take to be the correct theory of moral obligation1, but you want to be a decent person, you want to avoid being morally blameworthy, you want to act at least in the spirit of the theory you believe, then you are probably trying to find out what is morally obligatory in the second instance. This sort of obligation is determined by how things appear from the perspective of the agent and so it is natural to think of it as a form of subjective obligation.[14]

With this distinction in place, we can understand how a person could recognize that he isn't going to be able to figure out what he morally ought to do, but at the same time sensibly hope to be able to act upon a distinctively moral recommendation concerning what to do. In order to make clear that there is no paradox here, he could phrase his hope this way: 'I want to do what I ought1 to do, but I can't get the information I need in order to figure out what it is. So, given that I cannot determine what I ought1 to do, what ought2 I do?'

There is nothing paradoxical about this. It may turn out that your moral obligation2 is different from your moral obligation1. What you really, absolutely, unconditionally objectively ought to do is what you ought1 to do; when you can't figure out what that is, and you seek a back-up recommendation pointing you towards something you will be able to do with a relatively clear conscience, then you are trying to discover what you ought2 to do. What you ought2 to do may be different from what you ought1 to do, but it will be about the best you can do given your unfortunate ignorance.

The third condition, then, is this: any acceptable practical level principle must give a *moral* recommendation for action; it must purport to say what *morality* requires2 of an agent as a back-up when that agent cannot identify the action that is morally obligatory1.

[14] It may appear that we can say that you have a subjective obligation to do something if you would have an objective obligation to do it if the world were precisely as you believe it to be. Parfit's concept of wrongness in the belief-relative sense as discussed in section 21 of his *On What Matters* seems to be like this. But this is problematic. Suppose your beliefs about the morally relevant features of your alternatives are 'gappy' – for some features, F, you neither believe that the situation has F nor that it lacks F. The world could not objectively be 'gappy' like that. Furthermore, it's hard to see how anything could be objectively obligatory in a situation in which the relevant alternatives were gappy with respect to a significant number of the properties that make for moral obligation.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

(d) Ideally, if a practical level principle directs an agent to perform some action, then it should be possible for the agent to perform that action. There might seem to be something strange about a moral principle that recommends a certain course of action, when in fact the agent will not be able to act on that recommendation. How helpful is a back-up plan when in fact it can't be followed?

Perhaps surprisingly, I think that it would be a mistake to impose this condition in this very robust form. Recall that principles at the practical level are supposed to recommend courses of action that are subjectively obligatory. That is, they are supposed to point us towards actions based on how things seem to us rather than based on how things are objectively.

Suppose a utilitarian thinks – mistakenly, as it happens – that he has certain alternatives; suppose it seems to him that one of them would be the best one to perform; but suppose that one is in fact one that he cannot perform. I want to say that from the subjective perspective of the agent, that alternative is obligatory2. Until he realizes that he can't do it, that's the one that he ought2 to aim for. As soon as he realizes that he can't do it, something else will become his obligation2.

So while we cannot endorse an objective version of the 'ought implies can' principle for practical level obligation, we can endorse a subjective version of that principle: if, as of some time, t, an agent, S, has a practical level obligation to perform an act, a, then, as of t, S must think that a is one of his alternatives. An acceptable practical level principle should be consistent with this condition, too.

(e) An adequate practical level principle must provide a way for the agent to avoid at least certain sorts of blame.[15] In some cases, a person is blameworthy for having done a certain act largely because he really could have done better; out of laziness, or selfishness, or lack of concern with morality, he just took the easier or more selfish path. Any of these conditions may make the person blameworthy. A conscientious person would want to be able to avoid that sort of blame. This is where practical level principles come in.[16]

[15] I recognize that my statement of this condition is vague. In her discussion of her Criterion 4, Holly Smith is similarly vague. Like me, she sees some connection between subjective obligation and blameworthiness, but avoids committing herself to any fully precise principle. She says that the concept of subjective rightness 'should bear appropriate relationships to assessments of whether the agent is blameworthy or praiseworthy for her act' (Smith, *Subjective Rightness*, p. 73).

