REVENGE OF WOLFMAN:
A PROBABILISTIC EXPLICATION OF FULL BELIEF

Jonathan Roorda[1]
(November 1995)

> "To some people, life is very simple . . . no shadings and grays, all blacks and whites. . . . Now, others of us find that good, bad, right, wrong, are many-sided, complex things. We try to see every side; but the more we see, the less sure we are."
> —Sir John Talbot, *The Wolf Man* (Universal Pictures, 1941)

> ". . . [N]or am I disturbed by the fact that our ordinary notion of *belief* is only vestigially present in the notion of degree of belief. I am inclined to think Ramsey sucked the marrow out of the ordinary notion, and used it to nourish a more adequate view. But maybe there is more there, of value. I hope so. Show me; I have not seen it at all clearly, but it may be there for all that."
> —Richard Jeffrey, "Dracula Meets Wolfman: Acceptance vs. Partial Belief"

What follows is a discussion of the nature of belief. This description, however, is apt to promote misunderstanding, so I want to begin with a brief outline of the broader philosophical project to which I take this paper to be a contribution.

Formal epistemology is the term I will use to refer to a particular philosophical approach to epistemological issues. The technique in question has played an important role throughout the history of philosophy, but it is only in the 20th century that it has been developed with enough methodological self-consciousness to separate it clearly from other epistemological approaches. The project of formal epistemology, roughly speaking, is to devise a representational system which allows us to characterize in a precise way those features of human opinion in which we are interested. The concepts employed in such a formal system will typically not be the fuzzy and ill-understood terms of epistemic characterization and evaluation which occur in ordinary language, but will instead be

[1] The author of this draft of a paper died in June 1996, at age 25. He would have revised it further before publication, or might have decided not to publish it at all. His two published papers are: `Fallibilism, ambivalence, and belief', *Journal of Philosophy* **94**(1997)126-55 and `Kitcher on theory choice', *Erkenntnis* **46**(1997)215-39. —Richard Jeffrey

rigorous explications of those ordinary concepts (a la Carnap) or entirely novel philosophical inventions. Built into the formal rules governing which representations count as well-formed or acceptable are broad normative constraints which we take rational states of opinion to necessarily embody; in other words, we distinguish between appropriate and inappropriate states of opinion by restricting the grammar of the formal system to permit only representations of normatively adequate opinions as grammatical. Using such a system, one can engage in a process of self-correcting inquiry: if, upon reflecting upon one's own opinions, one discovers that they cannot be translated into the formal system in any way which does not violate the rules of well-formedness, then inasmuch as one accepts the normative principles which motivate the choice of rules, one has good reason to regard one's state of opinion as inadequate and in need of revision. The ultimate goal of those who pursue the formal epistemology project is to produce a system familiar enough that people understand easily how to translate their opinions into the formal apparatus, well-motivated enough that the underlying normal principles seem plausible to a wide range of potential users, and powerful enough to provide unexpected and useful recommendations for updating one's opinion.

The above description will perhaps become clearer if we consider a few specific examples of attempts to work out the formal epistemology program in detail. Let us begin with what is surely the best-known formal model of opinion, the *belief model.* First, we define a domain of subject matter over which all possible states of opinion will range. For present purposes, I will employ the familiar device of possible worlds in this capacity. (Of course, there are other popular choices at this juncture; we could proceed equally well, for example, talking in terms of possible individuals (Lewis 1983) or maximal consistent sets of sentences in some logically regimented language (Carnap 1947).) Within the context of this paper, I will in general assume that the domain of possible worlds is of finite size. I wish to stress, however, that this is a decision made for expository convenience only. In the larger work of which the present paper is a summary, all of the results presented here are reconstructed in slightly more cumbersome form with no cardinality restrictions whatsoever. The added complexity would only detract from the clarity I seek here, however.

Given our domain of possible worlds, we can proceed to define *propositions* in the usual way as sets of worlds. These propositions form a Boolean algebra bounded by the impossible proposition (the empty set) and the necessary proposition (the set of all worlds). *Beliefs* can now be defined as attitudes which agents bear toward propositions; roughly speaking, to believe a proposition P is to hold that the actual world lies inside the set P and to rule out all worlds outside of P as contenders for actuality. (A genuine formal treatment of belief would of course require a much more rigorous definition; I am just going to presume that you have sufficient familiarity with the concept to know what I mean.) Finally, we impose the following constraints on well-formed beliefs: (i) that no agent believes P and believes Q without believing the intersection of P and Q; and (ii) that no agent believes the impossible proposition. These requirements are collectively referred to as *Deductive Cogency.* It is quick business to demonstrate that an agent satisfies Deductive Cogency if and only if there is some non-empty proposition K such that for all propositions P, the agent believes P just in case P is a superset of K. Let us call K the agent's *corpus;* the corpus can be regarded as the single strongest statement of the agent's beliefs, as the one belief from which all other beliefs deductively follow. The belief model places normative constraints on actual inquiry by requiring that one not knowingly adopt a state of opinion which cannot be cogently represented by a non-empty corpus. So, for example, if one finds oneself inclined to believe that P, that Q, and that $(-P \cup -Q)$, one can derive from the model the important result that these opinions cannot cogently be maintained simultaneously.

