Rejecting Ethical Deflationism*

Jacob Ross

One of the perennial challenges of ethical theory has been to provide an answer to a number of views that appear to undermine the importance of ethical questions. We may refer to such views collectively as "deflationary ethical theories." These include theories, such as nihilism, according to which no action is better than any other, as well as relativistic theories according to which no ethical theory is better than any other. In this article I present a new response to such deflationary ethical views. Drawing a distinction between acceptance and rejection, on the one hand, and belief and disbelief, on the other, I argue that we have strong reason to reject these theories, even if we do not have reason to disbelieve them.

In Section I, I clarify the question of what ethical theory we should accept, and I argue for the central importance of this question. In Section II, I discuss what I call "absolutely deflationary" ethical theories. These are theories according to which it matters *not at all* what we do or *not at all* what ethical theory we accept. I argue that it is generally rational to reject any theory of this kind. In Section III, I discuss what I call "relatively deflationary" ethical theories. These are theories according to which it matters *little* what we do or what ethical theory we accept. I argue that we have strong *pro tanto* reason to reject theories of this kind. And then, in Sections IV and V, I reply to some common objections to my arguments. Throughout, I will be arguing not that deflationary ethical theories are false but only that we should reject them from the practical point of view as a basis for guiding our actions.

* I would like to thank Frank Artzenius, Paul Bartha, John Broome, Ruth Chang, Stephen Darwall, Alan Gibbard, James Joyce, Jeff McMahan, Derek Parfit, Peter Railton, Andrew Sepielli, and Peter Vallentyne for very helpful comments on earlier drafts of this article. For generous help in formalizing the proof in the appendix, I am grateful to James Joyce and John Kolassa, and especially to Robert Berk, who showed me how a much longer version of the proof could be simplified to its present form. But my greatest debt of all is to Larry Temkin for invaluable comments on many earlier drafts of the article and for countless illuminating discussions.

Ethics 116 (July 2006): 742–768
 0014-1704/2006/11604-0004$10.00

I. THE PROBLEM OF THEORY ACCEPTANCE

Before considering the question of what ethical theories are worthy of acceptance and what ethical theories should be rejected, I must first define what I mean by 'acceptance' and by an 'ethical theory'. The term 'acceptance' is used in many ways by philosophers, but I will use it to refer to an attitude taken toward a theory or proposition in the course of practical reasoning or in the guidance of action. By "to accept a theory," in relation to a given decision problem, I mean "to guide one's decision on the basis of this theory." More precisely, to accept a theory is to aim to choose whatever option this theory would recommend, or in other words, to aim to choose the option that one would regard as best on the assumption that this theory is true. For example, to accept utilitarianism is to aim to act in such a way as to produce as much total welfare as possible, to accept Kantianism is to aim to act only on maxims that one could will as universal laws, and to accept the Mosaic Code is to aim to perform only actions that conform to its Ten Commandments.

In general, one will choose the option that a theory recommends if one reasons on the supposition that this theory is true. For example, if, in the course of making a given decision, one deliberates on the supposition that utilitarianism is the correct ethical theory, then one will choose an option that maximizes expected total welfare, whereas if one supposes that the Mosaic Code is the correct ethical theory, then one will choose an action the conforms to its Ten Commandments. Thus, we may say that one accepts a theory if one deliberates on the supposition that this theory is true or if one adopts it as a premise in practical reasoning. Similarly, we may say that one rejects a theory if one deliberates on the assumption that it is false or adopts its negation as a premise in practical reasoning.

It is possible to believe a proposition or theory to a greater or lesser degree, and degrees of belief or credence may be represented on a scale from zero to one, inclusive. And the same is true of acceptance. For while some have insisted that acceptance is an all-or-nothing phenomenon, if one understands accepting a proposition or theory as employing it as a premise in practical reasoning, then acceptance, like belief, can admit of degrees. For in practical reasoning we can treat a proposition or theory as having any probability between zero and one, inclusive. We may refer to this quantity as one's "degree of acceptance" in the proposition or theory in question. Thus, in solving a given practical problem, it is possible to take a number of alternative theories into consideration by assigning a nonzero probability to each, or in other words, by partially accepting each. I might, for example, aim to choose the option that would be optimal if there was a .5 chance that utilitarianism is true and a .5 chance that Kantianism is true. If I did this, I

would have a degree of acceptance of .5 in each of these theories. For any such set of alternative theories, we may refer to one's respective degrees of acceptance in these theories as an "acceptance distribution."

In what follows, I will be concerned primarily with the acceptance and rejection of ethical or evaluative theories, that is, of theories that can be employed in determining which of one's options one has most reason to choose or, more generally, in determining what relations of comparative choiceworthiness obtain among one's options. In the first instance, I will be concerned with theories about what one has most reason to choose, all things considered, rather than with theories that evaluate options in terms of one particular kind of consideration, such as moral or prudential considerations. However, everything I will say applies equally to the acceptance and rejection of moral theories in contexts in which only moral considerations are relevant. For in such contexts, the action that is morally best is the action that one has most reason to choose, all things considered. And similarly, what I will say applies equally to the acceptance of prudential theories in contexts in which only prudential considerations are relevant.

Further, the notion of acceptance that I have defined, as well as the principles of acceptance and rejection that I will later propose, all apply as much to descriptive theories, or to theories of the world, as to evaluative theories. For descriptive theories, like evaluative theories, can play a role in guiding our decisions, and so we can accept theories of either kind as a basis for practical guidance. For example, in the course of making various choices, such as choosing among designs for a bridge or choosing among trajectories for a lunar voyage, one might accept Newton's theory of gravity. If one does so, and if one is otherwise well-informed, then one will choose the option that would be optimal if Newton's theory is true. And it might be rational to proceed in this manner even if one believes that Newton's theory of gravity is false. This may be rational so long as one expects that the option that would be optimal according to Newton's theory would be nearly optimal according to the theory of gravity that one regards as true, say, the general theory of relativity.

In this case, the reason it is rational to accept Newton's theory of gravity, rather than Einstein's, is that the former is a much simpler theory and would thus be much easier to use as a basis for assessing one's options. Thus, while one might choose a slightly better option on the basis of Einstein's theory, this benefit would be outweighed by the far greater cognitive costs of employing this theory in reasoning. But there is another class of cases in which it can be rational for one to accept a theory that one does not believe or, more generally, for one's degrees of acceptance to differ from one's degrees of belief. These are cases in which one's credence is divided among a large number of theories, so

that attempting to reason probabilistically on the basis of all these theories would be inordinately difficult. It is this sort of case that I will be concerned with in what follows.

Cases of this kind abound. In deciding whether to send the trolley to the left or to the right, there may well be a large number of descriptive factors about which I am uncertain and that I regard as relevant to my decision, such as the number of people on each track, the probability that each person could escape an oncoming trolley, the effect of each person's death on her loved ones, and so on. Likewise, in deciding whether to spend the afternoon at the beach or at the gym, I might regard a number of factors as relevant, including the weather, the number of people at each location, the availability of parking at each location, and so on, and I might be uncertain about each of these factors. If, in making these decisions, I reasoned probabilistically, taking into account all of the possibilities for each of these factors, then it is likely that both the gym and the beach would already be closed, or that the people I might save would already have been run over by the trolley, long before I could reach a decision.