[16] In M. Zimmerman, 'Is Moral Obligation Objective or Subjective?', *Utilitas* 18 (2006), pp. 329–61, at 329, Michael Zimmerman says: 'conscientiousness precludes doing what one believes to be overall morally wrong'. I think this is not quite right. Dr Jill might believe that giving A is overall wrong; that giving B or C is overall right; but still she might think that under the circumstances it will be permissible for her to give A. If she refuses to do something she takes to be wrong, she runs the risk of doing something terrible. So, even though she is conscientious, she knowingly does something she takes to be overall morally wrong.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

In a case in which a person accepts a certain theoretical level principle, but realizes that he cannot get the information he needs in order to fulfil its recommendation, he may want a back-up principle that will direct him to a course of action such that if he does it, he will be able to defend himself against accusations of laziness, or selfishness, or lack of concern with morality. He will be able to say that it was impossible, at the time, for him to get the information he needed in order to identify in a helpful way the act that was really obligatory1; so, out of a concern for morality, and in an attempt to do the best he could under the circumstances, he fell back upon his practical level principle and did what it declared to be obligatory2. Where this sort of response is appropriately in place, blame of the sort envisioned is evaded.[17]

So the final condition concerning practical level principles is this: they should give recommendations for action such that, if the agent successfully acts on those recommendations, he will not be open to blame of the sorts described here.

IV. A FANTASTIC DIGRESSION

Suppose a conscientious moral agent believes in a form of act-utilitarianism and wants to do the right thing. Suppose also that this agent recognizes that she does not know the things she would need to know in order to apply the theory to her present predicament. She is morally perplexed. But suppose in addition that this agent has the opportunity to consult with a Utilitarian Moral Guide. The guide is a clear thinker who fully understands the workings of AU; he does not have any factual information beyond that available to the agent. Nevertheless, he is willing to help.

Let's assume that the perplexed utilitarian is Dr Jill (from the Jackson example). She explains her perplexity to the Utilitarian Moral Guide. The guide insists at the outset that Dr Jill describe her alternatives in helpful, action-guiding terminology. Thus he says, 'When you tell me about what you take to be your alternatives, be sure to describe each alternative in such a way that, if you subsequently decide that it is the one you should do, you will have no epistemic trouble about implementing your decision.' Dr Jill then explains her alternatives (Pills A, B and C) and she goes on to say that she

[17] This is not to suggest that blame of all sorts will be avoided. It's easy to imagine cases in which a person deserves blame for being ignorant of certain important morally relevant facts. At an earlier time, we may suppose, he could have obtained the relevant information. Now it is unavailable. As a result he cannot determine his obligation1, and falls back on a practical level principle. He does what is obligatory2. The condition then says that the agent cannot be blamed for doing the obligatory2 act; it allows that he might be blameworthy for failing to have learned the morally relevant facts at the earlier time.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

wants to give her patient, John, the utility maximizing treatment. Her problem is that she cannot get the information that would enable her to determine in a helpful way which of the three available prescriptions in fact would maximize utility. She is certain that pill A will give John a pretty good but less than ideal cure. She believes that one of B and C is a perfect cure pill that would lead to a much better outcome, and the other of B and C is a killer drug that would lead to a much worse outcome. Her problem is that she doesn't know which is the cure and which is the killer.

The Utilitarian Moral Guide recognizes that Dr Jill is afraid that if she tries to do what's best she may end up doing what's worst. After getting the information about Dr Jill's views about her alternatives and their values, he asks her to explain her view about the morality of putting John at this particular risk of death.

Suppose Dr Jill responds by saying that she thinks it would be morally wrong in this particular case to put John at serious risk of death unless it is absolutely necessary to save his life.