The belief model is the most familiar development of the formal epistemology program, but it is by no means the only possible development. In fact, it has been challenged (some would say eclipsed) in recent years by an important rival, the *probability model.* In this model, we begin once again with a finite domain of possible worlds representing the content of possible states of opinion. This time, however, we represent states of opinion not as individual propositions but as numerical measures distributed across the entire set of propositions. We can summarize the normative constraints of the model, jointly referred to as *Probabilistic Coherence*, as follows: (i) the entire measure must sum to 1; (ii) no measure may be distributed to the impossible proposition; and (iii) the measure assigned to the union of two disjoint propositions must equal the sum of the measures assigned to the individual propositions. Following de Finetti (1974), we will interpret these probability measures in terms of mathematical expectations. For each proposition P, let I(P), the *indicator variable* of P, be the random quantity which takes a value of 1 just in case the actual world is in P and 0 just in case the actual world is not in P. Then the probability which an agent assigns to the proposition P will be that agent's numerical estimate of the value of I(P): the higher the estimate, the more epistemic confidence the agent grants to the possibility that the actual world is in P. As with the belief model, it is easy to see how normative constraints on actual opinion can be derived from the axioms of the model: an agent who is inclined to assign a very high probability to P but only a moderately low probability to $P \cup Q$, for example, is subject to immediate correction on the grounds that no coherent probability measure can support such an assignment.

I hope I have conveyed enough in this thumbnail sketch to give some sense of how I am using terms such as "belief" and "probability" in this essay, and to make it clear in just what sense I take the belief model and the probability model to be rival theories. They are not, for example, intended to be competing empirical theories about the way in which opinions are actually stored in the head. Questions concerning the nature of mental representation are currently lively in cognitive science and the philosophy of mind, but the resolution of these questions has very little bearing on the usefulness of the models of opinion I am considering here. Likewise, these models are not supposed to offer a practical program for organizing all of our everyday reasoning; the computational costs involved in maintaining scrupulous adherence to the rules of these models are simply not feasible for ordinary thinkers, as Harman (1986) and Cherniak (1986) have stressed. The models of formal epistemology are purely philosophical tools, theoretical constructs which constrain our reasoning only inasmuch as we find them useful and reasonable to apply. There is (for present purposes) no factual question at stake of whether we "really" reason in terms of lists of sentences or arrays of numbers or what have you; there is only the pragmatic question of how we can go about devising and applying formal techniques for conceptualizing and improving our opinions by our own lights. So in a very important sense, the belief model and the probability model are not rivals at all. Each has its own advantages: the belief model cleaves closely to our most familiar everyday epistemic concepts and sustains fruitful parallels between the structure of private deliberation and public inquiry, while the probability model gives us a rich taxonomy of fine epistemic distinctions and can be embedded in a larger theory of mathematical expectation to give a powerful unified account of preference and decision. We are free to use each model as we see fit within the domain of application where it seems most useful, maintaining an overarching Carnapian attitude of tolerance rather than forcing a commitment of allegiance to one or the other.

On the other hand, the domains of the belief and probability models are too closely related for us to pretend that we can apply the two theories in strict isolation. And herein lies the problem I am concerned with: the descriptions of an agent's opinions in terms of beliefs and of probabilities place mutual constraints on one another, constraints which we ought to be able to capture in a hybrid theory employing both terminologies. Our two formal models began as explications of loose concepts familiar from ordinary life, and specifically as attempts to formalize the constraints that judgments expressed in terms of such concepts place upon one another. Beginning with simple intuitive insights such as that one ought not believe P and –P at the same time, we construct the deductive logic which is the basis of the belief model; and from similar primitive data such as that one ought not place a great deal of confidence in P and –P at the same time, we construct the inductive logic at the heart of the probability model. The crucial point which I want to stress here is that the same body of intuitive judgments also places constraints on joint assignments of beliefs and probabilities. Surely it strikes most of us as incoherent, or at least normatively suspect, for an agent to express a set of simultaneous judgments of the form, "I believe that P, but I am much more confident that –P than I am that P." Neither of our present models has anything to say about the inappropriateness of this judgment, however. What we need is a translation manual to take us between the two theories, an explicit way of expressing the concepts and judgments of one model using the terminology of the other. This is the project I want to address here: the possibility of defining, in purely probabilistic terms, a set of necessary and sufficient conditions for believing a proposition.

At this point, the more impatient among you may feel that I have wasted a good deal of time on an elaborate development of what is essentially a non-problem. After all, most standard presentations of probabilistic epistemology already contain a simple account of full belief which has a great deal of historical and intuitive support. This familiar account, which I shall refer to as the *Received View,* maintains that to believe a proposition is just to assign a probability value of 1 to it. The Received View trades on the metaphor which Ramsey (1990) used to introduce the concept of subjective probability, that of partial belief. If we think of the scale of probability values as a continuum of intermediate epistemic states falling partway between belief and disbelief, it seems only natural to use the extreme points of the continuum, 1 and 0, to represent the total epistemic commitments of which intermediate probabilities are partial counterparts. The Received View can plausibly be regarded as the dominant position within probabilistic epistemology; its truth has been tacitly assumed by virtually all writers on the subject who have not addressed the question of the analysis of belief explicitly. (For example, we can attribute some form of the Received View to nearly every probabilist who has regarded conditionalization as the appropriate response to the acquisition of new information.) Moreover, we can construct a powerful argument in favor of the Received View, using the following premises:

(P1) To believe a proposition P is to maintain that the actual world is in P. (Definition of belief)

(P2) To maintain that the actual world is in P is to maintain that the value of the indicator variable I(P) is 1. (Definition of indicator variables)

(P3) To maintain that the value of the indicator variable I(P) is 1 is to estimate the value of the indicator variable I(P) to be 1. (Definition of estimation)

(P4) To estimate the value of the indicator variable I(P) to be 1 is to assign a probability of 1 to P. (Definition of probability)

(C) To believe a proposition P is to assign a probability of 1 to P. (The Received View)

Let us call this argument the *ur-argument.* In light of the obvious strength of its premises, why can't we just pack up and go home now?

