

# *What To Do When You Don't Know What To Do*

*Fred Feldman*

## *1. A Puzzle in Ethics*

The fundamental project of normative ethics is the attempt to discover, properly formulate, and defend a principle stating interesting necessary and sufficient conditions for the moral rightness of actions. More simply: the attempt to figure out what makes right acts right. Lots of philosophers have come up with lots of competing views in this area. Among the more popular views are many forms of utilitarianism, Kantianism, and Rossianism, as well as the Divine Command Theory, Virtue Ethics, Social Relativism, and various forms of Rights Theory.

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 the principle he accepts. Yet no one has the kind of detailed information that would be required to implement any of these principles. Thus, these principles are practically unhelpful: we cannot use any of them in real life to determine concretely what we are supposed to do. Some put this point by saying that the principles fail to be “action-guiding”.

Let me illustrate this with a story about something that happened to me. Some friends of mine formed a biotechnology company. They were going to engage in some very tricky business involving the insertion of some human genes into the DNA of cows; then they were going to clone the resulting calves. They hoped they would be able to extract a human-like blood component from the cows. This could be used in human medical interventions. They thought (correctly, as it turned out) that they could save lives and make a fortune. They asked me to serve as chair of the company's ethics advisory board. They thought I would be a natural for the job, since I had defended a certain theory about

right action – a form of utilitarianism -- for a long time. I would just have to apply the theory to the company's operations and declare that what they were doing was morally OK.

Since these people were dear friends, I agreed to serve. But I immediately ran into insuperable problems. They wanted me to write up a report saying that in my professional opinion, there was nothing morally wrong with their project of inserting some human genes into cow DNA. I found myself utterly stymied. I did not know what would happen if the genetic manipulation were undertaken. I did not know if it would work and save lives, or if it would lead to some unexpected disaster. I did not know the probabilities of any outcomes. I did not know the values of the outcomes (though some imagined outcomes seemed pretty wonderful, and others seemed pretty horrendous). So, though I had a moral theory, I did not have (and knew that I could not get) the information I would need in order to apply the theory to the case about which I had to make a pronouncement.

I resigned my position on the board.

This problem is not confined to act utilitarianism. Many other theories in normative ethics confront epistemic problems just as intractable as those faced by act 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. The same is true of various forms of virtue ethics. The implementability problem confronts anyone interested in normative ethics, regardless of the normative theory he or she holds.

## ***2. Two-Level Theories in General***

One possible solution to this problem involves a move to a "Two Levels Approach". If we adopt this sort of approach, we will say that a complete theory in NEB involves two distinct components. The first component is the familiar normative principle – that

component states alleged necessary and sufficient conditions for the moral rightness of actions, just as traditional moralists have assumed. That principle – whatever it may be – will be unimplementable. The second component of the theory will be a decision procedure, or a handy and useful guide to action. It will be something that will be immediately helpful as the agent tries to choose actions. It will have to be implementable. When the two components are properly merged, 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”.<sup>1</sup>

In any plausible package, the two components must be properly connected. For each theoretical level principle, there must be a certain practical level principle that is correct, or appropriate, or 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.

### *3. 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 will be 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 and popular (and because I think something like this is true), I will use act utilitarianism as my sample theoretical level principle. The claims I make about the features of the associated practical level principle

---

<sup>1</sup> 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.

in this case should carry over pretty directly in other cases in which we start with a different theoretical level principle.

According to act utilitarianism (AU), an act is morally right if and only if it maximizes utility. It should be obvious 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.<sup>2</sup> In spite of this, many of us take it to be the true theory in normative ethics. When we consider possible cases described in complete detail, it seems to us that the right thing to do is always whatever would be best. Let's not debate that question here.<sup>3</sup> I introduce AU primarily as an example of a possible theoretical level principle.

We seek one main feature in a theoretical level principle: it should be true. It should state 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 practical level principle if it is to be a suitable partner for AU as the theoretical level principle in a Two-Level Theory.

The first condition is:

a. Usefulness; implementability; applicability. 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. My theoretical level principle is not practically helpful in this situation. So I need some practical level principle that will offer guidance in this condition of irremediable ignorance. Obviously,

---

<sup>2</sup> The classic discussion of this point can be found in Moore's discussion in *Principia Ethica* ([1903] 1993: 211-214 (sect. 99)). See also Frazier (1994). Bales (1971) contains a nice discussion of how the implementability problem arises for AU. Others have written extensively on this more recently, but no one has added anything of substance.