Uncertainty concerning what ethical theory is correct leads to similar difficulties. For in general, the more we reflect on questions of ethical theory, the greater is the number of ethical theories among which our credence is divided. What initially appears to be a single ethical theory often turns out to be specifiable in a number of ways, each of which has some plausibility. And when a problem arises for an initial formulation of a theory, it is often possible to solve this problem by modifying the theory in any of several ways, revealing once more a multiplicity of theories, each having some degree of plausibility. It would be impractical, however, to try to take each of these theories into consideration in making each decision. Suppose, for example, that when faced with the options of going to the gym and going to the beach, or when faced with the options of sending a trolley to the left or sending it to the right, I attempted to carry out a complex probabilistic calculation, taking into account the values assigned to my options by each and every version of egoism, consequentialism, Kantianism, virtue ethics, and so on, that I find even remotely plausible. If I did this, then once again it is unlikely that I would be able to make either decision in time.

Thus, our credence is often divided among a large number of descriptive or evaluative theories. And in such cases, because the decision problem we are actually faced with is inordinately complicated, we must employ some sort of heuristic in order to render our problem tractable. There are many conceivable heuristics that we might employ, but what we often do in practice is to ignore, or exclude from consideration, many of the alternative theories or hypotheses that we regard as having some degree of plausibility. Thus, in deciding between the

beach and the gym, I might take into account only the possibilities of sun and of rain and exclude from consideration the possibility of an unanticipated solar eclipse, even though I have a nonzero degree of credence in this possibility. Similarly, in deciding between sending the trolley to the left and sending it to the right, I might take into account only the ethical theories of Kant and of Mill and exclude from consideration Schopenhauer's ethical theory, even though I am not certain that the latter is false.

Very often, we bypass probabilistic reasoning altogether by reasoning on the basis of a single theory or set of assumptions, even though there are alternative theories or sets of assumptions in which we have credence. For example, in deciding whether to go to the beach or to the gym, rather than assigning probabilities to each possibility for each relevant factor, I may simply reason on the basis of certain assumptions, for example, that it will be sunny, that the beach will be more crowded than the gym, and that parking will be scarce in either location. Similarly, in deciding where to send the trolley, rather than assigning probabilities to various alternative sets of normative assumptions that I find plausible, I may simply suppose one such set of assumptions. That is, I may suppose that a certain set of considerations are relevant—say, the number and the age of the people on either track—and that these considerations are to be weighted in a certain manner, even though I acknowledge the plausibility of many alternative conceptions of what considerations are relevant and of how they are to be weighted.

Thus, so long as the idea of a 'theory' is construed sufficiently broadly to encompass any of the alternative sets of assumptions, or conceptions of the relevant factors, that we might adopt in practical reasoning, what I have described as the heuristic of theory rejection is one that we employ all the time in practical reasoning. For in the course of such reasoning, we almost always reject, or exclude from consideration, many of the alternative descriptive or evaluative conceptions in which we have a nonzero degree of credence. However, we seldom reflect on the question of which among these alternative conceptions it is rational to reject. One reason may be that we assume that the answer to this question is obvious. For it may seem obvious that the theories we should reject are those that we regard as least likely to be true. And it may seem equally obvious that if we are to accept a single theory as a basis for guiding our actions, we should accept whichever of the contending theories we regard as most probable.

These assumptions often appear to be taken for granted in ethical inquiry. Since the time of Aristotle, many philosophers have insisted that the ultimate aim of ethical inquiry is practical.[1] We want to do the

1. See Aristotle, *Nicomachean Ethics*, 1103b27.

right thing, or to do good, and so we assess alternative ethical theories not simply in order to satisfy a theoretical interest but in order to find a theory by which to guide our lives. In addition, it is often assumed that the way to assess ethical theories is to consider the epistemic reasons for and against them and thus to determine which one is most likely to be true. Thus, it is assumed that if we figure out which ethical theory is most likely to be true, we will thereby ascertain by which ethical theory we should guide our actions.

These assumptions, however, do not stand up to scrutiny. Reflecting on the case we considered earlier involving Newton's and Einstein's theories of gravity, it is clear that, at least in relation to some kinds of theories, we must qualify the view that we should always accept whatever theory we regard as most probable. In the sections that follow, I will argue that this view must be not merely qualified in special cases but must be entirely abandoned. I will show that if the question we are asking is what theory we should accept as a guide to action, there are certain factors that have hitherto been overlooked but that are no less relevant to this question than the probabilities of the contending theories. Thus, I will argue that we must change the way we approach the problem of theory acceptance, both in ethics and in other domains.

II. REJECTING ABSOLUTELY DEFLATIONARY THEORIES

In this section, I will argue that it is rational to reject absolutely deflationary theories. I will begin by arguing that it is rational to reject theories that imply that it doesn't matter what one *does*, and I will then argue that it is rational to reject theories that imply that it doesn't matter what ethical theory one *accepts*.

Let us begin by considering theories according to which it doesn't matter what one does in a given choice situation, that is, according to which none of one's options is better or more choiceworthy than any other. We may call such theories "nondiscriminating." There are several ways in which a theory can be nondiscriminating. First, it might imply that each of one's options is equally good. We may say that such a theory is uniform in relation to the set of options under consideration. Such uniform theories allow for the simplest illustration of why absolutely deflationary theories should be rejected.

Suppose I think that a uniform theory is probably true, but I'm not absolutely certain, because I think there is a small chance that some nondeflationary theory may be true. We may suppose that I have a degree of credence of .99 in a uniform theory, T_U, and a degree of credence of .01 in some nondeflationary theory, T_L. Now suppose I'm trying to decide whether to send a trolley to the right, which would result in ten deaths, or to the left, which would result in five deaths. Theory T_L, let us suppose, implies that I should send the trolley to the

left, while T_U, being uniform, implies that either option would be equally good. I could therefore reason as follows.

Given one possibility (that T_L is true) it would be better to send the trolley to the left, and given the alternative possibility (that T_U is true) both options would be equally good. Thus, the option of sending the trolley to the left dominates the option of sending it to the right, and so the former option has a higher expected value. Therefore, I should send the trolley to the left.

And this is the same conclusion I would arrive at if I deliberated on the basis of T_L alone. Whenever I have credence in two theories, one uniform and one nonuniform, the conclusion I would come to on the basis of the latter theory alone will be the same as the conclusion I would come to by taking into account both theories and reasoning probabilistically. Therefore, no harm is done by excluding the uniform theory from consideration and accepting the nonuniform theory.

Another kind of nondiscriminating theory is ethical nihilism. This is the view that the notions of good and bad and of right and wrong are illusions and that, objectively speaking, no option or state of affairs is better than any other, nor are any two options or states of affairs equally good. Thus, while uniform theories assign the same value to all of our options, nihilistic theories don't assign values to any of our options.[2]

Now suppose once again that I have a degree of credence of .01 in T_L from the previous example, but this time I have a degree of credence of .99 in a nihilistic theory, T_N. And again suppose that I must decide between sending the trolley to the right and sending it to the left. In this case we could reason as follows.

According to T_L, it would be better for me to send the trolley to the left than to send it to the right. And so my credence in T_L gives me *pro tanto* subjective reason to send the trolley to the left. The only way this could fail to be the most rational option would be if my credence in T_N gave me a sufficiently strong countervailing subjective reason to send the trolley to the right. But T_N implies that there would be nothing valuable or disvaluable about either alternative. And so my credence in T_N gives me no subjective reason to favor either alternative. Hence the *pro tanto* subjective reason to send the trolley to the left is unopposed, and so this is the rational option.

