Is it rational to have faith?

Looking for new evidence, Good's Theorem, and Risk Aversion

0. Introduction

Although statements involving ascriptions of faith are quite common, faith and rationality are often thought to conflict; at best, they are thought to have nothing to say to each other. In this paper, I analyze the concept of faith: in particular, I give an account that unifies statements of faith in mundane matters and statements of religious faith. My account focuses on spelling out the sense in which faith requires going beyond the evidence, and I argue that faith requires stopping one's search for further evidence. I then turn to the question of whether it is rational to have faith, taking into account both epistemic rationality and practical rationality. I argue that faith need not conflict with epistemic rationality at all; however, whether faith can be practically rational hinges on which attitudes towards risk it is rationally permissible to adopt when making decisions: more specifically, it hinges on how one should respond to the risk of misleading evidence. I have argued elsewhere that practical rationality permits a wider range of attitudes towards risk than is commonly supposed, and if this is right, then it is also rationally permissible to have faith; indeed, depending on the attitude towards risk that one decides to adopt, it is sometimes rationally required.[1]

1. Faith Statements

I begin by sketching some characteristics that faith statements have in common. By "faith statements," I simply mean statements involving the term "faith." So, the following are all typical examples of faith statements: *I have faith in your discretion*; *he has faith that his spouse won't cheat on him*; *I have faith in you*; *she has faith that her car will start*; *he has faith that you won't reveal his secret*; *she acted on faith*; *it was an act of faith*; and, of course, *I have faith that God exists*, or simply *I have faith in God*.

The first thing to notice is that faith statements typically involve a proposition to which the actor involved acquiesces. This is obvious in the case of "faith that" statements: when a

person has faith that *p*, he acquiesces to *p*. It is equally clear in the case of statements whose structure indicates that they can be easily translated into "faith that" statements: for example, the statement "I have faith in your discretion" is equivalent to "I have faith that you will be discreet." It is less obvious in the case of statements claiming an individual has faith in a person; however, upon further inspection, having faith in a person typically consists in acquiescing to particular propositions about the person: for example, having faith in a person might involve acquiescing to the claim that the person will do the right thing, or will succeed at a task, depending on the context, and having faith in God might involve acquiescing to (at least) the claims that God exists and that God is good. By the same token, performing an act of faith or acting on faith seems to involve acquiescing to a proposition; which proposition one acquiesces to will be set by the context: for example, if setting down one's own weapons will be an act of faith, then this is because setting down one's own weapons involves acquiescence to the claim that the other person will then set down his.

The second thing to notice about faith statements is that the truth or falsity of the proposition(s) involved is ordinarily a matter of importance to the actor. For example, it does not seem apt to state that I have faith that the Nile is the longest river in Egypt, because I do not care whether or not this is true (I might care, but in that case it could be an apt statement to make). On the contrary, it appears that we cannot attribute faith to a person unless the truth or falsity of the proposition involved makes a difference to that person: *that my spouse won't cheat on me* or *that my car will start* are the kinds of things I might consider whether to have faith in precisely because it makes a difference to me whether or not these things are true.

Along the same lines, having faith typically involves an action: a person's having faith in something should make a difference to her behavior. As St. James famously wrote, "Faith without works is dead"; that is, if one has faith, then it will be manifest by a particular action. Unsurprisingly, then, whether one has faith is context-dependent: for example, I might have faith that my car will start in the context of simply driving to work, but I might not have faith that my car will start in the context of being relied upon to carry a life-saving organ to the hospital (as evidenced by the fact that I may double-check my engine or arrange for a backup mode of transportation in the latter case but not the former). A person might have faith in God when it comes to giving weekly donations to the poor, but might lack faith in God when it comes to allowing himself to be martyred. So one might have faith as relates to the performance of some actions, but not others.