The Utilitarian Moral Guide then makes his recommendation. He says: 'Among the things that you take to be your alternatives, throw out all the ones that seem to you to involve exposing John to unwarranted risk of death. Then, among the remaining alternatives, select the one that seems to you to be best. Do that. (Or if several of your remaining alternatives seem to be tied for first place, then just pick one of them at random and do it.)'

So in this example (which I will call 'Case A'), the Utilitarian Moral Guide recommends that Dr Jill give Pill A.

Note that the act selected in Case A is not the act that maximizes actual utility. As a result, it should be clear that the policy behind the Utilitarian Moral Guide's recommendation is not equivalent to AU. Nor is the policy behind the Moral Guide's recommendation equivalent to the idea that the Perplexed Agent should just go ahead and do whatever seems best. In this case, giving Pill A does not seem best to Dr Jill. It seems to her that giving A is second best, and that either giving B or giving C would be best; she just doesn't know whether it's B or whether it's C.

We should also note something about expected utility. While the act recommended by the Utilitarian Moral Guide is probably the one that maximizes expected utility, the Utilitarian Moral Guide did not tell Dr Jill that she should do what maximizes expected utility.[18] Telling

[18] I say that giving A 'probably' maximizes expected utility in this case. We can't tell for sure. Whether it does depends upon the details of the probabilities and values of the alternatives. In this example I have deliberately stipulated that Dr Jill does not have any beliefs about the precise numerical ratings either for the values or for the probabilities of the outcomes.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

her that would not have been helpful, since she does not have the information about the probabilities and values of outcomes that would be necessary for calculating expected utilities.[19] We may even assume that Dr Jill lacks the concept of expected utility.

We can make some further comments on the dialogue by considering the extent to which the recommendation given by the Moral Guide satisfies the conditions that I stated earlier.

Condition (a): *Helpfulness.* The recommendation given in this case by the moral guide satisfies the helpfulness condition. It tells Dr Jill which pill she should2 give and it gives her this recommendation in terminology that will make it easy for her to figure out what she is supposed2 to do. This must be the case since it is stipulated that when Dr Jill initially asked for assistance, she was required to describe her alternatives in terms that would subsequently be helpful to her. When she gets her recommendation, it will have to specify one or more of those alternatives, described in precisely those helpful terms that she herself provided at the outset.

Condition (b): *Non-repugnance.* In this case the Utilitarian Moral Guide recommends that Dr Jill give pill A. Of course, act-utilitarianism implies that it's wrong1 to give pill A; but since Dr Jill has no way of knowing what pill she ought1 to give, and pill A is safe and fairly effective, Dr Jill, though a utilitarian, should find this to be morally acceptable guidance.

In light of her replies to the Utilitarian Moral Guide, we can see that Dr Jill is not *simply* a utilitarian. Her moral view is somewhat more nuanced. She still believes that morality requires1 her to do what's best in each situation; but in addition she evidently believes that when she does not know what's best, morality requires2 her to behave in a way that is appropriately sensitive to risk. In the case at hand, when reflecting on the magnitudes of these particular risks, she thinks that the possibility of inflicting serious harm on John is just too great. Because she holds this more comprehensive moral view, giving the second-best drug does not seem morally repugnant to her in this case.

Condition (c): *Morality.* The guidance given by the Utilitarian Moral Guide in this case does indeed constitute *moral* guidance. He

[19] One of the most blatant violations of the helpfulness condition occurs in a popular proposal involving act-utilitarianism. Some assume that act-utilitarianism gives a fair account of our obligations1. They go on to suggest that when you don't know what maximizes actual utility, and hence do not know what you ought1 to do, you ought2 to do what maximizes expected utility. As I argued in 'Actual Utility, the Objection from Impracticality, and the Move to Expected Utility', *Philosophical Studies* 129 (2006), pp. 49–79, while it is hard to know what maximizes regular utility, it is even harder to know what maximizes expected utility. As a result, telling a person that he ought2 to do whatever maximizes expected utility will often be unhelpful; so this sort of answer violates the helpfulness condition.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

recommends a course of action for Dr Jill that will be acceptable to her as a morally conscientious advocate of AU. The guide has not digressed; he has not slipped into talking about what would be prudent, or what would be lawful. He is talking about the requirements of morality (though, of course, he is focusing on moral requirement *in the second instance*).