<sup>3</sup> I have debated it elsewhere. In fact I defend a variant form of AU based on the idea that we should adjust the values of consequences to reflect the extent to which recipients deserve the goods and evils that they receive in those consequences. See my "Adjusting Utility for Justice" for details.

in this situation, it would be pointless to turn to another principle that other principle is just as hard to implement as AU. Thus, the associated practical level principle must be “helpful”, “useful”, “implementable”, “action-guiding”. This may not be entirely clear.

Let us consider some defective practical level principles that fail because they characterize the agent’s obligation in unhelpful terms. Consider this:

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

This principle certainly does recommend some actions for the agent to perform. And the recommended actions would undoubtedly be good ones from the perspective of the agent’s favored 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 will be completely pointless to tell him (as PLP1 does) that he should do what an omniscient utilitarian would do. How he is 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 AU, then you should perform the act that *you believe to be required by AU*.

Although some philosophers<sup>4</sup> have suggested this solution to our puzzle, it will not work. Suppose an agent 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

---

<sup>4</sup> There are several passages in Mason (2003) in which she suggests that she means to answer our question with some answer along these lines. Thus, for example, in one place she first 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” (2003: 324).

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 and unhelpful to tell him that he ought to do what he believes to be required by AU – there is no such act.<sup>5</sup> Suppose we say, “Do the act that you think is right.” He will reply, “But that’s my problem! There isn’t any act that I think is right!”

Another popular answer to our question is suggested by this:

PLP3: When you cannot identify the act that is required according to AU, then determine which act, of those alternatives available to you, has the highest probability on your evidence of being the one that is required by AU; then perform that act.<sup>6</sup>

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 AU. 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 AU, 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 example involving Dr. Jill and her delightful patient, John.<sup>7</sup> John has a minor skin ailment. Three drugs – A, B, and C – are

---

<sup>5</sup> 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 kind of pointless way in which he does know what he should do; he should do what will be best. But that act description is completely unhelpful; the agent does not know in practical terms what he is supposed to do in order to do what would be best.

<sup>6</sup> Mason also said something like this?

<sup>7</sup> This example originated in Jackson (1991: 462-463).

available for treatment. Dr. Jill knows that A will be good enough but with some minor side effects. One of B and C will yield a perfect cure; the other will kill John. Dr. Jill 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 her theoretical level principle 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 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 want to fall back to an action even though we know for sure that it is not permitted by our favored theoretical level principle.

I have now described several ways in which a proposed practical level principle could fail to be helpful, or useful, or implementable.<sup>8</sup> This (I hope) 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 helpful, or useful, or implementable.

---

<sup>8</sup> Holly Smith has proposed another idea. Let's say that the “success rate” of a back-up principle is the percentage of cases in which the act recommended by the back-up principle is the same as the act recommended by the agent's theoretical level principle. 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 back-up principle has the highest success rate; (b) figure out what act that back-up 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 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. And if he cannot determine the success rates of these principles, he cannot identify the one he should try to follow.

b. Non-repugnance. 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 that maximizes utility. Surely it would be a miracle if we could find an easy-to-use principle that would manage to pick out precisely this action in every case.<sup>9</sup>

So I need a 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 challenged, I will have to grant that what I did was in fact not right according to my own principle. But I will have an excuse. I lacked some essential information. I knew that I would not be able to get that information. I did the best I could in light of my epistemic shortfall. So, while my action fell short according to my own moral principle, it would be unfair to blame me for having done it, or to claim that I should have tried harder or otherwise should have done better. I did the best I could under the epistemic circumstances.

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.

---

<sup>9</sup> This feature may be what motivated Smith to seek practical level principles that have outstandingly good success rates.



c. Morality. This brings us to a closely related further condition. This may be described as the “morality” condition. In order to satisfy this 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 then 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 AU 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 AU, 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 objective and subjective obligation may come in;<sup>10</sup> or perhaps where the distinction between “critical level obligation” and “intuitive level obligation” may come in.<sup>11</sup> I prefer to use new terminology here, so as to avoid confusing my distinction with the possibly distinct distinctions that others have already made in the literature. I prefer to say that the theoretical level principle provides information about *moral obligation in the first instance* and that the practical level principle provides information about *moral obligation in the second instance*. I abbreviate these as “obligation1” and “obligation2”.