Third, we make assertions of faith only when the outcome of the proposition involved is yet uncertain, or when the evidence we have does not yield complete certainty. For example,

when a friend is worried about the outcome of an exam, we might reassure her by saying "I have faith that you passed"; however, once she shows us that she got an A, we would no longer say "I have faith that you passed." Clearly, this is not because we are less willing to acquiesce to the claim that she passed, but because the matter is now settled. Similarly, it seems odd to claim faith in logical truths. "I have faith that 2+2 = 4" appears unnatural: because we are certain that 2+2 = 4, we don't need faith to "fill the gap," so to speak.

When a person has faith, he does not have conclusive evidence – evidence yielding certainty – in the proposition involved. Are there further restrictions on the evidence he has for a proposition that he can aptly be described as having faith in? I claim there are not. Indeed, a person may have no evidence at all for the proposition he has faith in, or evidence that tells *against* the proposition. For example, we could imagine someone saying "Although she's spilled all the secrets I've told her so far, I have faith that this time will be different," or "I don't think there's any evidence that God exists, but I have faith that he does." Therefore, it seems that the person's evidence can support the proposition to any degree, as long as it is not conclusive: that a person has faith that X implies nothing about his evidence for X, aside from its inconclusiveness. However, although statements in which the actor has faith despite no or contrary evidence do seem correctly described as cases of faith, they are not cases in which we are inclined to think that the actor is *correct* to have faith; rather, we think his faith is misplaced. We will later see that my account does justice to this intuition: faith turns out to be rational only if one's evidence supports the proposition in question to at least a particular degree – a degree that will depend on the context.

Finally, having faith seems to involve going *beyond* the evidence in some way (as I've pointed out, this is why we cannot have faith in a proposition for which we have conclusive evidence). Section three will discuss the manner in which one must go beyond the evidence in order to count as having faith.

We are now in a position to give a formal analysis/explication of faith, aside from spelling out the sense in which faith requires going beyond the evidence. As I've made clear, the term "faith" appears in many different grammatical constructions: you might have faith in a *person*, you might have faith in a *proposition*, you might perform an *act of faith*, or you might *act on faith*. We want to come up with an account that makes sense of all of these uses of the term. I've already pointed out that faith typically involves a *proposition* as well as an *action* that the truth or falsity of the proposition makes a difference to. What I propose, then, is to make "faith that *X*, expressed by *A*" the basic unit of analysis, where X is a proposition and *A* is an act, and define the other constructions in terms of this one.

It is important to build into our analysis the way in which the proposition and the act are related. I pointed out that a person can have faith in X only if he cares whether X is true or false, and presumably this is because the act performed is good if X is true and bad if X is false. So, as a first pass, we might say:

> A person has faith that X, expressed by *A*, only if that person performs act *A*, and A is better if
> X is true than it is if X is false.

However, this isn't quite enough to guarantee that *A* is an act that expresses faith in X: A might be better if X is true than it is if X is false, but one might do A for the sole reason that *A* is the only act available. Or one might do *A* because *A* is better than all the alternatives regardless of whether or not X is true. Walking to the store is much more enjoyable if it is sunny, but if I need to go to the store and don't have a car, I will walk to the store regardless of how likely it is to rain. In order to be an expression of faith in X, *A* must be based on X or depend on X in some way. So a more accurate characterization of the way the proposition and the act are related is:

> A person has faith that X, expressed by *A*, only if that person performs act *A*, and *A* is better if
> X is true than it is if X is false, and there is some alternative act *B* that is available to the
> person such that *B* is better than *A* if X is false.

In other words, when a person has faith, she takes a risk that will turn out well if and only if X turns out to be true.

Of course, this is not yet the whole story. But it does give us a way to identify what it is to have faith in a person, and to perform an act of faith, in terms of having faith that X, expressed by *A*:

> A person p has faith in another person q if and only if there is some act *A* and some
> proposition(s) X that is (are) *about* q such that p has faith in X, expressed by *A*.