2. One might object that this is a metaethical theory rather than an ethical theory. I need not take a stand on this question of terminology, since, for my purposes, the essential distinction is between theories that can consistently be held in conjunction and theories that constitute genuine alternatives. Since the kind of nihilism I am considering implies that no option is better or more choiceworthy than any other, it is clearly incompatible with, and hence constitutes a genuine alternative to, many more customary ethical theories.

I would choose this same option if I were to deliberate on the basis of T_L alone. Thus, taking into account T_N has no effect on what I ultimately choose, and so T_N can harmlessly be excluded from consideration in deciding how to act. That is, I can harmlessly reject the nihilistic theory and deliberate on the supposition that the nondeflationary theory, T_L, is true.[3]

So far I have argued that we may harmlessly reject nondiscriminating theories in simple cases in which our credence is divided between one discriminating and one nondiscriminating theory. But the same is true in the more general case in which our credence is divided among any number of discriminating theories and any number of nondiscriminating theories. In any such case, none of the nondiscriminating theories will do any work in determining which option we should choose. And so they can all be rejected, or excluded from consideration, without affecting the conclusion of our deliberation. And this is an important result, since it is generally only in cases in which our credence is divided among a large number of theories that the heuristic of theory rejection is necessary.

By a similar argument, we can show that it is always rational to reject certain kinds of ethical relativism. To show this, we must first draw a distinction between what we may call "acceptance-dependent" and "acceptance-independent" ethical theories. An acceptance-dependent theory is one according to which it matters not just what you do but also why you do it. More precisely, such a theory implies that an action can be better or worse depending on what theory or theories it is chosen on the basis of. Thus, an acceptance-dependent theory might imply that keeping one's promise on the basis of accepting the Categorical Imperative would be intrinsically more valuable than keeping one's promise on the basis of accepting ethical egoism. Since Kantianism has this very implication, Kantianism is an acceptance-dependent theory. Classical utilitarianism, however, implies that what matters are an action's causal consequences and not the theory or theories on the basis of which an action is chosen, and so this theory is acceptance-independent.

Until now I have been assuming that all the theories in which one has credence are acceptance-independent. But if we abandon this assumption, then we can consider relativistic theories. In this section, I will focus on a very strong kind of relativism, which we may call "absolute relativism." This is the view that one should do whatever action is rec-

3. As Peter Railton has pointed out in discussion, the above argument does not address a more radical form of nihilism that rejects the very idea of subjective practical rationality. This argument is concerned with a kind of nihilism that denies that there are any objective reasons for action but that concedes that if one held that there were such reasons, then it would be subjectively rational to act in accordance with them.

ommended by the ethical theory one accepts. More precisely, a theory is absolutely relativistic if and only if it implies the following two principles:

R1. Given that one accepts an ethical theory, T_A, the best thing one can do is to perform any one of the actions ranked highest by T_A.

R2. For any two ethical theories, T_A and T_B, accepting T_A and performing an action recommended by T_A would be no worse than accepting T_B and performing an action recommended by T_B.

Thus, an absolutely relativistic theory will imply that the Kantian should follow Kantianism and that the utilitarian should follow utilitarianism. Such a theory would also imply that to accept and follow Kantianism would be no worse than to accept and follow utilitarianism, nor would the latter be worse than the former.[4]

Let T_R be an absolutely relativistic theory. Thus, T_R will imply R1 and R2. But it must say more than this. For if T_R consisted simply in these two propositions, then although it would indicate what one should do if one accepts Kantianism or utilitarianism or the like, it would not indicate what one should do if one accepts T_R. And so T_R would be an incomplete ethical theory, and worse, it could never be action guiding, since it would never prescribe any action for an agent who accepts it. In order for T_R to be a complete and action-guiding ethical theory, it must indicate what one should do if one accepts T_R. There are an infinite number of ways in which T_R might be completed, and so T_R is but one of an infinite number of alternative absolutely relativistic theories. But for the sake of concreteness we may stipulate that T_R is completed with the following:

R3. Given that one accepts T_R, the best thing that one can do is any one of the actions recommended by whatever ethical theory is most commonly accepted in one's society.

Now suppose your credence is divided between the absolutely relativistic theory, T_R, and the nonrelativistic, discriminating theory, T_X. In this case, it would be rational for you to reject T_R and accept T_X. For, suppose T_X is true. In this case, there is a subset of top-ranking options, which does not depend on what theory you accept. And if you accept T_X, then we may assume that you will choose one of these top-ranking

4. What I am calling "absolute relativism" might also be called "absolute internalism," since it implies that what one has most reason to do is determined by one's subjective motivational set and more specifically by the ethical or evaluative theory that one accepts. Note, further, that what I am calling "absolute relativism" is incompatible with, and is hence an alternative to, many ordinary ethical theories, such as utilitarianism, according to which how one should act does not depend on what ethical theory one accepts. See n. 2, above.

options. But if you accept T_R, then you might or might not choose one of these top-ranking options: you will choose one of these options only if the ethical theory most commonly accepted in your society happens to agree with T_X. Thus, on the assumption that T_X is true, by accepting T_R you might do as well as you would by accepting T_X, or you might do worse, but you will not do any better. Suppose, however, that T_R is true. Since T_R is absolutely relativistic, it implies R2, and so it implies that accepting T_X and performing an action recommended by T_X would be no worse than accepting T_R and performing an action recommended by T_R. Thus, assuming that you will perform whatever action is recommended by the ethical theory you accept, T_R implies that you would do no worse on the basis of T_X than on the basis of T_R. Since accepting T_X might be better if T_X is true, and would be no worse regardless of what theory is true, you should accept T_X.

More generally, whenever one's credence is divided among any number of absolutely relativistic theories and any number of acceptance-independent theories, it is rational to reject all of the absolutely relativistic theories. To make this argument, we must first supplement our definition of an absolutely relativistic theory, so as to indicate what such a theory implies concerning situations in which an agent does not fully accept any one ethical theory but instead partially accepts a number of alternative ethical theories, or in other words, has an acceptance distribution in which positive probability is assigned to several such theories. Absolutely relativistic theories imply that what one should do is determined by what one accepts, and so given that one fully accepts a single theory, the action one has most objective reason to choose is whatever action one would choose if one were to deliberate rationally on the basis of this theory, namely, the option to which this theory assigns the highest value. Hence, absolutely relativistic theories imply R1. On a natural extension of this view, the option one has most objective reason to choose is the option one would choose if one were to deliberate rationally on the basis of one's *acceptance distribution*. And hence,

R1′. given that one has acceptance distribution D, the best thing one can do is to perform any one of the actions with the highest expected value relative to D.

Further, absolutely relativistic theories imply R2, and hence they imply that no ethical theory is a better guide to action than any other, in the sense that accepting any one ethical theory and doing what it recommends is no better than accepting any other such theory and doing what it recommends. On a natural extension of this view, no *acceptance distribution* is better than any other, in the sense that

R2′. for any two acceptance distributions, D_A and D_B, having D_A

and performing an action that is optimal relative to D_A would be no worse than having D_B and performing an action that is optimal relative to D_B.