The problem is that there are serious mismatches between the concepts of belief and probability 1, mismatches which get papered over by the glib treatment offered in the ur-argument. These discrepancies have been noted by most authors in the probabilist tradition who have explicitly addressed the problem of belief; indeed, the rejection of the

Received View by such authors has achieved nearly the same article-of-faith status that its acceptance has received from thinkers who have only implicitly dealt with the issue. Here, I want to focus on what I take to be the three strongest challenges to the Received View, presented in order from weakest to strongest. In fact, I think it will emerge quite shortly that these challenges are really just three different facets of the same fundamental problem, and that resolving this single discrepancy is the key to the successful analysis of full belief. Still, it will serve us best to treat the challenges separately at first.

Challenge 1. As has already been noted, one of the important functions of the probability model is to provide a rigorous explication of the intuitive idea that one can invest varying degrees of confidence in the truth of various propositions. At the heart of the first challenge is the observation that we often make these same comparative confidence rankings with respect to propositions all of which we believe. For example, consider the following four propositions.

(1) Millard Fillmore was born on January 7.
(2) Millard Fillmore was the 13th President of the United States.
(3) Millard Fillmore was the President of the United States.
(4) Millard Fillmore either was or was not the President of the United States.

I fully believe each of these propositions. But I am not prepared to judge that I am equally certain that each of them is true. On the contrary, I have deliberately arranged the propositions to reflect the increasing order of confidence I assign to them: I am more confident in (2) than in (1), more confident in (3) than in (1) or (2), and so on. If we take the probability model as our explication of epistemic confidence, however, the Received View does not allow me to sustain this parcel of judgments. Since I believe each of (1)-(4), I must assign a probability of 1 to each proposition, and hence I must be exactly as confident in the truth of (1) as I am in the truth of (4). But this seems just wrong: I am not as confident in my knowledge of Fillmore's birthday as I am in the truth of the tautology expressed by (4), and it is a mark against the Received View that it seems to require me to hold to such a judgment.

Challenge 2. The Received View runs into similar problems when we attempt to elucidate the relationship between belief and practical decision-making. According to a familiar rule of practical decision, we can embed the probability model in a more general decision theory by using an agent's probability judgments and subjective utility assignments to calculate the agent's expected net utility gain for various actions with uncertain outcomes. An immediate consequence of this approach is that any propositions which receive probability 0 can be effectively ignored in practical deliberation: the utility of a possible outcome with probability 0 will not have any effect on calculations of expected utility. This point can be expressed by the following maxim: one ought always to act as if all propositions to which one assigns probability 1 are true. Combined with the Received View, the maxim reads: one ought always to act as if all propositions one believes are true. But this version of the maxim violates our intuitions about the relationship between belief and decision: most of us, I take it, are willing to grant that it is sometimes permissible to hedge our bets, taking actions which are not utility-optimal on the assumption that all of our beliefs are true. This point is traditionally made through reference to some sort of wager at incredibly high stakes. Consider, for example, a gamble which earns me three dollars if Millard Fillmore was the 13th president but which costs me a horribly painful death if Fillmore was not the 13th president. Since I believe that Fillmore was the 13th president, the maxim requires me to accept this bet, but most of us would surely grant that I would not be irrational to turn it down. Examples of this sort have often been regarded as suspect at best, since the introduction of such wildly high stakes can have distorting factors on our intuitions; so I think it is worth noting that the challenge can be launched just as successfully even when nothing so portentous as life and death is at stake. Let us suppose, for example, that I am a contestant on a Jeopardystyle television quiz show, and I am given the opportunity to wager some or all of my point total on my knowledge of Millard Fillmore's birthday. Since I believe that Fillmore was born on January 7, the Received View recommends as the uniquely utility-optimizing decision to bet my entire bank on this answer. But it may well strike me as strategically prudent in such a situation to bet somewhat less than my total, keeping a safety margin in case my belief turns out to be wrong. In short, then, the concepts of belief and probability 1 play importantly different roles in practical decision, despite the attempts of the Received View to shoehorn them together.

Challenge 3. This is, I think, the most serious and general challenge to the Received View. Most of us, I take it, adopt an attitude of *fallibilism* toward at least some of our beliefs. That is, most of us grant that it is possible that some of our beliefs are false. More importantly, I think most of us would grant that this is not just a remote logical possibility but a *serious* possibility, a possibility which deserves to play a non-negligible role in our reasoning. At the very least, I hope that even those among us who are not fallibilists about their own beliefs would grant that fallibilism is a coherent attitude to take, that it is not irrational as such to take seriously the possibility that one is mistaken in at least some of one's opinions. (My hope is in this case possibly unfulfilled, since a number of misguided philosophers have suggested that fallibilism is incoherent or inappropriate.) But it is not at all clear that the Received View is compatible with fallibilism. One way of seeing the point is this. A natural way to elicit my commitment to fallibilism within the probabilist framework would be to present me with sets of propositions I believe and ask me to estimate the number of truths in the set. Inasmuch as I endorse fallibilism with respect to any given set of my beliefs, I should be willing to post an estimate for the number of truths in that set which is less than the number of propositions in the set. But a trivial theorem of the probability calculus (Jeffrey (1992) calls it "De Finetti's Law of Small Numbers") tells us that one's estimate of the number of truths in a collection of propositions must equal the sum of the probabilities one assigns to those propositions. So if we accept the Received View that all full beliefs receive probability 1, we must conclude that the only estimate I can rationally post for the proportion of truths among my beliefs is 100%. And this just seems wrong—I certainly don't feel rationally compelled to rule out as a serious possibility that anyone else has false beliefs, and I don't see why I myself should be the sole exception to this rule. I've been wrong before, I'll be wrong again, and although I must admit I don't see just where at the moment, I'm modest enough to grant that I may well be wrong about something right this instant.