So Bob might have faith in Mary because he has faith that Mary won't reveal a secret he tells her, expressed by the act of telling her his secret. Paul might have faith in God because he has faith that God exists and that God is good, expressed by the act of refraining from lying because he believes God has forbidden lying.

We can also identify what it is to perform an act of faith, or to act on faith:

> A person performs an act of faith (or acts on faith) if and only if he performs some act *A* such that there
> is a proposition X such that he has faith in X, expressed by *A*.[2]

---

[2] Another grammatical construction that involves "faith" is the claim that someone is faithful. I will not analyze this construction here; I suspect that is it used either in a way that is definable in terms of the other constructions, e.g., a person is faithful in performing a particular act because it is an act of faith, or to indicate a disposition, e.g., a person is a faithful person because they have faith in many things or are inclined to have faith.

With these preliminaries in mind, the rest of this essay will be devoted to spelling out what else faith in X, expressed by *A*, requires; and when (if at all) it is rational to have such faith. First, let me briefly explain what I take rationality to consist in; only then will we be in a position to see whether the requirements of faith conflict with the requirements of rationality.

2. Rationality

For our purposes, rationality has two components: epistemic rationality and practical rationality. I will begin with epistemic rationality.

I will assume a broadly evidentialist conception of epistemic rationality: one should proportion one's beliefs to one's evidence. One should not, for example, simply believe what one likes or believe what would make one happy; more generally, one should not take non-truth-conducive reasons as reasons for belief. I will also assume a subjective Bayesian account of partial belief: beliefs come in degrees; degrees of belief obey the probability calculus; and two people can (rationally) have different degrees of belief in a proposition if and only if they have different evidence or different priors.[3] An important feature of this account for our purposes are that a rational person can only change his degrees of belief in response to evidence – and *must* change his degrees of belief in response to new evidence. Furthermore, I will assume that rational agents are *non-dogmatic*: the only propositions a person can believe to degree 1 are logical or analytic truths and possibly one's evidence itself.[4] The justification for the latter feature is that if you assign degree of belief 1 to a proposition X ($p(X) = 1$), then there is no piece of evidence that could make you lower your degree of belief in X: for every piece of evidence E, it is either the case that $p(X \mid E) = 1$ or $p(E) = 0$. But, presumably, we think that any proposition, as long as it's not a logical or analytic truth – or a proposition one could not possibly be wrong about, as evidence, appropriately described, may be – should be open to falsification by further evidence; that is, we can never get so much evidence for something that no further evidence could tell against it.

---

[3] Depending on how one defines evidentialism, it might be strictly speaking incompatible with subjective Bayesianism, since the latter allows any set of prior degrees of belief to count as rational, so that two people could share a body of evidence and still (rationally) have different degrees of belief. On the other hand, a natural version of evidentialism implies that there is only one rational way to respond to a given body of evidence. We could alleviate this problem by pointing out that rationality is a feature of belief *change*, or that evidentialism and subjective Bayesianism agree about what to do once the priors have been set. Thomas Kelly makes this point in his entry on "Evidence" in the *Stanford Encyclopedia of Philosophy*.
[4] We can relax the assumption that evidence gets assigned probability 1, as Jeffrey does in introducing Jeffrey conditionalization, but since it does not make a difference to my argument, I will make the assumption for simplicity.

To count as epistemically rational, you must proportion your beliefs to the evidence, and you cannot claim certainty about anything except logical or analytic truths or propositions you could not possibly be wrong about. Practical rationality, on the other hand, involves taking the means to one's ends. It is this notion of rationality that decision theory attempts to spell out. Standard decision theory (expected utility theory) claims that rational agents maximize expected utility: that is, they prefer the act with the highest mathematical expectation of utility, relative to their utility and probability functions. So if we think of an act as a gamble that yields a particular outcome in a particular state of the world – e.g. $\{O_1, E_1; O_2, E_2; \dots ; O_n, E_n\}$ is the act ("Savage act") that yields O_i if E_i is true, for each i – then the value of this act is $\sum_{i=1}^{n} p(E_i)u(O_i)$

I am sympathetic to the general aim of decision theory, and I think that expected utility theory is largely correct in its analysis of practical rationality. However, I think that expected utility theory employs too narrow a criterion of rationality: in particular, because it incorrectly analyses the phenomenon of risk aversion, it unnecessarily limits the attitudes decision makers can take towards risk. I prefer an alternative decision theory that is a generalization of expected utility theory.[5] I will briefly describe my theory and the motivation for it, though I will not argue that it is the correct account of rationality here.