We may thus define an absolutely relativistic theory as one that implies R1′ and R2′. Since R1′ and R2′ imply R1 and R2, it follows that any absolutely relativistic theory will imply R1 and R2 and so will satisfy the earlier definition.

Now suppose your credence is divided among a set of absolutely relativistic theories and a set of acceptance-independent theories. If any one of the absolutely relativistic theories is true, then no acceptance distribution is better than any other. Suppose, however, that the true theory is not absolutely relativistic. The option with the highest expected value, conditional on this assumption, is precisely the option one would choose if one were to reject, or exclude from consideration, all the absolutely relativistic theories in which one has credence. Thus, conditional on the assumption that the true theory is not absolutely relativistic, rejecting all the absolutely relativistic theories would be an optimal strategy. Thus, both on the assumption that the true theory is absolutely relativistic and on the contrary assumption that it isn't, we should expect the strategy of rejecting all the absolutely relativistic theories to be no worse than the strategy of taking into consideration all the theories in which one has credence. Therefore, it is rational to reject all the absolutely relativistic theories.

III. REJECTING RELATIVELY DEFLATIONARY THEORIES

In the previous section, I argued that in a broad range of cases one can harmlessly exclude from consideration the absolutely deflationary theories in which one has credence.[5] However, even after rejecting all such theories, it is likely that the remaining ethical theories in which one has credence will be too numerous to all be included in deliberation. Hence, it is likely that one will need to exclude from consideration some of the theories in which one has credence that are not absolutely deflationary. And if a theory is not absolutely deflationary, then in arguing for its rejection, we cannot appeal to the sort of dominance argument we employed in the last section. There we argued that we can rationally reject absolutely deflationary theories, since doing so would be worse according to some theories in which we have credence and better under none. But if a theory, T_X, is not absolutely deflationary, then it may well imply that one would be better off accepting it than by rejecting it and

5. There is one exceptional type of case in which it can be irrational to reject an absolutely deflationary theory. This can arise if one has a sufficient degree of credence in an acceptance-dependent theory that implies that an action chosen on the basis of an absolutely deflationary theory would be better than an action chosen otherwise.

accepting some other theory. And if rejecting T_X would be worse according to at least one of the theories in which we have credence, namely, T_X itself, then we cannot show that T_X should be rejected using a dominance argument.

Indeed, in most situations involving significant uncertainty, dominance arguments are insufficient to determine which option we should choose, since in such cases there is seldom any option that is best or equal-best under every condition that we regard as having some probability. If, for example, I must choose between going to the gym and going to the beach, I may be unable to rule out the possibility that there will be a thunderstorm and that it would thus be best for me to go to the gym, and I may be equally unable to rule out the possibility that it will be the most beautiful day of the year and that it would thus be best for me go to the beach. In this case, given my uncertainty concerning the weather, neither option will dominate. Further, even if I have no uncertainty concerning the weather, or concerning any other descriptive state of affairs, it is unlikely that any option will be best or equal-best according to every ethical theory in which I have credence. If, for example, I know that it will be the most beautiful day of the year, then according to an ethical theory on which I should seek personal enjoyment, it would be better for me to go to the beach. But I have some credence, albeit extremely little, in an ascetic ethical theory according to which I should avoid enjoyment, and on such a theory it would be better for me to go to the gym. So once again, neither option will dominate. And where it is not the case that the practical option recommended by one theory dominates the practical option recommended by the other, it will not be the case that the acceptance of the former theory dominates the acceptance of the latter.

A standard approach to choice under uncertainty, in cases in which none of one's options dominate, is to employ decision theory and to choose the option that maximizes expected utility or expected value. This approach is normally employed in cases in which we are uncertain about what outcome would result from each available action, not about the value of each of these outcomes. For example, in a standard decision problem in which I must choose between going to the beach and going to the gym, I might be uncertain as to whether it will be sunny or rainy, in which case I will be uncertain as to whether going to the beach would result in my getting a suntan or in my getting wet. But it is assumed that I have no uncertainty concerning the values of getting wet and of getting a suntan. Assuming I regard sunny weather as three times as probable as rainy weather, my decision problem might be represented as follows:

	Sun (probability = .75)	*Rain* (probability = .25)
Gym	Get exercise (value = 8)	Get exercise (value = 8)
Beach	Get suntan (value = 12)	Get wet (value = 0)

In this case, the expected value of going to the gym is $.75 \times 8 + .25 \times 8 = 8$, while the expected value of going to the beach is $.75 \times 12 + .25 \times 0 = 9$, and so in order to maximize expected value I should go to the beach.

Such an approach can also be employed, however, in cases where one has credence in a number of alternative ethical or evaluative theories and where these theories disagree concerning the relative values of one's options. Suppose, for example, that I have a degree of credence of .75 in hedonism and a degree of credence of .25 in asceticism. In this case, my decision problem might look like this:

	Hedonism (probability = .75)	*Asceticism* (probability = .25)
Gym	Get exercise (value = 8)	Get exercise (value = 8)
Beach	Get suntan (value = 12)	Get suntan (value = 0)

Once again, the expected value of going the gym is eight, while that of going to the beach is nine, and so I should go to the beach. This case is exactly analogous to the previous one, except that here we are treating alternative ethical theories in the way in which we had earlier treated alternative states of nature.

In cases of this kind, where one's credence is divided among a plurality of alternative ethical theories, one will only be able to determine the expected values of one's options if each of these theories assigns quantitative values to these options and if these values are commensurable, or can be represented on the same scale. The problem of commensurability will be taken up in the final section, but for the moment there is a more basic problem to address. And this is that many ethical theories, especially deontological theories, do not seem to assign quantitative values at all. For, deontological theories are normally understood not as evaluating options quantitatively but rather as designating certain types of action as right and others as wrong.

Any plausible deontological theory, however, will recognize that the values of actions admit of degrees. It will recognize that among actions that are wrong or impermissible, some are worse than others and that among actions that are right or permissible, some are better than others.[6] Further, on any plausible deontological theory, there will be situations in which, relative to an agent's epistemic situation, it is un-

6. For a more extensive discussion of degrees of moral rightness, and of choice under ethical uncertainty in general, see Ted Lockhart's *Moral Uncertainty and Its Consequences* (Oxford: Oxford University Press, 2000).

certain to which normatively relevant kind a possible action would belong. It may, for example, be unknown to an agent whether by acting in a certain manner she would be killing an innocent person, say, or breaking a promise. Now suppose every action I might perform has a certain probability of violating a prima facie moral requirement: for example, if I act in one way, I run a high risk of failing to fulfill a promise, but if I act in another way, I run a low risk of killing an innocent person. It can be shown that any ethical theory that satisfies certain coherence conditions and that is rich enough to indicate what an agent should do in situations of this kind can be represented in terms of a value function that assigns quantitative values to outcomes.[7]

Or more precisely, any such theory can be represented in terms of a function that assigns quantitative values to outcomes relative to agents and to times. Thus, for example, hedonistic egoism can be represented by a function such that the value it assigns to an outcome relative to an agent, *s*, and a time, *t*, is proportional to the total quantity of pleasure *s* experiences in this outcome over the course of her lifetime. On such a theory, the value of an outcome will vary from agent to agent but not from time to time. And a purely deontological theory can be represented by a function such that the value it assigns to an outcome relative to *s* and *t* is determined entirely by whether, in this outcome, *s* acts in accordance with certain moral requirements at *t*. On such a theory, the value of an outcome will vary both from agent to agent and from time to time. Of course, for many purposes, representing deontological theories in terms of such value functions would be both cumbersome and misleading, but for our purposes it will prove useful.