And here, I think, is really the basic problem. The concepts of belief and of probability 1 just differ in their relation to the important third concept of serious possibility. To assign probability 1 to a proposition really is to rule out all alternatives to that proposition as serious possibilities. This is just a consequence of the mathematics of estimation: to formulate an estimate is to calculate a probability-weighted average of a set of alternate possibilities, so a possibility which receives no probability receives no weight. It just doesn't figure; it might as well not even be in the set of alternatives. But to believe a proposition isn't as strong a commitment; it doesn't force us to wholly disregard all alternatives. The direction of our reasoning is often shaped in profound ways by serious contemplation of the possibility that some of our beliefs are false, and we take this to be a rational state of affairs. We saw one application of this difference in the second challenge. When we reason about practical decisions, we don't need to worry about the probability 0 options, but we do need to worry about possibilities which we fully believe will not arise. The third challenge is an extension of this point from the sphere of practical rationality to the wider sphere of rationality in general. We need to find within the probability model a way of representing full beliefs which allows us to take seriously the possibility that our beliefs are false. The Received View must therefore be rejected.

In response to these challenges, a number of philosophers have proposed alternatives to the Received View. I won't attempt an exhaustive review of the relevant literature here, but prominent players in this game include Henry Kyburg, Paul Teller, John Harsanyi, Robert Nozick, Ronald de Sousa, Mark Kaplan, Patrick Maher, Richard Foley, and Isaac Levi. All of these authors have proposed probabilistic analyses of full belief which allow for believed propositions to be assigned probabilities less than 1. The various accounts on offer vary widely in the details: some authors have proposed a simple numerical threshold, others have suggested that belief be relativized to one's decision context or other circumstances, others have proposed "cognitive utility functions" which allow one to determine decision-theoretically what to believe, and so on. Elsewhere, I review each of these proposals with some care and argue that they all fail on their own terms, i.e., that none of them cleaves closely enough to the concept of belief to count as an adequate explication. For the purposes of this discussion, however, I just want to point out that none of the available accounts does sufficient justice to the ur-argument for the Received View. Despite the above challenges, we are still faced with a very strong deductive argument, none of whose premises we have yet said anything to undermine. None of the authors I have named really takes the ur-argument seriously, though; certainly, none of them says anything which directly challenges it, and none of them offer arguments for their own accounts which are more compelling than the ur-argument. There simply seems to be something incoherent about simultaneously maintaining that a proposition is true and estimating the indicator of its truth-value to be some quantity less than 1. The depressing conclusion I invite you to draw is that no explication of belief is possible within the confines of the probability model. That model simply does not have the conceptual resources to draw the necessary distinction between taking a proposition to be true and using that proposition as one's standard of serious possibility. The task we have set ourselves is insoluble.

All is not lost, however. In the true spirit of the formal epistemology project, we respond to the demonstration of conceptual inadequacy in our probability model by tinkering with the model. Toward that end, I want to propose a third epistemic model, the *extended probability model.* In this new model, states of opinion are represented not as single probability functions but as sets of probability functions. This extension of the standard probability model is certainly not a novel move on my part; variants on the idea have been suggested by a host of authors in the probabilist tradition, and few contemporary treatments of probabilism fail to pay at least lip service to the idea. Nevertheless, I want to take a bit of care in my development of the extended model, since my interpretation of the model differs in certain crucial ways from the standard interpretations on offer.

Perhaps the best way to introduce the idea behind the extended model is by further developing the analogy between the belief and probability models. Recall that in the belief model, there is a one-to-one mapping between propositions and states of opinion, since each available belief state is associated with a propositional corpus. We can thus distinguish between two types of belief states: fundamental belief states, i.e., those whose corpus contains exactly one possible world, and derivative belief states, i.e., those whose corpus contains more than one possible world. Suppose that, instead of presenting the standard belief model in the above introduction, I had instead developed a restricted version of that model in which the only admissible states of opinion are fundamental belief states. In this *limited belief model,* a rational agent either believes or disbelieves every proposition; there is no matter on which the agent suspends judgment. In other words, the limited model mandates that one be hyper-opinionated, that is, that one be willing to stake a definite position on one side or the other of every issue. Such a model of human opinion would not be completely worthless; it might in fact serve as a reasonable idealization with respect to a wide range of epistemic questions. Still, as we developed this model, we would surely encounter severe strains and mismatches between our ordinary conception of belief and the limited model's explication, strains which would prevent us from effectively applying the model to a number of important issues. The solution to the problem is obvious: we can relax the requirement of hyper-opinionatedness by allowing that there may well be propositions which a rational agent neither believes nor disbelieves. The standard belief model accomplishes this by allowing for derivative as well as fundamental belief states. An agent whose corpus contains more than one possible world may be regarded as being in a state of *ambivalence* with respect to a number of fundamental belief states. Such an agent identifies to a certain extent with each fundamental belief state contained in his corpus, but is fully committed only to those judgments which are common to all of those fundamental states. On those issues where the fundamental states disagree, the agent suspends judgment. The standard belief model can thus be regarded as an extension of the limited belief model; it contains the latter model as a limiting case, but also allows for states of opinion which are less hyper-opinionated than those of the limited model.