*Moral obligation1* is the obligation codified by the correct theory of absolute moral obligation; it is your moral obligation “in the first instance”. If someone believes in AU, he thinks that AU gives the correct account of necessary and sufficient conditions for moral rightness1. *Moral obligation2* is your moral obligation “in the second instance”, or your “fall-back” obligation. If you are having trouble identifying the right action

---

<sup>10</sup> Cite Frances, Holly, Ellie and others here.

<sup>11</sup> Cite Hare here.

according to what you take to be the correct theory of moral obligation<sup>1</sup>, 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.

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 have 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 ought<sup>1</sup> to do, but I can't get the information I would need in order to figure out what it is. So, given that I am probably not going to do what I ought<sup>1</sup> to do, what ought<sup>2</sup> I do?"

There is nothing paradoxical about this. It may turn out that your moral obligation<sup>2</sup> is different from your moral obligation<sup>1</sup>. What you really, absolutely, unconditionally ought to do is what you ought<sup>1</sup> to do; when you can't figure out what that is, and you seek a back-up recommendation pointing you toward something you will be able to do with a relatively clear conscience, then you are trying to discover what you ought<sup>2</sup> to do. What you ought<sup>2</sup> to do may be different from what you ought<sup>1</sup> to do, but it will be about the best you can do given your unfortunate ignorance.

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 toward actions based on how things seem to us rather than based on how things are

objectively. (Informally, we may think of these as actions that would be obligatory if the world were objectively just as it seems subjectively to us.)

Suppose a utilitarian thinks – mistakenly, as it happens – that he has certain alternatives; suppose it seems to him that one of them would have the best outcome; but suppose that action is in fact one that the agent cannot perform. I want to say that from the subjective perspective of the agent, that alternative is obligatory<sup>2</sup>. Until he realizes that he can't do it, that's the one that he ought<sup>2</sup> to aim for. As soon as he realizes that he can't do it, something else will become his obligation<sup>2</sup>.

So while we cannot endorse the full-blooded “ought implies can” principle for practical level obligation, we can endorse a somewhat weaker 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.

The condition, then is this: a satisfactory practical level principle should direct the agent to perform an action only if it seems to the agent that he is at that time able to perform that action.

e . An adequate practical level principle must provide a way for the agent to avoid at least certain sorts of blame.<sup>12</sup> Recall that our main motivation for seeking a practical level principle is that we sometimes cannot figure out in a helpful way what is required by our theoretical level principle. In these cases, if we are morally conscientious, we don't want simply to give up on morality; we still want to do something that will be morally acceptable, given our irremediable ignorance.

There is a connection to blameworthiness here. In some cases, a person is blameworthy for having done a certain act largely because he really could have done better; out of

---

<sup>12</sup> In her discussion of what she calls “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” (2010: 73).

laziness, or selfishness, or lack of concern with morality, he just took the easier 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.

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 fulfill 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 do the act that was really obligatory<sup>1</sup>; 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 he took to be obligatory<sup>2</sup>. Where this sort of response is appropriately in place, blame of the sort envisioned is evaded.

So our 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.

#### ***4. 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 is aware of the fact that he does not know the things he would need to know in order to apply the theory to his present predicament. He 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.

Imagine their discussion:

*Perplexed Agent:* I believe that I should<sup>1</sup> do the best I can, but because of ignorance concerning the details of my alternatives and their values, I don't know what I should<sup>1</sup> do in my current specific situation. I'd like to do the right thing – I want to be a morally decent person -- but I'm perplexed. Can you help? Can you tell me what I should<sup>2</sup> do?

*Utilitarian Moral Guide:* Maybe I can help, but first you are going to have to tell me more about your situation. I will need to know what you take your alternatives to be; and, insofar as is possible, I will need to know something about what you take to be the values of these alternatives. In addition, if you are uncertain about the values of your alternatives, I will need to know something about your attitude toward risk in this situation. If you tell me all this, maybe I can help.