An agent displays risk averse behavior just in case she prefers receiving a set of goods for certain to taking a deal that yields those goods on average (e.g., she prefers $50 to a coin flip between $0 and $100). Expected utility theory accounts for these preferences by claiming that agents who are risk averse must not value goods linearly: for example, the difference in value between $0 and $50 must be greater than the difference in value between $50 and $100. Thus, preferences that are risk averse with respect to money imply, on EU theory, a diminishing marginal utility function. More generally, EU theory explains risk aversion as a particular way of valuing *outcomes*.

By contrast, we could instead explain risk aversion as *sensitivity to global properties of gambles*. The global sensitivity explanation says that even holding fixed how much one likes various outcomes, and thus holding fixed the average value of a gamble, which outcomes constitute the gamble and how they are arranged matter: for example, an agent might care about the minimum value of a gamble, or the maximum, or the variance. That is, properties of gambles that do not supervene on any particular outcome or chance of getting that outcome – global

---

[5] For a more complete explanation of my preferred theory (and its relationship to other non-expected utility theories, as well as arguments that it, not expected utility theory, is the correct theory of rationality, see my "Risk and Rationality," manuscript, 2009.

properties of gambles – matter. This does not rule out marginal diminishment in the value of goods. People who have diminishing marginal utility functions can still be sensitive to global properties; I suspect that most people's utility functions do diminish marginally for very large quantities of money, and yet that many people are globally sensitive. So, whereas the expected utility theorist claims that risk averse behavior always entails a diminishing marginal utility function, I claim that it might indicate global sensitivity instead (or some combination of the two).

In expected utility theory, utility and probability functions vary from agent to agent, but utilities and probabilities interact in a set way to yield the value of a gamble. As mentioned above, the utility of a gamble is its mathematical expectation: the utility of each possible outcome weighted by the probability of obtaining that outcome. So, the utility of a gamble between A and B, where the agent gets A if event E obtains and B if event ~E obtains, is $p(E)(u(A)) + p(\sim E)(u(B))$. Assuming an agent's degrees of belief obey the probability calculus, this is equivalent to $p(E)(u(A)) + (1 - p(E))(u(B))$, which is in turn equivalent to $u(B) + p[u(A) - u(B)]$. Taking B to be the less (or equally) desirable option, this latter formulation merely says that the value of a gamble will be the minimum value it guarantees plus the amount by which the agent might do better, weighted by the probability of doing that much better.

We can generalize this equation to allow for different ways of aggregating possible outcomes. In particular, we could interpret agents who are sensitive to global properties as having preferences that accord with a *weighted* expected utility function: the desirability of the above gamble, assuming A is at least as good as B, will be $u(B) + \mathbf{r(p(E))}[u(A) - u(B)]$, where r is the agent's "risk function" or "weighting function," adhering to the constraints $r(0) = 0$, $r(1) = 1$, r is non-decreasing, and $0 \leq r(p) \leq 1$ for all p.[6] In effect, the interval by which the agent might improve her lot above what she is guaranteed to get shrinks not by her chance of getting the better prize, but by a *function* of this chance, which reflects her attitude towards risk. Thus, the value of