The extended probability model is related to the standard probability model in exactly the same way that the standard belief model is related to the limited belief model, and the justification for the extension is roughly the same in both cases. The crucial insight is that a single probability distribution is in fact an absurdly precise state of opinion, representing a far stronger set of epistemic commitments than most of us are ordinarily willing to engage in. This is most readily seen if we employ the useful operational device of regarding one's probability assignments as judgments about fair betting odds. Suppose that you have two acquaintances whom you hold with equal regard, so you have no reason to prefer that one benefits at the other's expense. Both friends have utility functions which are linear with respect to monetary values, i.e., all that they value is money, and they place the same value on a given increment of monetary gain no matter how much money they already have. These friends want to wager against each other on the proposition P, but they are willing to defer to your judgment about what odds they should stake on P's truth. So they come to you and ask, "What if I were to stake x dollars on P and she were to stake 1–x dollars on not-P? Would that give one of us an unfair advantage?" If you are really committed to helping them devise a fair bet, you should give your approval to their proposal just in case x is the probability you assign to P. But once we frame the issue in this way, it should be obvious that there are very few propositions such that you would be willing to give your approval to exactly one number x, rejecting all others as too high or too low. Such precision is justified for bets involving carefully designed and tested games of chance such as dice and roulette wheels, perhaps, but not for most of the messy issues in life. Suppose, for example, that the hypothetical friends asked me to help them devise a bet on whether Bob Dole will win the 1996 Presidential election. If they suggested 20:1 odds either way, I would surely reject the proposal; given what I know about the state of politics, both of these bets strike me as patently unfair. But as the odds become more balanced, my ability to confidently reject the proposals becomes shakier. Even odds seem fair to me; but so do odds which give Dole a slight edge, say 11:9. Perhaps with a little statistical and political research and some voter surveying I could sharpen my opinion considerably, but it seems doubtful to me that even in a fully informed state I would be willing to commit myself to a single set of fair odds. Since the probability model requires me to assign a single probability value to the proposition that Dole will win, it forces me to be more hyper-opinionated than I regard as rational for someone in my epistemic position. The advantage of the extended probability model is that it allows me to represent my opinion as a set of probability distributions, one for each package of betting odds which I regard as fair.

Another way of understanding the extended probability model, one which does not make use of the operational definition of probabilities in terms of betting odds, is to reflect on the phenomenon of ambivalence in one's opinion. Permit me to offer a personal example. The issue of scientific realism is one which has vexed me for a number of years. When I reflect on the question of whether electrons (as described by our best contemporary scientific theories) really exist, for example, my opinion exhibits a certain instability. On Mondays, Wednesdays, and Fridays, I reflect on the tremendous predictive and technological successes which have been achieved by modern physicists, I marvel at the electronic images generated by my television picture tube, and I am inclined to assign a very high probability to the proposition that there are electrons. On Tuesdays, Thursdays, and Saturdays, on the other hand, I reflect on the history of science as an unbroken chain of extremely successful and powerful theories which turned out to be wildly wrong in nearly every detail, I remind myself of the arrogance of assuming that I'm living at the end of history, and I am moved to assign a depressingly low probability to the proposition that our current physical theories are even remotely right. One way of representing what's going on in my case would be to maintain that I don't really have an opinion about whether there are electrons. I have two opinions, and a disposition to change my mind a lot. This is the tactic a defender of the probability model would have to take. But this description just doesn't seem true to the facts to me. My vacillation on the electron question doesn't strike me as of a piece with the genuine changes of mind I periodically undergo, such as when I acquire new information or accept the consequences of a novel argument. Rather, it seems to me that both of my stances on the electron issue are partial reflections of a single deeper state of opinion, a state which has remained fairly stable over the last few years, my vacillations notwithstanding. Psychological factors such as the immediate salience of certain facts or arguments may make one facet of that opinion strike me as momentarily compelling, but if I step back from these distorting factors and enter a more self-reflective state, I find myself partially identifying with both judgments simultaneously. In this respect, epistemic ambivalence shares important features with emotional ambivalence, a phenomenon which has been given an insightful treatment by Greenspan (1980). She writes,

> [W]e may be said to have or exhibit a particular emotion (and indeed, I might add, to exhibit it consciously) over a span of time which includes, but is not limited to, the times (supposing there are some) when I am actually experiencing it. Thus, if I waver, over time, between happy and unhappy feelings about my rival's victory (I momentarily feel bad when I first hear the news on the phone, say; but then I immediately consider my friend's good fortune and momentarily feel good), we would reasonably conclude that I have "mixed feelings" throughout the overall time span involved, and not that I am continually changing my mind.

Exactly similarly, I want to suggest that I am at least partially committed to the view that it is quite likely that electrons exist even during the periods of time when I do not find this view immediately compelling. But this suggests that my state of opinion is most adequately represented not by a single probability function but by a set of probability functions, each representing a precise state of opinion with which I am willing to partially identify. At various times, the contingencies of my psychological set may lead me to take some of those precise states of opinion more seriously than others, but it is always possible for me (at least in principle; in practice I often lack the mental discipline) to step outside my momentary frame of mind and recognize the broader range of opinions which I am willing to call my own.

At this point, it's worth making explicit the differences mentioned earlier between the interpretation of the extended probability model I'm floating and the standard interpretation. Most authors who have introduced the idea of sets of probability functions (or similar devices such as interval-valued probability functions or qualitative probability rankings) have done so in an attempt to "humanize" probabilism, to make the model more directly applicable to real human opinion with all its limitations. The idea is that a single probability function which assigns an exact value to every proposition is too complex for an ordinary person to store and manipulate in everyday reasoning. To circumvent this problem, it is suggested that probabilistic opinions might be represented in the form of a reasonable number of probabilistic constraints such as "A has a probability greater than 0.6" or "B has a higher probability than C" or "D and E are probabilistically independent." (In some versions of the story, these constraints are not directly given but are derived from other features of the agent's epistemic state such as her preference rankings.) We can then mathematically represent the agent's opinion as the set of all probability functions which satisfy the constraints; the agent's opinions are just those judgments which all of the functions have in common. This "supervaluational" interpretation has, I think, served to obscure the tremendous gain in representational power which we attain by moving from the standard to the extended probability model. Since the extended model is seen as a weakening rather than a strengthening of the standard model, many authors have tended to continue to reason in terms of single probability functions, tacitly assuming that anything expressible in the idealized standard model can be generalized to the more down-to-earth extended model. This is especially obvious in the literature on full belief, where a number of authors have gone to a good deal of trouble to develop elaborate versions of the extended model, and immediately proceeded to offer analyses of full belief which utilize only the resources of the standard model. On the present interpretation, however, there is no suggestion that the extended model is more realistic or humanized than any other formal epistemic model. A set of probability functions is still an absurdly complex mathematical object, no more amenable to direct storage and manipulation by the feeble human cortex than a single probability function (indeed, probably much less so). The central point is this: even an idealized agent with unlimited computational powers may well adopt a state of opinion which can only be represented by a set of probability functions. Even in situations where such an agent is able to restrict her opinion to a single function, it may not be rational to do so: one may simply be in an evidential position where ambivalence between a number of distributions is the preferable choice. To put it another way, there is no reason to think that my uncertainty about the fair betting odds for Dole's election, or my vacillation on the scientific realism issue, are consequences of my computational limitations. Give me a supercomputer brain, but leave me in the same evidential position, and my probabilistic expression of my opinion is likely to be exactly the same. (On this point, I take it that Levi (1980) is the author whose interpretation is closest to my own.)