*Perplexed Agent:* Good. What do you need to know?

*Utilitarian Moral Guide:* What, as you see it, are the options among which you are choosing? When you tell me about these alternatives, please be sure to describe each of them in a way that will be helpful to you if in the end you choose to do it. That is, in each case describe the action in terms such that, if you decide to do it, you will have no epistemic trouble about implementing your decision. And, insofar as you have any views about it, I need to know what you take to be the values of these options.

*Perplexed Agent* (who in this case turns out to be Dr. Jill): OK. Here's the problem. I have a lovely patient, John. He has a skin condition. I have three pills, A, B, and C. I am certain that A will give him a pretty good but less than ideal cure. I believe 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. My problem is that I don't know which is the cure and which is the killer.

*Utilitarian Moral guide:* I see. I think I will be able to help. I need one further bit of information: what is your view about the morality of putting this patient at serious risk of death?

*Perplexed Agent:* I think that would be wrong if it can be avoided. I don't think it would be right to put John at serious risk of death unless it is absolutely necessary to save his life.

*Utilitarian Moral Guide:* OK. Then I have a suggestion: among the things that you take to be your alternatives, throw out all the ones that seem to you to involve exposing John to serious 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.)

*Perplexed Agent:* Great! Then I will forget about B and C. They are just too risky. That leaves only A, so I will prescribe A. Thank you so very much for this help. You are a godsend.

*Utilitarian Moral Guide:* No problem. Glad to help. In fact, if you think about it, you will see that you could have figured this out for yourself. You didn't need me. I just helped you to pull together some things that you already knew and to reach a conclusion that was available to you from the start.

So in this example, the Utilitarian Moral Guide concludes that Dr. Jill ought<sup>2</sup> to give Pill A.

Note that the act selected in this example 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 B or 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.<sup>13</sup> Telling 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.<sup>14</sup> 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 should give and it gives her this recommendation in terminology that will make it easy for her to figure out what she is supposed 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.<sup>15</sup>

Condition (b): *Non-repugnance*. In this case the Utilitarian Moral Guide recommends that Dr. Jill give Pill A, and there is nothing repugnant about that. Of course, Act Utilitarianism implies that it's wrong to give Pill A; but since Dr. Jill has no way of

---

<sup>13</sup> 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. If the probability of the +100 outcome of giving B is sufficiently high, giving B might maximize expected utility.

<sup>14</sup> I think one of the most blatant violations of the helpfulness condition occurs in a popular case involving Act Utilitarianism. Some assume that Act Utilitarianism gives a fair account of our obligations. They have gone on to suggest that when you don't know what maximizes actual utility, and hence do not know what you ought to do, you ought to do what maximizes expected utility. As I tried to show in my (2006), while it is hard to know what maximizes regular utility, it is even harder to know what maximizes expected utility. Since telling a person that he ought to do whatever maximizes expected utility will often be unhelpful, this sort of answer violates the helpfulness condition.

<sup>15</sup> The Perplexed Agent is not permitted to say that her alternative set is: {a1 – Maximize utility; a1 – Don't maximize utility}. While that might be a legitimate alternative set, the act descriptions here are unacceptable. The agent has to use descriptions that she will later find helpful and action guiding.

knowing what pill she ought<sup>1</sup> to give, and Pill A is safe and fairly effective, Dr. Jill, as 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 requires<sup>1</sup> 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 requires<sup>2</sup> 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 seems to her to be non-repugnant in this case.

Condition (c): *Morality*. The guidance given by the Utilitarian Moral Guide in this case does indeed constitute *moral* guidance. He 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*. In this case we have no problem. 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 ought<sup>2</sup> to do something that she can't do.<sup>16</sup>

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

---

<sup>16</sup> There is much more to be said about this condition. In some cases the Guide will end up recommending a course of action that the agent will not be able to pursue. I discuss this in detail in a longer version of this paper.



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.<sup>17</sup>

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 obligation<sup>2</sup> 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.