---

[6] My theory is closely related to two existing theories in the literature, Schmeidler's Choquet expected utility (described in Hong and Wakker (1996) and Kobberling and Wakker (2003)), and Quiggin's anticipated utility (described by Quiggin (1982); later called rank-dependent utility or cumulative utility) – and can be make equivalent to these theories, each under a particular assumption. Like these theories, my theory employs a weighting function that is sensitive to the ordering of outcomes. However, Choquet utility employs a weighting function of states, not of *probabilities* of states – i.e. it does not directly include an agent's judgments about probability at all. And anticipated utility uses an "objective" probability function, and imposes additional constraints. For more on the relationship between my theory and other non-expected utility theories, see my "Risk and Rationality." Sources: Hong, Chew Soo and Peter Wakker (1996). "The Comonotonic Sure-Thing Principle." *Journal of Risk and Uncertainty* 12, pg. 5-27. Kobberling, Veronika and Peter P. Wakker (2003). "Preference Foundations for Nonexpected Utility: A Generalized and Simplified Technique." *Mathematics of Operations Research* 28:3, pg. 395-423. Quiggin, John (1982). "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization* 3, pg. 323-343.

a gamble will be the minimum value the gamble guarantees plus the amount by which the agent might do better, weighted by this function of the probability of doing that much better. If this function has a low value, then any improvement over the minimum will be heavily discounted; thus, the minimum will weigh heavily in the evaluation of the gamble. If it has a high value, then any improvement over the minimum will be amplified; thus, the maximum will weigh heavily. If r(p) = p, then the agent in an ordinary expected utility maximizer. On this theory, an agent is risk-averse in utiles – that is, he exhibits risk-averse behavior even with a linear utility function – just in case he has a convex r-function. I call such an agent *risk-avoidant* (to preserve the neutrality of the term risk aversion). One example of such a function is r(p) = p^2, and I will use this function in my example later in this paper. Since this theory includes a weighting function that reflects the agent's attitude towards risk, I call it **risk-weighted expected utility theory.**[7] My view is that it is rationally permissible to adopt any risk function, as long as it adheres to the constraints mentioned above.

There is a natural generalization of this theory to gambles over multiple options. I noted above the expected utility of an act $\{O_1, E_1; O_2, E_2; \dots ; O_n, E_n\}$. This is equivalent to:

$$u(O_1) + (\sum_{i=2}^{n} p(E_i))(u(O_2) - u(O_1)) + (\sum_{i=3}^{n} p(E_i))(u(O_3) - u(O_2)) + \dots + p(E_n)(u(O_n) - u(O_{n-1}))$$

If $u(O_1) \leq \dots \leq u(O_n)$, this equation states that an expected utility maximizer will consider the utility of the worst possible outcome, consider how much he stands to gain over the worst outcome if he gets the next-worst outcome, and weights that amount by the probability of getting an outcome at least as good as that (second-worst) outcome; then he will consider how much he stands to gain over the second-worst outcome if he gets the next-worst outcome, and weights that amount by the probability of getting an outcome at least as good as that (third-worst) outcome; and so forth. Again, my theory will generalize this equation to make the amount by which a possible improvement is weighted a *function* of the probability of improving. So the *risk-weighted* expected utility of an act $\{O_1, E_1; O_2, E_2; \dots ; O_n, E_n\}$, where $u(O_1) \leq \dots \leq u(O_n)$, will be

$$u(O_1) + r(\sum_{i=2}^{n} p(E_i))(u(O_2) - u(O_1)) + r(\sum_{i=3}^{n} p(E_i))(u(O_3) - u(O_2)) + \dots + r(p(E_n))(u(O_n) - u(O_{n-1}))$$

---

[7] Let me briefly say something about why *this* modification of the standard theory, rather than a different modification that takes account of global properties. First, I want to remain within a consequentialist framework – a framework that says that only the outcomes matter – while allowing (unlike the standard theorist) that there are different ways of aggregating possible consequences. Second, I find my proposal to accord with how I intuitively think about risk, and experimental research shows that I am not alone: Kahneman, Daniel and Amos Tverksy (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47, pg. 263-291. Finally, my theory follows from a set of axioms that are themselves intuitive constraints on rational preference – more intuitive, I claim, than those of the standard theory.