Condition (d): *Possibility.* Dr Jill thinks that giving pill A is one of her alternatives. In fact, it is; she can give pill A. So the recommendation given by the Utilitarian Moral Guide does not tell her that she ought2 to do something that she can't do.

Condition (e): *Blameworthiness.* The recommendation given by the moral guide in this case purports to give a recommendation for action, based upon what Dr Jill believes and knows about her situation. I am inclined to think that the recommendation directs her to do something such that if she were to do it, it would not be reasonable to blame her for doing it. I think the blameworthiness condition is satisfied in this case. But I recognize that this is a tangled question. More needs to be said about it.

Let me now state my general conclusion about this example: the recommendation given by the Utilitarian Moral Guide is plausible. In fact it does seem that Dr Jill's fall-back obligation2 in this case is to give Pill A. She will be able to adopt the Guide's recommendation, and she will retain her status as a morally conscientious person if she does so. While some might want to blame her for being in such a pickle in the first place, no one could reasonably blame her for following the Guide's advice when in the situation as described. By following that advice, she would be doing the best she could given her ignorance of some morally relevant information.

Some variations on the theme

Now let us turn to some examples that illustrate what happens when the agent is in slightly different subjective situations. Imagine that the dialogue in Case B starts out just like the dialogue in Case A. Dr Jill has the same patient, the same set of drugs, the same troubling lack of information about the merits of B and C. But when the utilitarian guide gets to the part about the morality of risk, the discussion takes a different turn:

Utilitarian Moral guide: You evidently want to do what's best, but you are afraid that if you try to do what's best you will end up doing what's worst. I need one further bit of information: what are your views about the morality of putting your patient at serious risk of death in this situation?

Perplexed Agent: My considered opinion is that a partial cure in this case is morally unacceptable. I think I have to go for broke. I am not afraid of putting John at risk, if that's required in order to have a shot at a perfect cure. I do

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

not think it would be right to do something that will end up with John only partially cured when there is a chance of getting a perfect cure.

Utilitarian Moral Guide: I won't comment on your policy concerning this sort of case. I will simply tell you what, given that you have those moral views, you ought2 to do. Among the things that you take to be your alternatives, toss out all the ones that insofar as you understand the situation, would at best lead to a partial cure. In addition, toss out any that would definitely lead to death. The remaining alternatives seem to you to be ones that might lead to a perfect cure. That would be B and C. Since you have no basis to choose between B and C, you can simply choose at random between them. Flip a coin if you like. No matter what you choose, there is a chance that John will be perfectly cured.

So in Case B, the Guide recommends that Dr Jill ought2 to give either B or C.

The distinctive feature of Case B is Dr Jill's view about the morality of risk. Here, as before, she thinks morality requires1 her to aim for the best outcome. The distinctive feature in this case is this: when Dr Jill reflects on the amounts of harm and benefit that might result from the different courses of action that she takes to be available, she thinks morality calls upon her to expose her patient to this particular risk in order to achieve the best possible result. She thinks that this is what she has to do as a conscientious person. Given that she has these moral views, it's permissible2 for her to give B or C.

Maybe she shouldn't have those views. Maybe she should be more risk averse – especially when it is John who is going to be exposed to that risk. But she isn't. She thinks it's worth the risk even though there's no certainty of a complete cure. I do not have these views about risk, but given that she sincerely does, it seems permissible2 for her to act on them. Perhaps it was not permissible1 for her to have allowed herself to have those views. However, in the present instance we are evaluating her actions based on her current mental state; we are not evaluating the processes by which she got into her current mental state.

Let us now turn to another sort of case.