We are now ready to return to our central question. Given the newfound resources of the extended probability model, can we give an adequate probabilistic analysis of full belief? My contention is that we can. I will offer my analysis in two stages, defining first *strong beliefs* and then *weak beliefs.* For any probability distribution p, let us define the *corpus* of p to be the intersection of all the propositions to which p assigns probability 1. And for a set S of probability distributions, let us define the *minimal corpus* of S to be union of the corpuses of the probability distributions in S. The minimal corpus of S is the strongest proposition to which every distribution in S assigns probability 1. Finally, define the *strong beliefs* of S to be the supersets of the minimal corpus of S. In other words, the strong beliefs of S are exactly those propositions which receive probability 1 on every distribution in S. The minimal corpus represents S's standard of serious possibility; there is no state of opinion consonant with S which assigns positive probability to any possibility incompatible with the minimal corpus. Were we to end our analysis here, we would thus be faced with all the same challenges which plagued the Received View: our explication would not allow for the possibility of full beliefs whose denials are taken to be seriously possible, hence it would not allow for the possibility of coherent fallibilism. We need to define a notion of *weak belief* to complement the concept of strong belief.

As it turns out, the key concept needed to understand the nature of weak belief is the concept of *presupposition.* Stalnaker (1973) and Lewis (1979) have defended a view of language according to which conversations are governed by the presupposition that certain propositions are true. As Lewis puts it, "At any stage in a well-run conversation, a certain amount is presupposed. The parties to the conversation take it for granted; or at least they purport to, whether sincerely or just 'for the sake of argument.' " This latter point is crucial: in order to participate in a conversation, you need not believe that all of that conversation's presuppositions are true; you may in fact find some of the presupposed material highly doubtful, or even disbelieve it altogether. All that is required is that you participate in the conversation as if you take the relevant presuppositions for granted. This may involve a certain amount of make-believe. Within the framework of probabilism, we can understand presupposition in terms of conditional probabilities: when I participate in a conversation whose total presupposition is B, the judgments I express about a given proposition A will not be governed by my unconditional probability (or probabilities) for A, but rather by the conditional probability (or probabilities) I assign to A given B. More formally: let S be the set of probability distributions which represents my actual state of opinion, and let S_B, the *relativization* of S to B, be the set of probability distributions which contains all and only those distributions q such that for some distribution p in S, $q(A) = p(A|B)$ for all propositions A. Then when I participate in a conversation whose total presupposition is B, I in effect pretend to be a person whose state of opinion is representable by the set S_B. If in fact I do take B for granted, i.e., if B is a strong belief for me, then S and S_B are the same set, and no pretense is necessary. Otherwise, the opinions I report will not necessarily correspond with the opinions I actually hold. (Note: I have talking about presupposition as if it were a purely linguistic phenomenon, but this is largely for convenience of exposition. We often presuppose things in our private deliberations as well as in our public conversations, e.g., when we momentarily adopt a working hypothesis in order to trace its consequences. In either case, the probabilistic analysis is the same: presupposition corresponds to relativization.)

Now, however, I want to distinguish between two importantly different sorts of cases.

Case 1: suppose I am participating in a conversation in which it is presupposed that Phil Gramm will win the 1996 Presidential election. I am asked to express my opinion on the likelihood that a flat federal income tax will be adopted. Now as it happens, I think that the adoption of a flat tax is quite unlikely; all of the distributions in my probability set assign this eventuality a probability well below one-half. At the same time, though, I think that the probability of a flat tax given a Gramm presidency is at least moderately likely. It is the latter conditional probability which governs my assertions in the conversation in question, so in order to participate appropriately in the conversation, I must report opinions with which I do not really identify. (Of course, there is no deception involved here; all parties to the conversation understand that I am honestly reporting my conditional probabilities and not sneakily misrepresenting my unconditional probabilities. The claim here is just that the literal content of what I say, e.g., "A flat tax before the year 2000 is moderately likely," does not accurately report any opinion I actually hold.)