#### *4.1 A Variation on the Example*

Now let us turn see what happens when the agent is in a slightly different subjective situation. Imagine that the dialogue starts out just as it did in the previous example. 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:* I see. 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. Still, I may be able to help. 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:* After discussing this possibility with John and thinking more carefully about the situation, I have concluded that a partial cure is morally unacceptable. I think I have to go for broke. I am not afraid of putting him at risk, if that's required in order to have a shot at a perfect cure. I do not think it would be right to do something that will end up with my patient only partially cured when there is a chance of getting a perfect cure.

---

<sup>17</sup> In a longer version of this paper I do discuss the blameworthiness criterion.

*Utilitarian Moral Guide:* OK. I won't comment on your policy concerning this sort of case. We can discuss that on another occasion. I will simply tell you what, given that you have those moral views, you ought<sup>2</sup> to do. And that is fairly clear: 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 (obviously) 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.

*Perplexed Agent:* Thank you so very much for this help. You are a godsend.

*Utilitarian Moral Guide:* No problem. Glad to help. In fact, if you think about it, you will see that you could have figured this out for yourself. You didn't need me. I just helped you to pull together some things that you already knew and to reach a conclusion that was available to you from the start.

So in this revised case, the outcome is that Dr. Jill ought<sup>2</sup> 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 requires<sup>1</sup> 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 have a chance at achieving 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 permissible<sup>2</sup> 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 permissible<sup>2</sup> for

her to act on them. Perhaps it was not permissible<sup>1</sup> 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.

### *5. A Two-Level Moral Theory*

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

In every case, the UMG'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 UMG would recommend that he 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 he thinks that morality requires him to avoid putting people at such serious 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 he perform one of the ones he 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 his judgment in his current 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 he 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 ought<sup>1</sup> to perform an act iff it maximizes utility.

Level 2: If you cannot determine what you morally ought<sup>1</sup> to do, then you morally ought<sup>2</sup> to perform the act that the utilitarian moral guide recommends.

Of course there really isn't any UMG and so no actual agent is able to consult with such a person. Thus it would be impossible to implement the Level 2 component of this theory. Furthermore, since the concept of the UMG would be unfamiliar to many morally conscientious people, the Level 2 principle here might not be helpful. Fortunately, it is possible to formulate the theory without engaging in fantasy. Talk of the UMG was mere heuristic. In order to state the actual view, we must first describe a decision procedure that the perplexed agent can follow on her own, without the help of a UMG. Here is such a decision procedure:

Step One: consider the acts that you take to be your alternatives – described in “helpful”, “action-guiding” terms;

Step Two: consider, insofar as your epistemic state permits, what you take to be their values or perhaps just their relative values;

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 should<sup>2</sup> be done. The procedure constitutes moral guidance rather than etiquetteal or legal or prudential guidance; even if the resulting guidance would not be equivalent to the implications of AU, 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, she 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 my Two-Level Theory:

Level 1: You morally ought<sup>1</sup> to perform an act iff it maximizes utility.

Level 2: If you cannot determine what you morally ought<sup>1</sup> to do, then you morally ought<sup>2</sup> to perform an act iff it is an outcome of the Utilitarian Decision Procedure.

## ***6. Unfinished Business***

I have now stated my two-level normative theory. I have explained its application to two versions of an important case. I have claimed (briefly, I admit) that it's OK. But I am well aware of the fact that many questions remain unanswered. By way of conclusion, let

me briefly indicate some of these questions, just to let you know that I am not totally oblivious.

i. It would be good to have a properly developed system of deontic logic for the various fall-back concepts of obligation, permission, and prohibition. It would also be good to have an account of the ways in which the logic of the fall-back concepts interacts with the logic of the familiar first-order deontic concepts.

ii. it would be good to have a clear account of the way in which the proposed concept of fall-back obligation relates to the more familiar concept of conditional obligation.

iii. In this paper I have defended a view about a principle of subjective obligation that is intended to be linked with a utilitarian principle of objective obligation. We might wonder how it would work if the principle of objective obligation were a Rossian principle about the maximization of prima facie rightness; or if it were a Virtue Ethical principle about the maximization of the manifestation of virtue.

iv. I have explained a view about the subjective obligations of people who believe in AU. I have hinted about the subjective obligations of people who believe in Rossianism or Virtue Ethics. But what about people who do not believe in any principle about objective obligation? What are they supposed<sup>2</sup> to do?

v. How does my proposed two-level theory differ from the two-level theories of Hare, Smith, Mason, Brink and others who have written on this topic?

vi. What do I want to say about people who have accepted really terrible principles of objective obligation? Are they required<sup>2</sup> to get as close as possible to the terrible behavior required by their terrible theories? – and if so, is that a problem for this theory?

vii. In a longer version of this paper I offer replies to a barrage of questions that Shelly Kagan asked in a paper presented at the New England Consequentialism workshop.