In Case C, Dr Jill doesn't have any idea of the utility of any alternative. Nor does she have any views about relative magnitudes of the utility differences between alternatives. However, she thinks she has certain alternatives; and among these there are some that she thinks would be better than others. She has no views about the extents to which the different alternatives would put people at risk of serious harm. After eliciting this information, the Utilitarian Moral Guide recommends that Dr Jill pick at random among the ones she thinks might be best.

Case C is intended to highlight an important difference between my approach and some other popular approaches. Note that in Case C, Dr Jill does not have any beliefs about the mathematical features of

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

the case, aside from the fact there are certain alternatives that (as she sees it) would lead to better consequences than others. Thus, it would be pointless for the Utilitarian Guide to recommend that Dr Jill perform the act that maximizes utility, or the act that maximizes expected utility, or the act that she thinks will maximize utility or expected utility – for in this case there are no acts about which Dr Jill has any such beliefs. Given her very limited information, the best recommendation the Guide can give is just to choose an alternative that, as she sees it, might be among the best available. I think that if Dr Jill follows this advice, she will be in the clear morally. She will have done something permissible2.

V. A TWO LEVEL MORAL THEORY

As I have described him, the Utilitarian Moral Guide has certain features. He is calm and methodical; he understands act-utilitarianism; he is always willing to engage in dialogue with perplexed utilitarians. But it is important to recognize that the Utilitarian Moral Guide is not omniscient about empirical facts. He doesn't know more about the perplexed agent's situation than the agent herself knows.[20] The Utilitarian Moral Guide helps by focusing the agent's thoughts on the considerations that are relevant in her perplexity.

In every case, the Utilitarian Moral Guide's recommendation would be consistent with certain policies:

(1) If the agent is convinced that certain alternatives are better than others, then other things being equal, the Utilitarian Moral Guide would recommend that she perform one of the ones the agent takes to be among the best available.
(2) If the agent doesn't have an actual ranking of alternatives, but thinks that some alternatives are riskier than others, and thinks that in the case at hand morality requires her to avoid putting people at risk of harm, and believes that there are alternatives that would avoid putting anyone at serious risk of serious harm, then, other things being equal, the Utilitarian Moral Guide would recommend that she perform one of the ones she takes to be less risky.
(3) Where the implications of (1) seem to conflict with the implications of (2), the Utilitarian Moral Guide would try to elicit from the agent some indication of her feelings in her current

[20] Thus I am not defending a version of the Ideal Observer Theory (as typically understood). The Utilitarian Moral Guide understands AU; he is calm and methodical. But he does not have access to any factual information beyond that available to the perplexed agent.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

case of the relative moral importance of doing what's best versus avoiding risk. He will recommend that the agent abide by the policy that she thinks is more important in the present instance.

(4) If the perplexed agent has no clue about the values of alternatives, then the Utilitarian Moral Guide would recommend that the agent pick at random.

We might think that the Two Level Theory suggested by these fantastical reflections would be something like this:

Level 1: You morally ought1 to perform an act iff it maximizes utility.

Level 2: If you cannot determine what you morally ought1 to do, then you morally ought2 to perform the act that the Utilitarian Moral Guide recommends.

Of course there really isn't any Utilitarian Moral Guide and so there is no such act as the act that the Utilitarian Moral Guide recommends. Thus it would be impossible to implement the Level 2 component of this theory. Furthermore, it would not help to say that you ought to perform the act that the Utilitarian Moral Guide would recommend if he were to exist. The problem here is that the perplexed agent might not know what a Utilitarian Moral Guide is, and might not be able to figure out what such a creature would recommend if he were to exist.