Since I will not argue for my theory here, I do not expect the reader to endorse it wholeheartedly as the correct theory of rationality. However, what I conclude about faith will be of interest even to those who are skeptical about risk aversion: it will turn out that faith can be rationally permissible only if it is rationally permissible to be risk averse in the way I suggest. Therefore, the fate of risk aversion and that of faith are linked: each is rationally permissible to adopt if and only if the other is. It is also worth noting that others have proposed non-expected utility theories to account for the way people actually behave – though as far as I know, few have endorsed these as accounts of *rational* preferences – so my argument here will also be of interest to those trying to understand what might motivate actual agents.

Before I turn to the main analysis, it is worth clearing up two worries about rationality in the senses I've sketched. The first worry is that they might conflict. [[Note: In the next draft, I will briefly explain this worry]]

The second worry is that these criteria of rationality might not be strong enough. My notion of epistemic rationality seems to permit adopting any prior degrees of belief, including any conditional priors: priors about the relationship between particular hypotheses and particular pieces of evidence. For example, it seems that on my account, a person may rationally believe he has been abducted by aliens, as long as he also believes that the evidence he has supports this to the degree that he believes it. Similarly, my notion of practical rationality seems to permit adopting any set of values, including, as Hume famously referred to, a preference for the destruction of the whole world over the scratching of one's finger. Both epistemic rationality and practical rationality, as I define them here, are notions of *consistency*: the only restriction on degrees of belief is that they are consistent with one another, and the only restriction on preferences is that they are consistent with one another, given one's degrees of belief about the likelihood of each possible state of the world and one's level of risk aversion. However, there is another notion of rationality, which rules out believing one has been abducted by aliens and rules out preferring the destruction of the world to the scratching of one's finger. We might call this notion *reasonableness*.

I cannot fully respond to this worry in depth here. But it is important that these two notions of rationality come apart quite readily. Consistency restrictions are structural: they rule out particular patterns of belief or desire, regardless of the content of these attitudes. On the other hand, reasonableness restrictions are substantial: they rule out particular beliefs or desires,

regardless of which other beliefs or desires one has.[8] Therefore, we can talk about what they require separately. Furthermore, since the consistency sense of rationality exhausts the subjective sense of rationality, my analysis fully answers the question of whether it can ever be rational *from an agent's own point of view* to have faith. I will show in this paper that it can be subjectively rational to have faith: that is, that there are some situations in which it is rational (in the consistency sense) to have faith. If we want to answer the further question of whether it is reasonable to have faith – that is, of whether a person has objective reasons not to have faith – we can address this separately. I should note, though, that nothing in my argument relies on the content of the propositions believed or desired, so the question whether faith is reasonable (if I've convinced the reader that it is rational) can be answered by asking whether there are any contents for which is it reasonable to have the patterns of belief and desire presented in my examples.

3. Going Beyond the Evidence

So far, we have as a necessary condition of faith:

A person has faith that *x*, expressed by *A*, only if that person performs act *A*, and *A* is better if *x* is true than it is if *x* is false, and there is some alternative act *B* that is available to the person such that *B* is better than *A* if *x* is false.

However, we haven't yet spelled out an important component of faith: that faith requires going *beyond* the evidence in some sense. This section will be devoted to making sense of this requirement. I will consider several initially plausible ways to make sense of this requirement; however, I argue that all of these attempts fail to reveal a genuine requirement of faith.

A very naïve reading of faith takes faith in X[9] to require believing X to a higher degree than one thinks is warranted by the evidence. Under this reading, faith has a Moorean feel to it, assuming that thinking X is warranted by the evidence to a particular degree is equivalent to something along the lines of believing that *X* is likely, to that degree, to be true. For this reason, it is hard to see that this could be a plausible requirement of faith – not because we are starting

---

[8] There is an analogy here with the rationality of full belief. There are important consistency conditions that need to be in place for an agent to count as a rational believer –namely, her beliefs need to obey the laws of logic. There may be additional constraints beyond these structural constraints, but if her beliefs do not obey the laws of logic – e.g. if she believes *p*, *q*, and *if p then ~q* – then surely she is irrational. And the question of what the consistency constraints are (and what follows from them) is both important in its own right and analyzable apart from questions about what (if any) substantive constraints there are on rational belief.