Case 2: now I am participating in a conversation in which it is presupposed that either Bill Clinton or Bob Dole will win the 1996 election. My conditional probabilities for a flat tax given a Clinton or Dole victory are all quite low, so in this conversation I express a fair amount of skepticism about the prospects of a flat tax. But, as I noted before, my unconditional probabilities for a flat tax are also all quite low; so in this case, as opposed to case 1, I am not expressing an opinion which I would not also be willing to express even in the absence of the presupposition in question. Moreover, this is true not just for the flat tax issue, but for any topic which might come up in the conversation. As far as I can tell, there is no probabilistic judgment with which I identify conditional on a Clinton or Dole victory but with which I do not identify unconditionally. So in this second conversation, I am never misrepresenting myself when I speak. Of course, this is not to say that I am free to express the full range of my opinions within this conversation. For example, at least some of the states of opinion with which I identify assign a probability of less than 1 to a Clinton or Dole victory. That is, I do not strongly believe that Clinton or Dole will win; I allow as a serious possibility that another candidate might prevail. But I am not free in the course of the conversation to say, "You know, there's some chance that neither Clinton nor Dole will win"; even though I partially identify with this judgment, the rules of the conversation prevent me from expressing it without upsetting the established presupposition. In general, then, the opinions I may express in this conversation are only a subset of the full range of opinions with which I identify.

I have isolated a difference between the two cases, a difference in the attitudes I bear towards the propositions that Clinton or Dole will win and that Gramm will win. And now I want to claim that this difference just is the difference between belief and unbelief. I believe that Clinton or Dole will win; I do not believe that Gramm will win. My belief that Clinton or Dole will win simply amounts to the fact that when I presuppose that Clinton or Dole will win, I do not thereby move outside of the range of opinions with which I already identify. Presupposing something I already believe doesn't take me anywhere beyond my starting point. Likewise, my failure to believe that Gramm will win simply amounts to the fact that when I presuppose that Gramm will win, I am pretending to identify with states of opinion with which I do not in fact identify. Of course, as already noted, I do not strongly believe that Clinton or Dole will win. On the contrary, there are odds at which I think it would be quite reasonable to bet against this eventuality. I am willing to admit as a serious possibility that I am wrong in my conviction that

Clinton or Dole will win. In short, my belief that one of those two gentlemen will win is a weak belief.

Let us approach this more formally. For a set of probability distributions S and a proposition B, we will say that S is *closed under relativization* to B just in case S_B is a subset of S. And we will define the *maximal corpus* of S to be the intersection of the corpuses of the distributions contained in S. Finally, we introduce the *Closure Condition:* a set S satisfies the Closure Condition just in case S is closed under relativization to all supersets of its maximal corpus. The Closure Condition is not a condition of coherence or any other sort of normative requirement; it is perfectly acceptable to fail to satisfy it. It is, however, a necessary and sufficient condition for the possession of weak beliefs. On the analysis I'm offering, if one's probability set fails to satisfy Closure, then one's beliefs are just one's strong beliefs. In such a case, one is an infallibilist; one takes one's beliefs as one's standard of serious possibility and hence does not acknowledge the possibility that any of one's beliefs are false. If, however, one satisfies Closure, then one's beliefs are all and only the supersets of one's maximal corpus. Those propositions which are supersets of both the maximal and the minimal corpus are strongly believed; one is infallibilist with respect to those beliefs. Those propositions which are supersets of the maximal corpus but proper subsets of the minimal corpus are weakly believed; one is fallibilist with respect to those beliefs.

What are the properties of weak beliefs? First, it is a trivial consequence of the Closure Condition that all weak beliefs receive probability 1 on at least some of the probability measures in the given set. Unlike strong beliefs, however, they do not receive probability 1 from all measures in the set; some measures will assign weakly believed propositions a lower probability. None of the measures, however, will assign a weakly believed proposition a probability of 0. In other words, all of the states of opinion I identify with represent the propositions I weakly believe as being at least seriously possible; and some of the states of opinion I identify with take those weak beliefs to set the standard for serious possibility. This strikes me as roughly the right way to probabilistically characterize my attitude toward most of the propositions I believe. Consider, for example, my belief that Millard Fillmore was born on January 7. When I am in some frames of mind, this proposition strikes me as absolutely, certainly true—this date was in fact Fillmore's birthday, and there are no real grounds for doubt. In other, more skeptical frames of mind, however, I am more inclined to regard the proposition with some degree of doubt—after all, my only evidence for this belief is an encyclopedia entry I read more than ten years ago; my memory has failed me before, and encyclopedias can make mistakes. On the other hand, there is an upper bound to the amount of doubt I am willing to grant this proposition even in my most skeptical moods—neither my memory nor the World Book's editors are *that* unreliable. To make the point another way, let us return to the operational definition of probability in terms of fair betting odds. Once again, suppose that I have two friends who want to bet on the date of Fillmore's birthday and are relying on my judgment as the fairness of the proposed bet. There is no set of odds so high that I would regard them as unfair to the friend who is betting that Fillmore's birthday is January 7: after all, Fillmore's birthday *is* January 7, so even if this friend stakes a fortune against a negligible gain, he's not doing anything irrational so far as I can see. On the other hand, I think there are some odds which are sufficiently high that they are fair to the friend betting against January 7 as well. Once again, my conviction that the proposition is true may well be based on faulty evidence, and taking the long shot that I am wrong may well be worthwhile provided that the potential return is sufficiently sweet. Finally, there are odds which are sufficiently low that they strike me as patently unfair to the friend betting against January 7. If, for example, he's staking $20 to win $1 in case I'm wrong, I think he's making a big mistake, and I would certainly caution him against such a wager. So the explication I'm offering at least seems to get the basic probabilistic judgments underlying belief right.