Fortunately, all this talk of the Utilitarian Moral Guide is mere heuristic and can be eliminated. In order to state the actual view without resorting to fantasy, we must describe a decision procedure that the perplexed agent can follow on her own, without the help of a Moral Guide. Since the agent under consideration here is a utilitarian, she would need a Utilitarian Decision Procedure. It goes like this:

Step One: Consider the acts that you take to be your alternatives – described in 'helpful', 'action-guiding' terms.[21]

Step Two: Consider, insofar as your epistemic state permits, what you take to be their values or perhaps just their relative values.[22]

[21] Note that Step One does not require the agent to *list* all of her alternatives. There might be millions of them. In many cases it will be sufficient for the agent to consider whole groups of alternatives under suitable general descriptions. For example, suppose an agent has been asked to pick a number between one and one million. There is no need for her to consider picking one, picking two, picking three, etc. Since she will have no epistemic trouble in any case, she can describe her alternatives as a group by saying 'I have to pick a number between one and one million.' This will be sufficiently action guiding.

[22] Step Two does not require that the agent consider each individual alternative separately. In the numbers case imagined in the previous footnote, she might simply consider that there is no number, n, such that her evidence gives her reason to suppose that picking n will yield more utility than picking any other number.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

Step Three: If you haven't got useful information about the actual values of your alternatives, then consider how your views about the morality of risk apply to your present situation; and, in light of all this,

Step Four: identify the acts in this particular case that seem most nearly consistent with the general policy of maximizing utility where possible while avoiding things that put people at excessive risk of serious harm; and then,

Step Five: perform one of them.

Conscientious use of this decision procedure would yield a conclusion about what should2 be done. The procedure constitutes moral guidance rather than etiquettical or legal or prudential guidance; even if the resulting guidance would not be equivalent to the implications of act-utilitarianism, it would not be morally repugnant from the perspective of a utilitarian. The guidance would emerge in helpful terms, so that the agent would know how to perform the designated act; the agent would at least think that she will be able to perform the recommended action; and if the agent were legitimately unable to determine the implications of act-utilitarianism for his current situation, and were to make use of and then act upon the output of this decision procedure, he would not be open to moral blame in the specified ways. Hence, this proposal satisfies many of the conditions laid out at the outset – or at any rate *versions* of those conditions.

With all this in place, I can now state a Two Level Theory with AU at the first level:

Level 1: You morally ought1 to perform an act iff it maximizes utility.

Level 2: If you cannot determine what you morally ought1 to do, then you morally ought2 to perform an act iff it is an outcome of the Utilitarian Decision Procedure.

VI. EXTENSIONS TO ROSSIANISM AND VIRTUE ETHICS

According to the normative theory of Sir William David Ross (as presented in *The Right and the Good*) an action is morally right iff it maximizes net prima facie rightness.

Ross thinks that there are certain features of actions that tend to make those actions morally right. Among these he lists *being the keeping of a promise*, *being a case of justly distributing some good or evil, being a case of showing gratitude, being a case of making reparations for a past misdeed, being a case of conferring a benefit on someone.* These are the prima facie rightmaking characteristics. An act can have one or more of these, or it can have none. For each such characteristic, an act can have it to a great extent, or to a smaller extent. We can give each act a score indicating the total extent to which it has each of the

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

prima facie rightmaking characteristics. We can sum these extents. We can call that the act's 'pfrightness'.

The same is true with respect to prima facie wrongmaking characteristics. We can call the sum of these in any case the act's 'pfwrongness'.

If you subtract an act's pfwrongness from its pfrightness, you get its net pfrightness. The right act according to Ross is the one in a set of alternatives that maximizes this value. Here is the famous passage in which Ross states 'the universal nature of all acts that are right':

> right acts can be distinguished from wrong acts only as being those which, of all those possible for the agent in the circumstances, have the greatest balance of prima facie rightness, in those respects in which they are prima facie right, over their prima facie wrongness, in those respects in which they are prima facie wrong.[23]

It should be clear that typical agents, in typical cases, are not able to determine what they should do by direct application of Ross's theory. That's because typical agents do not know precisely what alternatives are available to them; and even if they know this, they do not know precisely how much pfrightness and pfwrongness each alternative has. In many cases they do not know which, among their alternatives, has more net pfrightness than the others. So they are in an epistemic pickle.