So long as the various ethical theories in which we have credence can be given an appropriate quantitative representation, it will be possible to employ decision theory in determining what choices would be most rational under ethical uncertainty. And since one kind of choice we can make under ethical uncertainty is the choice of what ethical theory to accept, it follows that we can employ decision theory in making choices of this kind. But before considering such choices, let us consider a more straightforward problem, one in which we don't know what *descriptive* theory is true and we must decide which such theory to accept.

Suppose you are on board a pirate ship, and you are going to be taken to the Mysterious Archipelago. The Mysterious Archipelago consists of twenty-six islands, conveniently labeled A–Z. The pirates will be dropping you off on one of these islands and returning to pick you up one hour later. You must choose to which island you will be taken. In

7. The value function in question will be unique up to any positive linear transformation. This follows from Ramsey's representation theorem. See Richard Jeffrey, *The Logic of Decision*, 2nd ed. (Chicago: University of Chicago Press, 1983), chap. 3.

making this choice, there are two incompatible theories or accounts of the archipelago that are available for your consultation, but you are only allowed to consult one. You know that one of the two accounts is accurate, and you regard them as equally probable. You know that the first, the Cannibal Account, says that half the islands are occupied only by cannibals and the other half are empty, and you know that it indicates which islands are which. You also know that the second, the Blackfly Account, says that half the islands are occupied only by blackflies and the other half are empty, and you know that it indicates which islands are which. Since you must rely on one or the other of these accounts in deciding what island to be taken to, which should you rely on?

It seems clear that you should rely on the Cannibal Account. For, suppose the Cannibal Account is true and hence that half the islands are occupied by cannibals. If you rely on the Cannibal Account, you will choose one of the islands that it indicates to be empty. And since we are assuming that this account is true, you will thereby choose an island that really is empty, and so you will avoid being eaten. If, however, you rely on the Blackfly Account, you will choose one of the islands that it indicates to be empty. But since we are assuming that this account is false, the island you thereby choose might not be empty but might instead be occupied by cannibals, and so you might end up being eaten. Suppose, however, that the Blackfly Account is true. In this case, if you rely on the Blackfly Account, you will choose one of the empty islands, and so you will avoid being bitten by blackflies. But if you rely on the Cannibal Account, and thus choose one of the islands it indicates to be empty, you may end up on an island infested with blackflies, and so you may end up being bitten. However, being eaten by cannibals is much worse than being bitten by blackflies. Thus, on the assumption that the Cannibal Account is true, accepting this account would have a much greater expected value than accepting the Blackfly Account; whereas on the assumption that the Blackfly Account is true, accepting the Blackfly Account would have only a moderately greater expected value than accepting the Cannibal Account. Therefore, so long as you regard the two accounts as equally probable, you should accept the Cannibal Account. Moreover, it would surely be more rational to accept the Cannibal Account even if the Blackfly Account were slightly more probable. Indeed, on reflection, it seems clear that it would be more rational to accept the Cannibal Account unless the Blackfly Account had a *much* greater probability, one close to certainty.

Thinking about the above example, one might initially suspect that where it is rational to accept the less probable of two theories, this is because we are risk averse, and we want to accept a theory that will steer us away from the worst possible outcome. But we can see the falsity of this assumption if we consider a modified version of the case. Suppose

the two available accounts of the Mysterious Archipelago are the Treasure Account and the Coconut Account, and you know that one or the other of these accounts is true. The Treasure Account says that half the islands contain only treasure and that the other half are empty, and it indicates which islands are which, while the Coconut Account says that half the islands contain only coconuts and that the other half are empty, and it likewise indicates which islands are which. Once again, you know that the pirates will be returning to pick you up one hour after dropping you off, and this time you know that they will permit you to keep whatever you find on the island.

In this case, it seems clear that you should follow the Treasure Account unless you regard the Coconut Account as much more probable. And this is not based in a desire to avoid the worst possible outcome, for in this case, this outcome, namely ending up empty-handed, is equally likely whichever theory you accept. To see why you should accept the Treasure Account, it will be useful to discuss the case quantitatively.

Let us assign a value of zero to the outcome of finding nothing, a value of two to the outcome of finding coconuts, and a value of one thousand to the outcome of finding treasure. Since each account implies that half the islands in the archipelago contain goods and the other half are empty, you know that half the islands contain goods regardless of which account is true. But prior to consulting either account, you have no idea what the two accounts say about each individual island, and so initially you regard each island as having a .5 chance of containing goods. Assume, further, that the Treasure Account and the Coconut Account are independent, in the sense that knowing what one account says provides you with no information about what the other account says. Thus, even if you know that one account implies that there are goods on a given island, your degree of credence that the other theory implies the same will remain .5. Let p represent the subjective probability of the Treasure Account. Since you are certain that one of the two accounts is true, it follows that the subjective probability of the Coconut Account is $1 - p$. What we want to find out is how probable the Treasure Account must be in order for it to be more rational to accept it than to accept the Coconut Account. To do this, we must find a formula for the value of accepting each account.

If you accept the Treasure Account and it is true, then you will find treasure, for a value of one thousand. But if you accept the Treasure Account and it is false, then you stand a .5 chance of finding coconuts, for a value of two, and a .5 chance of finding nothing, for a value of zero. Thus, the expected value of accepting the Treasure Account, on the condition that the Coconut Account is true, is one. Since the probabilities of the Treasure Account and of the Coconut Account are p and

$1 - p$, respectively, it follows that the unconditional expected value of accepting the Treasure Account is $p \times 1{,}000 + (1 - p) \times 1$. By analogous reasoning, it can be shown that the unconditional expected value of accepting the Coconut Account is $(1 - p) \times 2 + p \times 500$. And thence it can be shown that the expected value of accepting the Treasure Account will exceed that of accepting the Coconut Account unless p, which is the subjective probability of the Treasure Account, is 1/501 or less, and in this case $1 - p$, which is the subjective probability of the Coconut Account, will be 500/501 or more. In other words it will be more rational to accept the Treasure Account than the Coconut Account unless one regards the latter as at least five hundred times as probable as the former.

Since the value of finding treasure exceeds that of finding nothing by one thousand units of value and since the value of coconuts exceeds the value of finding nothing by two units of value, five hundred is precisely the ratio by which the difference in value between the better and worse outcomes according to the Treasure Account exceeds the difference in value between the better and worse outcomes according to the Coconut Account. A similar argument would show that in the Cannibal/Blackfly case, if we assume that the two alternative accounts of the archipelago are mutually independent and if the difference in value between the better and worse options is ten thousand times as great according to the Cannibal Account as it is according to the Blackfly Account, then it will be more rational to accept the Cannibal Account unless the Blackfly Account is at least ten thousand times as likely.

In general, we have reason to prefer theories that assign much higher values to some options than to others and which thus imply that our options differ significantly in their values. For it is theories of this kind that imply that we would be much better off accepting them than accepting some other theory. Thus, in deciding what theory to accept, although we have *pro tanto* reason to prefer more probable theories to less probable ones, we also have *pro tanto* reason to prefer theories according to which the differences or dispersion among the values of our options is higher to theories according to which this dispersion is lower. We may call this the "dispersion principle."