[9] I mentioned that I take myself to be analyzing *faith in X expressed by A*. In the first several proposals for the sense in which faith requires going beyond the evidence, the act does not figure into the proposal. Therefore, for readability, I will say "faith in X" when I really mean "faith in X expressed by *A*, for some particular *A*."

from the presumption that faith must be rational, but because it seems hard to imagine someone actually having faith in this sense, and yet having faith seems to be a fairly common occurrence.

The first serious candidate for an analysis of the sense in which faith goes beyond the evidence is that faith in X requires being determined not to abandon one's belief in X no matter what evidence one encounters. This is the sense adopted by Kierkegaard in his Concluding Unscientific Postscript to Philosophical Fragments, according to Robert Adams.[10] Adams explains one of Kierkegaard's arguments that religious faith cannot be based on objective inquiry, which Adams calls the "postponement argument." According to Adams, Kierkegaard argues that religious commitment requires faith in particular historical claims, such as whether Jesus declared Peter and his successors to be infallible, but that objective inquiry can never yield certainty about historical propositions; therefore, Kierkegaard concludes, anyone who seeks to base his religious commitment on objective inquiry postpones his faith forever. Although Adams does not mention it, the same thing could be said of basing one's religious commitment on objective inquiry into the truth of scientific claims.

On Adams's interpretation of Kierkegaard, authentic religious faith requires total commitment to particular beliefs, and total commitment to a belief requires commitment to not revise it in the future. In Bayesian terms, this means that having faith in a proposition X requires having $p(X) = 1$. However, historical (or scientific) inquiry always leaves room for doubt – so historical inquiry can never yield $p(X) = 1$. In our terms, since X is contingent, it would never be epistemically rational to have $p(X) = 1$: we should always be open to the falsification of X by further evidence. So if having faith really does require being determined not to abandon one's belief come what may, then there really is a conflict between faith and epistemic rationality.

Luckily from the point of view of someone who hopes faith and rationality are compatible, there are three serious objections to the Kierkegaardian view that faith in X requires being determined not to abandon one's belief in X under any circumstances, or that faith in X requires having $p(X) = 1$. Adams himself thinks that this is too strong of a requirement, and the first objection I'll present is one that he poses. Adams points out that this requirement does not seem to be something that religious ethics would endorse (when, clearly, faith is something that it does endorse): "It has commonly been thought to be an important part of religious ethics that one ought to be humble, teachable, open to correction, new inspiration, and growth of insight, even

---

[10] Adams, Robert. "Kierkegaard's Arguments against Objective Reasoning in Religion," *The Monist*, Vol. LX, No. 2 (April, 1976), 228-243. Adams cites Soren Kierkegaard, Concluding Unscientific Postscript. trans. David F. Swenson; introduction, notes, and completion of translation by Walter Lowrie. (Princeton, N.J.: Princeton University Press, 1941).

(and perhaps especially) in important religious beliefs."[11] Nor do the ethics involved in friendship seem to require that we remain determined not to abandon our belief in a friend's trustworthiness come what may, even if these ethics do require having faith in our friends.