More importantly, the analysis survives all three of the challenges which scuttled the Received View. Since the analysis does not require that all of my beliefs receive exactly the same probability assignments, it does not require that I place the same amount of epistemic confidence in all of my beliefs. Consider, for example, the proposition that Fillmore was the 13th President and the proposition that Fillmore was President. On all of the probability distributions I identify with, the second proposition receives at least as much probability as the first, and on some of those distributions, the second proposition receives more than the first. So we have a natural way of representing the fact that I am more confident in my belief that Fillmore was President than I am in my belief that he was the 13th President. Second, the present analysis does not imply that rationality requires me always to act as if all my beliefs are true. As was made clear in the Fillmore's birthday example, I think that there are odds at which it is fair to bet that some of my weak beliefs are false. So decision theory does not rule out as irrational any course of action I might take which can be represented as such a bet. It is also true, note, that on the present analysis I am never rationally required to act as if one of my beliefs is false; it is always rational for me to act on the assumption that all my beliefs are true. But I don't think there is any solid reason to think that this represents an inappropriate degree of license. Finally and most importantly, this explication allows me to give coherent expression to my fallibilist convictions: there are judgmental states I identify with according to which the best estimate for the proportion of my beliefs that are true is considerably less than 100%. Of course, I also identify with other judgmental states according to which 100% of my beliefs are true. But again, this strikes me as exactly as it should be. Critics of fallibilism have always suggested that there is something inconsistent about the fallibilist stance: on the one hand, I am maintaining that each proposition I believe is true, and on the other hand, I am maintaining that my beliefs collectively contain some untruths. This is the point behind Makinson's (1965) famous paradox of the preface. I would suggest in response that fallibilism is not an instance of incoherence but simply of ambivalence: in some states of mind, I am certain that these propositions are true; in other states of mind, I am more skeptical about them; and on balance, I recognize both states of mind as rational and appropriate ones with which I am willing to identify.

Finally, let us return to the ur-argument which I offered in support of the Received View. I suggested that all of the available analyses of belief in the literature fail because they fail to engage the strong intuitions which drive the ur-argument. Does my analysis fare any better? I submit to you that it does. In defense of this position, I want to suggest that the first premise of the ur-argument, which states that to believe a proposition P is to maintain that the actual world is in P, contains a subtle ambiguity. Our analysis of ambivalence helps us to see this ambiguity and to repair it. Recall Greenspan's point about emotional ambivalence: I can be said to have an emotion even during those periods when I am not actually in the grip of the conscious experiences associated with that emotion; and so during a period in which my emotional experience vacillates between two different states, I am best described as having both emotions simultaneously throughout the period in question. In a certain important sense, my emotional state transcends the vagaries of my momentary experience. In much the same sense, I want to suggest that belief is a state of opinion which transcends the vagaries of my momentary psychological set. I do not believe a given proposition only during those moments in which I hold it with firm and absolute conviction, eschewing as patently false all alternatives; I can continue to believe a proposition even as I cycle through moments of certainty and doubt, of trust and skepticism. Consider the game show example once again: if I decide not to bet my entire bankroll on my ability to state Millard Fillmore's birthday correctly, this doesn't mean that I have suddenly ceased to believe that Fillmore was born on January 7. It merely means that the salient features of my immediate circumstances make me find the skeptical facets of my underlying opinion more compelling than the dogmatic facets. Premise 1 of the ur-argument obscures this distinction by employing the ambiguous concept of "maintaining": I can maintain that something is true in the strong sense of being immediately gripped by the judgment that it is true or in the weaker sense that this judgment is represented in my underlying state of opinion. Suppose we restate Premise 1 to make clear that the weaker sense is intended, something like:

(P1*) To believe a proposition P is to identify with the judgment that the actual world is in P.

If we rephrase the other premises in the obvious way, we arrive at the revised conclusion:

(C*) To believe a proposition P is to identify with the judgment that the probability of P is 1.

But on the explication I am offering, this conclusion is satisfied. I do identify with the judgment that a given proposition I weakly believe has a probability of 1. Of course, I also identify with other probabilistic judgments about this proposition; and at various times I may find myself more moved by these judgments than by the more confident assignments of probability 1. But I don't think any of the intuitions which drive the ur-argument provide us with a reason to reject the analysis on these grounds.

So that's the basic story I'm pitching. Clearly, there's a lot more work to be done. The formal properties of weak beliefs need to be explored in much greater detail. The phenomenon of judgmental ambivalence needs to be given a much more careful treatment than I've managed here. Something needs to be said about the complex relationship between the range of judgments I identify with and the range of judgments I express in a given conversational context; the simple story I've told about presupposition barely scratches the surface. And there are lots of interesting directions in which the account might be expanded; I am particularly interested in coming up with something to say about the dynamics of weak belief, perhaps devising a generalization of conditionalization which represents the appropriate constraints on adding a weak belief to one's corpus. I also think there may turn out to be interesting relationships between the analysis of belief and the analysis of the indicative conditional; an account of what it is to accept such a conditional will, I think, resemble my account of what it is to accept a proposition. I hope to address most of these issues in the larger work of which this paper is a preliminary survey. But I hope what I have said here has been sufficient to convey at least the flavor of the problem and a sense of the direction in which I take the correct solution to lie.

References

Carnap, Rudolf (1947). *Meaning and Necessity.* Chicago: University of Chicago Press.

Cherniak, Christopher (1986). *Minimal Rationality.* MITY: Cambridge, Mass.

de Finetti, Bruno (1974). *Theory of Probability,* vol. I. New York: Wiley .

Harman, Gilbert (1986). *Change in View.* Cambridge, Mass: MIT.

Jeffrey, Richard (1970). "Dracula Meets Wolfman: Acceptrance vs. Partial Belief" in *Induction, Acceptance, and Rational Belief,* Marshall Swain (ed.), pp. 157-185. Reidel: Dordrecht, 1970

Jeffrey, Richard (1992). *Probability and the Art of Judgment.* Cambridge: Cambridge University Press.

Levi, Isaac (1980). *The Enterprise of Knowledge.* Cambridge: MIT Press.

Lewis, David (1983). *Philosophical Papers,* vol. I. New York: Oxford University Press.

Ramsey, Frank (1990). *Philosophical Papers.* Cambridge: Cambridge University Press.