I have not yet specified how dispersion is to be measured, nor have I indicated how a theory's dispersion combines with its probability in determining the expected value of accepting the theory. In the cases we have considered, each theory implies that there are only two kinds of options, a good kind and a bad kind, and each theory implies that there are equal numbers of options of each kind. In such simple cases, we can measure a theory's dispersion in terms of its range, or the difference between the values it assigns to the better and worse options. But in the more general case, where we are choosing among any set S of options, and where the theories in which one has credence assign

any number of different values to these options, the range will not suffice as a measure of dispersion. It turns out that in this general case, the best measure of a theory's dispersion for purposes of theory selection is the expected difference between the highest value it assigns to any of the options in S and the average value it assigns to these options.[8] We may refer to this interval as the theory's "disparity." Thus, we have reason to accept a theory if we expect that this theory implies that the best option is much better than the average option, and we have reason to reject a theory if we expect that it implies that the best option is not much better than the average option. In the appendix, I demonstrate the general result that where the theories in which one has credence are mutually independent in the manner defined above, the expected value of accepting a given theory varies with the product of its subjective probability and its disparity. Thus, if one's credence is divided between two theories, T_X and T_Y, then it will be more rational to accept T_X than T_Y whenever the product of T_X's probability and disparity is greater than the product of T_Y's probability and disparity.

So far we have been considering cases involving the acceptance of descriptive theories, but the dispersion principle applies equally to the acceptance of ethical or evaluative theories. Consider a simple illustration in which our credence is divided between two ethical theories. The first is a time-neutral theory according to which we should be equally concerned about the welfare of individuals regardless of when they will live. And the second is a time-relative theory according to which we should be concerned only about the welfare of our contemporaries. Suppose we want to accept one of these two theories in the course of making a political decision among a number of alternative proposals for reforming social security. Suppose, finally, that all of these proposals would have very similar consequences in the short run, so that none of our contemporaries will be affected significantly by our choice, but in the long run, the consequences of these different proposals would diverge dramatically, because the different proposals would affect future generations in very different ways.

In this case, the stakes are likely to be much higher according to the time-neutral theory than according to the time-relative theory. For since the various proposals will have very different consequences for future generations, it is likely that the time-neutral theory will imply

8. The reason I refer to the expected difference between these values is that in circumstances in which theory selection is a rational heuristic, the actual difference between these values will generally be unknown. For in general, theory selection is only useful when we have not yet determined the values assigned to each of our options by the various theories in which we have credence. And in such cases, we are unlikely to know the precise difference between the maximum and average values assigned to our options by these theories.

that it matters a great deal which of these proposals we adopt. But since the various proposals will have very similar consequences for our contemporaries, it is likely that the time-relative theory will imply that it matters little which proposal we adopt. Thus, if the time-neutral theory is true, then we could end up adopting a much worse proposal if we accept the time-relative theory than if we accept the time-neutral theory. But if the time-relative theory is true, then we are likely to adopt a proposal of similar value regardless of which theory we accept. Therefore, unless we regard the time-relative theory as much more probable than the time-neutral theory, it would be more rational to accept the latter than the former in choosing among our options.

Since a theory's dispersion is defined in terms of a set of practical options under consideration, it follows that the relative dispersions of contending theories can vary from one decision problem to another. Therefore, the dispersion principle implies that what theory we have most reason to accept can likewise vary from decision problem to decision problem. For example, in situations in which our options would have widely divergent consequences in the short run, but where these differences are likely to wash out in the long run, the stakes are likely to be much higher according to the time-relative theory than according to the time-neutral theory. In such cases, there will be reason to accept the time-relative theory, because it will have the higher dispersion. Likewise, in a situation in which we are starving, the difference in value between finding coconuts and finding nothing will exceed the difference in value between finding treasure and finding nothing, and so the Coconut Account will have a higher dispersion than the Treasure Account and may therefore be more choiceworthy. It follows that the question, "What theory should we accept?" may admit of no general answer, and we may instead have to ask what theory we should accept in making a given decision or in making a given type of decision. Thus, while we have seen that there are some ethical theories that are almost always rational to reject, there may be no ethical theory that is always rational to accept.

In the last part of the previous section, we dropped the assumption that all the theories in which we have credence are acceptance-independent, and we considered a particular kind of acceptance-dependent theory, namely, absolute relativism. We may likewise conclude this section by dropping the acceptance-independence assumption and considering what we may call "moderately relativistic" theories. While an absolutely relativistic theory is one according to which one will act perfectly no matter what ethical theory one accepts, so long as one performs an action recommended by the accepted theory, a moderately relativistic theory is one according to which one is likely to act in a way that is not far from optimal no matter what theory one accepts, so long as one

performs an action recommended by the accepted theory. Thus, while an absolutely relativistic theory implies that it matters not at all what ethical theory one accepts, a moderately relativistic theory implies that it matters little what ethical theory one accepts. A moderately relativistic theory might, for instance, imply that one should act in accordance with a near variant or amended version of whatever theory one actually accepts, so that, by following the theory one actually accepts, one is likely to act in a way that is nearly optimal.

Now suppose our credence is divided between two theories, one moderately relativistic theory, T_M, and one nondeflationary, acceptance-independent theory, T_X. And assume that T_X does not have a very low dispersion. Since T_X is acceptance-independent and nondeflationary, it implies that certain actions would be worse than others regardless of what ethical theory they are chosen on the basis of. Thus, T_X implies that if we accept some other ethical theory, such as T_M, we may end up performing a relatively bad action. Since we are assuming that T_X does not have a very low dispersion, T_X implies that the option we would choose on the basis of T_M might well be considerably worse than the option we would choose on the basis of T_X. Therefore, assuming that T_X is correct, the expected value of accepting T_X will be considerably greater than that of accepting T_M. But since T_M is moderately relativistic, it implies that we are likely to act in a way that is not far from optimal regardless of what theory we accept. And so assuming that T_M is correct, we should not expect to do much better by accepting T_M than by accepting T_X. Therefore, unless we regard T_M as much more probable than T_X, it would be more rational to accept the latter.

IV. THE PROBLEM OF INTERTHEORETIC VALUE COMPARISONS

In the final two sections, I will consider two objections to the arguments I have presented. Both of these objections concern the argument for the rejection of relatively deflationary theories, not the argument for the rejection of absolutely deflationary theories. The first may be stated as follows:

> The argument for the rejection of relatively deflationary ethical theories assumes that we can compare the disparities of different ethical theories. That is, it assumes that it makes sense to say that the margin by which the best option exceeds the average option is greater according to one ethical theory than another. And this presupposes that value differences can be compared across ethical theories. But such comparisons are unintelligible. It is only possible to compare value differences within a theory, not across theories. For example, it makes sense to say that according to Peter Singer's moral theory, the difference in value between causing and preventing pain to an animal is equal to the difference in value between

> causing and preventing the same amount of pain to a human,[9] and it likewise makes sense to say that according to a traditional moral theory, the former value difference is smaller than the latter. But it does not make sense to say that the difference in value between causing and preventing pain to an animal is greater according to Singer's moral theory than it is according to traditional morality. For there is no common value scale shared by both theories.[10]