## *Bibliography*

- Bales, Eugene. 1971. "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 8: 257-265.
- Brink, David O. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Doviak, Daniel. 2009. *Being Good, Doing Right, Faring Well*. Ph. D. Dissertation. University of Massachusetts at Amherst.
- . Forthcoming. "A New Form of Agent-Based Virtue Ethics." *Ethical Theory and Moral Practice*.
- Ewing, A. C. 1947. *The Definition of Good*. London: Routledge and Kegan Paul.
- Feldman, Fred. 2006. "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility." *Philosophical Studies* 129: 49-79.
- Frazier, Robert L. 1994. "Act Utilitarianism and Decision Procedures." *Utilitas* 6: 43-53.
- Hare, R. M. ????. *Moral Thinking: Its Levels, Method, and Point* Oxford: The Clarendon Press of Oxford University Press.
- Howard-Snyder, Frances. 1997. "The Rejection of Objective Consequentialism." *Utilitas* 9: 241-248.
- . 2005. "It's the Thought that Counts." *Utilitas* 17: 265-281.
- Hudson, James L. 1989. "Subjectivization in Ethics." *American Philosophical Quarterly* 26: 221-229.
- Jackson, Frank. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101: 461-482.
- Kagan, Shelly. 1998. *Normative Ethics*. Boulder, CO: Westview Press.
- \_\_\_\_\_. 2010. "The Paradox of Methods" draft of 11/1/10, presented at the New England Consequentialism Workshop on Wednesday, November 17, 2010, between 4:30-6:00 p.m., at the Edmond J. Safra Center for Ethics (124 Mt. Auburn Street, 5th floor, Cambridge MA 02138).

- Lockhart, Ted. 2000. *Moral Uncertainty and its Consequences*. Oxford: Oxford University Press.
- Oddie, Graham and Peter Menzies. 1992. "An Objectivist's Guide to Subjective Value." *Ethics* 102: 512-533.
- Mason, Elinor. 2003. "Consequentialism and the 'Ought Implies Can' Principle." *American Philosophical Quarterly* 40: 319-331.
- Moore, G. E. [1903] 1993. *Principia Ethica*. Revised Ed. Ed. Thomas Baldwin. Cambridge: Cambridge University Press.
- , 1912. *Ethics*. Oxford: Oxford University Press.
- Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13: 134-171.
- Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.
- , 1939. *The Foundations of Ethics*. Oxford: Clarendon Press.
- Sepielli, Andrew. 2009. "What to Do When You Don't Know What to Do." In *Oxford Studies in Metaethics: Volume 4*. Ed. Russ Shafer-Landau. Oxford: Oxford University Press, 5-28.
- Smart, J. J. C. 1973. "An Outline of a System of Utilitarian Ethics." In J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, 3-74.
- Smith, Holly. 1983. "Culpable Ignorance." *Philosophical Review* 92: 543-571.
- , 1991. "Varieties of Moral Worth and Moral Credit." *Ethics* 101: 279-303.
- , 2010. "Subjective Rightness." *Social Philosophy and Policy* 27: 64-110.
- Timmons, Mark. 2002. *Moral Theory: An Introduction*. Lanham, Maryland: Rowman and Littlefield.
- Widerker, David. 1991. "Frankfurt on 'Ought Implies Can' and Alternative Possibilities." *Analysis* 4: 222-224.
- Zimmerman, Michael J. 1997. "A Plea for Accuses." *American Philosophical Quarterly* 34: 229-243.
- , 2004. "Another Plea for Excuses." *American Philosophical Quarterly* 41: 259-266.



-----, 2006. "Is Moral Obligation Objective or Subjective?" *Utilitas* 18: 329-361.

-----, 2008. *Living with Uncertainty*. Cambridge: Cambridge University Press.