We can imagine a Perplexed Moral Agent who has accepted Ross's Theory. He wants to do the right thing, but does not know what it is. He seeks help from a Rossian Moral Guide. Their dialogue looks like this:

> *Perplexed Agent:* I believe that I should1 do what maximizes net pfrightness, but because of ignorance concerning the details of my alternatives and their amounts of pfrightness and pfwrongness, I don't know what I should1 do in my current specific situation. I'd like to be a decent person, but I'm perplexed. What should2 I do?

> *Rossian Moral Guide:* I will need to know certain things about what you take your alternatives to be; and, insofar as is possible, I will need to know something about what you take the net pfrightness of these alternatives to be. In addition, if you have strong views about the morality of risk, I will have to know about that, too.

Of course, you believe that what you really ought1 to do is whatever maximizes net pfrightness. You recognize that you don't know what that is but still you think that that's really what morality requires1 of you at all times. I cannot help you figure out what that is. So if I give you some help, it will concern only what you ought2 to do given that you think you ought1 to do the act that maximizes net pfrightness, but you are ignorant in the ways you will specify

[23] Ross, *The Right and the Good*, p. 41.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

and are thus unable to figure out what you ought1 to do. My recommendation will be limited to what you ought2 to do.

The Rossian Moral Guide will then proceed in pretty much the same way as the Utilitarian Moral Guide. He will elicit information about the agent's beliefs about what alternatives are available to him, and what their net pfrightness is. Where the agent does not have beliefs about that, the Rossian Moral Guide will ask about beliefs about alternatives that are particularly high-ranking in terms of net pfrightness. He may ask about which pfobligations the agent takes to be most important in this case. If the agent is uncertain about the extent to which certain acts might involve serious violations of prima facie obligations, the Guide will ask for estimates of likelihood and for an expression of the Agent's views about the morality of running that sort of risk in the case at hand. In the end he will make a recommendation based on the agent's conception of his current moral perplexity.

As before, all this talk of moral guides is mere heuristic. It can be eliminated. According to the non-fantastical version of the theory, if an agent accepts Ross's theory, he will think that his moral obligation in the first instance is to perform an act that maximizes net pfrightness. When he cannot figure out what act that is, he may fall back on his moral obligation in the second instance. That act will be the output of the Rossian Decision Theory. The Rossian Decision Theory will be structurally just like the Utilitarian Decision Theory described above in section V, but with the obvious substitution of the Rossian concepts for the utilitarian concepts.

According to another theory in the normative ethics of behaviour, actions should be evaluated by appeal to the amounts of virtue and vice that the agent would manifest in performing them. Thus, for example, suppose someone has a chance to save a baby from a burning building. If he rushes in and pulls out the baby, his action will manifest a lot of courage. It will also manifest a certain amount of benevolence. If these are virtues, then the act gets some positive points for manifesting these virtues. The precise number of points would be determined by the precise magnitudes of the virtues being manifested. An act that manifests great courage would get more points for that than would an act that manifests just a tiny bit of courage.

Suppose someone has a chance to skulk away and avoid getting involved in saving anyone from a fire. Suppose he does this because he is afraid of getting injured. Then his act manifests the vice of cowardice. As a result, the act gets some negative points for manifesting this vice. The precise number of points would be determined by the precise amount of cowardice being manifested.

Suppose we have a list of all the virtues and vices. Suppose we have a way of determining, for each possible action, how much virtue it would

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

manifest and how much vice it would manifest. Assume that these amounts can be represented numerically. We could then subtract the total amount of vice that would be manifested by some possible action from the total amount of virtue that would be manifested by it. The result is the VV Quotient of the action. To say that an act maximizes this value is to say that no alternative has greater VV Quotient than it has.