In response to this objection, it should first be noted that the claim that one cannot compare value intervals across ethical theories has very counterintuitive implications. As an illustration, consider the following argument: "I don't know whether Singer's moral theory is correct or whether the traditional morality I was raised with is correct. They both seem about equally plausible. I have a choice between ordering a veal cutlet and ordering a veggie wrap. If Singer's moral theory is correct, then I should order the veggie wrap so as not to contribute to a system that causes great suffering to animals. However, if the traditional theory is correct, then I should order the veal cutlet so as to support the local farmers. But if Singer's theory is correct, then ordering the veggie wrap would be much better than ordering the veal cutlet, while if the traditional theory is correct, then ordering the veal cutlet would be only slightly better. And so given my uncertainty, I should order the veggie wrap." This seems like a perfectly good argument. But if intertheoretic value comparisons were impossible or unintelligible, then we would have to reject this argument, and it is difficult to see how a rational choice could be made between our options.[11]

Moreover, the impossibility of rational choice in situations of this kind would have serious consequences. For I have argued that if inter-

9. See Peter Singer, *Animal Liberation*, 2nd ed. (New York: Avon, 1990).

10. James Hudson raises this problem in "Subjectivization in Ethics," *American Philosophical Quarterly* 26 (1989): 221–29. Lockhart's response to this problem is to propose what he calls the Principle of Equality among Moral Theories (*Moral Uncertainty*, 84). According to this principle, the difference in moral value between the best and worst option in any given situation should be considered equal across all moral theories. But this principle is incompatible with the fact that different moral theories can disagree concerning which of two choice situations is more morally significant. For example, utilitarianism might imply that more is at stake in deciding what charity to give money to than in deciding whether to make a false deathbed promise, while Kantianism might imply the reverse. And from this it would follow that the two theories cannot agree concerning how much is at stake in both choice situations. Thus, Lockhart's principle must be false.

11. Of course, in the absence of intertheoretic value comparisons, one might simply follow the theory one finds most plausible, so that if one regarded the traditional morality as ever so slightly more plausible than Singer's theory, one would order the veal cutlet, and if one regarded the two theories as equally plausible, one would flip a coin. But as is shown by this case, and others we have considered above, this hardly seems like a rational solution.

theoretic value comparisons are impossible, then the only options one could rationally rule out would be options that are worse according to some theories in which one has credence and better according to none. But I have also argued that where one has credence in a sufficiently broad range of ethical theories, we should not expect there to be many options that can be eliminated on this ground. Thus, given the epistemic situation most of us find ourselves in, the denial of the possibility of intertheoretic value comparisons would imply that among most of our options there is no basis for rational choice. In other words, it would imply the near impotence of practical reason.

Fortunately, inter theoretic value comparisons are no more difficult to explicate than intratheoretic value comparisons. The latter can be explicated in terms of their practical implications. To say, for example, that according to Singer's moral theory, a given amount of human suffering is equally bad as the same amount of animal suffering is to say, among other things, that according to Singer's theory, we should be indifferent between producing a given amount of human suffering and producing the same amount of animal suffering, other things being equal. Likewise, to say that according to the traditional moral theory, human suffering is a thousand times as bad as animal suffering is to say, among other things, that we should be indifferent between a probability of p that a given quantity of animal suffering is produced and a probability of $p/1{,}000$ that the same quantity of human suffering is produced, other things being equal. In other words, intratheoretic value comparison can be explicated in terms of claims about what choices would be rational on the assumption that the theory in question is true.

Similarly, we can explicate intertheoretic value comparisons in terms of claims about what choices would be rational assuming that the ethical theories in question had certain subjective probabilities. Thus, to say that the difference in value between ordering the veal cutlet and ordering the veggie wrap is one hundred times as great according to Singer's theory as it is according to the traditional moral theory is to say, among other things, that if one's credence were divided between these two theories, then it would be more rational to order the veggie wrap than the veal cutlet if and only if one's degree of credence in Singer's theory exceeded .01.

If we can explicate value comparisons in terms of what they imply about rational choice under uncertainty, then we can figure out what we take the correct value comparisons to be by asking ourselves what sorts of choices we would regard as rational. Employing this method, we can figure out the implicit value ratios that obtain both within and across the ethical theories in which we have credence. Thus, if I have credence in a particular ethical theory that I have never explicitly formulated in quantitative terms, and if I want to determine what this

theory implies about the relative disvalues of human and animal suffering, then I can simply ask myself what choices I would regard as rational over gambles involving human and animal suffering, assuming I had a degree of credence of 1 in this theory. Likewise, if I have credence in Singer's theory and in a traditional moral theory, and if I have already determined the value comparisons that obtain within each of these theories, then I can figure out how to commensurate the value scales of the two theories by asking myself what choices I would regard as rational given certain probability assignments to the two theories.

It appears, therefore, that since the inter- and intratheoretic value comparisons are susceptible to similar forms of analysis, the former should be regarded as no more mysterious than the latter. It may often be impossible to make intertheoretic value comparisons with precision, especially when the ethical theories in question are very different from one another. Likewise, it may often be impossible to make intratheoretic value comparisons with precision, especially when the options or outcomes being compared are very different from one another. But fortunately, for most practical purposes, we don't require perfect precision in either kind of comparison.

There are some cases in which the task of commensurating two alternative ethical theories will be greatly facilitated. These are cases in which, for some pair of options, we know that the difference between their values is the same according to both ethical theories. This will obtain whenever there is a pair of options that does not differ in relation to any of those factors about which the two theories disagree. Suppose, for example, that I am uncertain what is the correct theory of rights. My credence is divided between two such theories, T_1 and T_2. Suppose, however, that I have a background theory, T_B, that evaluates my options in relation to all considerations other than those deriving from rights. And suppose I am fully confident that this background theory is true. Thus, my credence is divided among two complete ethical theories, the first, which we may call T_{B+1}, consisting in the conjunction of T_B and T_1, and the second, which we may call T_{B+2}, consisting in the conjunction of T_B and T_2. Now suppose there is a pair of options, i and j, such that, according to both T_1 and T_2, no one's rights are at stake in the choice between i and j (i and j might, e.g., be the options of giving either of two alternative gifts). Since no rights are at issue, T_B alone will suffice to evaluate these options, and so T_{B+1} and T_{B+2} will agree concerning their values. Therefore, these alternative ethical theories will agree concerning the difference between the values of these options. We may now define "one unit of value" as the magnitude of this difference. And having thus defined a common unit of value for the two theories, it will follow that so long as we can compare the value intervals within each

of these theories, there will be no difficulty comparing value intervals between the two theories.

One motivation for the view that intertheoretic value comparisons are impossible may be that it follows from an assumption normally made in decision theory. The assumption in question is that if two value functions are linearly equivalent (i.e., if one is a positive linear transformation of the other), then for practical purposes there is no difference between them, and so they should be regarded as equivalent tout court. But this assumption should be questioned. Let T_X and T_Y be two evaluative theories whose value functions are linearly equivalent. Now it is true that having *full* credence in T_X will be equivalent, for practical purposes, with having *full* credence in T_Y. But for reasons we have already seen, having *partial* credence in T_X need not be practically equivalent to having the same degree of credence in T_Y. And so the scale of a value function can matter from the practical point of view.