A plausible form of Virtue Ethics then says:

VE: An act is morally right iff it maximizes VV Quotient.[24]

It should be clear that typical agents, in typical cases, are not able to determine what they should do by direct application of VE. That's because typical agents do not know precisely what alternatives are available to them; and even if they know this, they often do not know, for each alternative available, precisely how much virtue and vice they would be manifesting if they were to perform that alternative. More deeply, they may not know for certain which character traits are virtues and which are vices. As a result of all this, they rarely can know which, among their alternatives, has greater VV Quotient than the others. So if they have accepted VE, they will often be unable to identify the alternative required by their own moral theory.

We can imagine a Perplexed Moral Agent who has accepted Virtue Ethics. He wants to do the right thing, but does not know what it is. He seeks help from a Virtue Ethical Moral Guide. Their dialogue looks like this:

Perplexed Agent: I believe that I should1 do what maximizes VV Quotient, but because of ignorance concerning the details of my alternatives and the virtues and vices I would be manifesting if I were to perform them, I don't know what I should1 do in my current specific situation. I'd like to be a decent person, but I'm perplexed. What should2 I do?

Virtue Ethical Moral Guide: I will need to know certain things about what you take your alternatives to be; and, insofar as it is possible, I will need to know something about what you take the VV Quotients of these alternatives to be. In addition, if it seems to you that some of your alternatives might involve manifesting a lot of vice, I will have to know about those alternatives and I will have to know how you feel about the morality of manifesting vice in this particular case, too. If you tell me all this, maybe I can help.

The Virtue Ethical Moral Guide will try to elicit information about the agent's beliefs about his alternatives and the amounts of virtue and vice he would manifest in their performance. He will also try to

[24] My formulation of Virtue Ethics derives from things that Daniel Doviak says in his 'A New Form of Agent-Based Virtue Ethics', *Ethical Theory and Moral Practice* 14.3 (2011), pp. 259–72.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press

elicit information about the agent's views about the nature and relative importance of different virtues. 'Do you think that courage is a virtue? How about patriotism? How about piety? And if these conflict, do you have a view about which is more important in the present instance?' He will also ask the agent about his attitude towards risk – in this case the risk of manifesting a virtue or a vice in the performance of some action. Having elicited these views from the agent, the Virtue Ethical Moral Guide will make a recommendation much like the one made by the Utilitarian Moral Guide or the Rossian Moral Guide. He will give an appropriate recommendation concerning the agent's moral obligation2 on the assumption that VE gives the correct account of moral obligation1.

As in the cases of act-utilitarianism and Rossianism, all mention of the moral guide can be eliminated. If a person believes in Virtue Ethics, he will think that his obligation in the first instance is always to do the act that maximizes VV Quotient; his obligation in the second instance is to perform an act that is the outcome of the Virtue Ethical Decision Procedure. Although there is not sufficient space to explain it here, my view is that the answer given in the Rossian case as well as in the Virtue Ethical case would be consistent with the conditions that I outlined earlier.

The Two Level structure that I have described can be extended so as to apply to various forms of Kantianism, Rights Theory, and other views in the normative ethics of behaviour. I leave it to the interested reader to work out the details.[25]

ffeldman@philos.umass.edu

[25] I am grateful to several friends for helpful criticism and suggestions. I especially want to thank Kristian Olsen, Owen McLeod, Casey Knight, Chris Meacham, Pete Graham, Brad Skow and Michael Zimmerman. Papers by Holly Smith, Elinor Mason and Michael Zimmerman on this topic have also been very helpful. An ancestor of this article was presented as the Keynote Address at the annual meeting of the Creighton Club at Hobart and William Smith College in Geneva NY on 13 November 2010. Another version was defended at a meeting of the New England Consequentialism Workshop at the Safra Center of Harvard University on 9 February 2011. Another version was presented at Huron University College in London Ontario on 28 October 2011. This last version was also presented as the Keynote Address at the November 2011 meeting of the New Jersey Regional Philosophical Association. I am honoured to have had the invitations and I am very grateful to the participants in those discussions for many helpful comments.

https://doi.org/10.1017/S0953820812000155 Published online by Cambridge University Press