It can also matter in other ways, quite apart from issues raised by evaluative uncertainty. For value is normative not only for action but also for a number of attitudes. Consider disappointment. Disappointment is warranted whenever an action has an outcome that is worse than the agent reasonably expected. And the degree of disappointment that is warranted is proportional to the difference between the actual value and the expected value. Two linearly evaluative theories can disagree concerning the magnitude of this difference: one might imply that the actual outcome is much worse than should be expected, while the other might imply that it is only slightly worse. These theories will therefore disagree concerning the degree of disappointment that is warranted.

V. AN ARGUMENT FOR FANATICISM?

The most common objection to my position is that it seems to justify fanaticism. The objection may be stated as follows:

> Your view implies that for any reasonable ethical theory we might consider accepting, there is a crazy or fanatical theory that is more worthy of acceptance. Let T_X be any such reasonable theory. We can easily construct another theory, T_Y, whose disparity exceeds that of T_X by any margin one likes. One might, for example, construct a theory according to which wiggling one's toes is incredibly good, much better than the average option. So long as we construct T_Y so as to make this difference sufficiently great, it will turn out that we have more reason to accept T_Y than T_X.

If I claimed that how worthy a theory is of acceptance is determined entirely by its disparity, then it would indeed follow that for any reasonable theory, there is a fanatical theory that is more worthy of accep-

tance. Indeed, it would also follow that for any fanatical theory, there is a still more fanatical theory that is more worthy of acceptance and hence that no theory is most worthy of acceptance. But I have argued that how worthy a theory is of acceptance depends not only on its disparity but also on its subjective probability. And I have noted that in simple cases, a theory's subjective probability and disparity combine multiplicatively in determining the expected value of accepting it. This implies that no matter how high a theory's disparity may be, it will be unworthy of acceptance if its subjective probability is sufficiently low. Further, if we take a theory that we regard as plausible and modify it arbitrarily so as to increase its disparity, we are likely to regard the resulting theory as less probable than the theory from which it was derived. So long as, in making such modifications, the subjective probability of the resulting theory diminishes faster than its disparity increases, it will be irrational to accept the resulting theory.

More serious problems arise, however, when we consider what we may call ultrafanatical theories. These are theories whose disparity exceeds that of any reasonable theory by an infinite ratio. So long as such a theory has a positive, noninfinitesimal probability, the product of its probability and its disparity will be greater than that of any reasonable theory, and so my view implies that this theory will be more worthy of acceptance than any reasonable theory.[12] Naturally, this Pascalian conclusion may seem counterintuitive.

One possible response is to deny the validity of the argument leading to this conclusion. For one might insist that the decision-theoretic reasoning applies only when values are continuous and hence when there are no infinite value ratios. But a better response may be to endorse the Pascalian conclusion, however counterintuitive it may seem at first. For suppose one's credence is divided between a single ultrafanatical theory, T_P, and a number of nonfanatical theories. In this case, the difference between the maximum and average value assigned to one's options by T_F will be immeasurably greater than the corresponding difference for any of the other theories in which one has credence. Thus, in determining which option has the greatest expected value, the contribution made to this estimate by T_F will swamp those made by all the other theories. Therefore, the option to which T_F assigns the highest value is likely to be the option with the highest overall expected value.

12. What if an ultrafanatical theory has an infinitesimal probability? Suppose, e.g., that one has credence in two theories, T_F and T_X, such that the disparity of T_F exceeds that of T_X by an infinite ratio, but its probability is exceeded by that of T_X by an infinite ratio. In this case, there is no simple answer to the question of what theory one should accept. Insofar as this question has a well-defined answer, this answer will depend on the magnitudes of the infinitudes involved. See Alan Hájek, "Waging War on Pascal's Wager," *Philosophical Review* 112 (2003): 27–56.

And so it is likely that if one were to accept T_P, one would choose the same option one would choose if one were to take into account all the theories in which one has credence. Thus, we should expect that including the other theories in deliberation would simply make one's decision problem more difficult without much affecting what one ultimately decides. Since the aim of accepting a single theory is to simplify one's cognitive task while arriving at the best decision possible, it follows that in the present case it would be rational to accept T_P. If, therefore, one is subject to rational criticism in this case, it is not in choosing to accept T_P but rather in having a positive, noninfinitesimal degree of credence in a theory that is so fanatical that its contribution to the expected values of one's options swamps that of all other theories.

In conclusion, in deciding what theory to accept, it is a mistake to consider only the probabilities of the contending theories, for we must also take into account their degrees of disparity. But it is equally a mistake to ignore the probabilities of the contending theories and to opt for whatever theory has the greatest disparity. In making a rational choice among contending theories, it will often be necessary to weigh these factors against one another. But in deciding whether to accept or reject absolutely deflationary theories, such as nihilism and absolute relativism, no such weighing is necessary, for it is rational to reject such theories regardless of their probabilities and regardless of the probabilities and disparities of the alternative theories.

Appendix
Estimating the Value of Accepting a Theory

In the following, S denotes the set of options under consideration, i denotes an option belonging to S, and A, B, and so on, denote theories.

For each theory A, let $t_A(i)$ be the value of choosing option i according to A. We suppose that for each theory A we can treat $\{t_A(i) : i \in S\}$ as jointly distributed random variables for which

$$\mathrm{E}t_A(i) = \mu_A, \ \forall\, i \in S. \tag{A1}$$

Moreover, we assume that for any two different theories, A and B, we can treat

$$\{t_A(i) : i \in S\} \text{ and } \{t_B(i) : i \in S\} \tag{A2}$$

as stochastically independent. From equation (A1), and from the law of iterated expectation, it follows that if I is a randomly chosen option that is stochastically independent of the values $\{t_A(i) : i \in S\}$, then

$$\mathrm{E}t_A(I) = \mu_A. \tag{A3}$$

Let I_A be the option one would choose if one were to accept theory A. Formally,

$$I_A = \arg\max_{i \in S} t_A(i),$$

so that

$$t_A(I_A) = \max_{i \in S} t_A(i).$$

Let

$$M_A = \mathrm{E}\{t_A(I_A)\} = \mathrm{E}\left\{\max_{i \in S} t_A(i)\right\}.$$

That is, M_A is the expectation for the maximal value assigned to the options in S by theory A.

Let $v(A)$ be the unconditional expected value of accepting theory A and hence of choosing the option it recommends, namely, I_A. Thus, where H ranges over the set of theories in which one has credence,

$$v(A) = \sum_H p_H \mathrm{E} t_H(I_A) = p_A M_A + \sum_{H \neq A} p_H \mathrm{E} t_H(I_A) = p_A M_A + \sum_{H \neq A} p_H \mu_A,$$

where the last equality follows from the independence assumption (eq. [A2])—which entails that I_A and $\{t_H(i) : i \in S\}$ are independent if $A \neq H$—and equation (A3).

Then

$$\begin{aligned} v(A) > v(B) &\Leftrightarrow p_A M_A + \sum_{H \neq A} p_H \mu_H > p_B M_B + \sum_{H \neq B} p_H \mu_H \\ &\Leftrightarrow p_A M_A + p_B \mu_B > p_B M_B + p_A \mu_A \\ &\Leftrightarrow p_A (M_A - \mu_A) > p_B (M_B - \mu_B) \\ &\Leftrightarrow p_A d_A > p_B d_B, \end{aligned}$$

where d_A is the disparity of theory A. In other words, the expected value of accepting theory A will exceed that of accepting theory B if and only if the product of the probability and disparity of theory A exceeds that of theory B.