The second objection to this requirement is that it distorts the phenomenology of acting on faith. Anyone who is acting on faith feels like she is taking a risk of some sort: again, the act A that you are doing on faith (in X) is supposed to be better than the alternatives if X is true and worse than the alternatives if X is false. But if one is certain that X is true, then doing A is not a risk at all! On the contrary, A is simply an act that, from your point of view, will undoubtedly turn out well. It is like the act of taking a bet on which you win $100 if water is H_2O and lose $100 if it is not. One might reply that from an objective standpoint, doing A is a risk; but even if that is the case, the view still fails to explain the phenomenology of acts of faith, since they feel risky even from an internal perspective. Alternatively, one might say that the risk involved is not in the act itself, but in fully believing something that you think might not be true. However, in addition to the Moorean worries mentioned above, I suspect the phenomenology in believing fully something you think might not be true is not one of taking a risk but of performing mental gymnastics – and an act of faith seems like something one could do with a perfectly clear head.

Finally, this requirement seems incompatible with the phenomena of *losing* faith. If faith requires total commitment, come what may, then if a person ceases to have faith, it must be the case that she was never fully committed in the first place; therefore, it must be the case that she never truly had faith. But that doesn't seem correct: on the contrary, it seems that a person can have an authentic faith at some point, and then lose her faith at a later point. Indeed, this seems particularly obvious in the case in which a person who has faith in X is faced with undeniable evidence against X. For example, it seems perfectly consistent with my having faith that a friend will keep my secret that I would lose that faith if I came across him revealing my secret to someone else. It seems consistent with Jesus's disciples having faith that the man in front of them is indeed Jesus resurrected in the flesh, that if their hands were to simply pass through he body they would reconsider their faith. Surely faith does not require stubbornly affirming X if you have a perceptual experience whose content is ~X.

One might argue that in these cases, losing faith seems consistent with having faith in the first place only because the actor loses faith in response to being in a situation he previously thought was epistemically impossible. This suggests that we might modify the original view to make it more plausible. We might accept that faith in X does not require being determined to stick to one's belief in X under any circumstances, or does not require setting p(X) = 1. But, we

---

[11] Adams, pg. 6.

might claim, faith does require disregarding the possibility of encountering evidence against X – even if one wouldn't stick to one's belief if, per impossible, one encountered such evidence. That is, faith requires treating the circumstances under which one would abandon one's beliefs as epistemically impossible. Indeed, Adams suggests this possibility.[12] This would explain how someone could have faith and then lose faith: if a person finds himself in a situation he previously thought impossible, he can no longer think that situation is impossible (that is, he can no longer have faith) – but that does not mean he consider it possible before, just that he erred in considering it impossible.

However, having faith at the present time seems not only compatible with losing one's faith at a later time, but also compatible with *realizing at the present time* that one *might* lose one's faith at a later time. It seems that I could realize ahead of time that in the event that I come across my friend revealing my secret, I would lose faith that he keeps my secret – and it seems that I could realize that this event is possible (if extremely unlikely) without this realization detracting from my present faith. That is, it seems that I could countenance the possibility of losing faith in my friend's discretion without thereby losing that faith.

Furthermore, this 'modified' view is actually equivalent to the original view, for subjective Bayesian agents. If every piece of evidence that would lower one's degree of belief in *x* is treated as epistemically impossible (that is, if for all E such that $p(X \mid E) < p(X)$, we have $p(E) = 0$), then one must assign a degree of belief 1 to X.[13] So it is not really a weaker view, after all.

What the failure of these initial analyses seems to suggest is that the sense in which faith requires going beyond the evidence isn't one in which the person with faith must believe something different than her evidence suggests (e.g., have more certainty that is warranted). This is good news for the person who hopes that faith and rationality are compatible, since believing something different than one's evidence suggests is straightforwardly epistemically irrational, as I've defined epistemic rationality. It seems possible for two people to have the same evidence for X, and the same degree of belief in X, such that one has faith in X expressed by some particular act *A* and the other does not have faith in X expressed by *A*. Consider the following case: Ann and Ben have both been friends with Dan for a long time. They have the exact same evidence about Dan's secret-keeping abilities, and they both have p(Dan will keep a secret) = 0.9. However, Ann decides to trust Dan with her secret, and Ben does not. It seems as if Ann has faith that Dan will keep her secret, whereas Ben does not have faith that Dan will keep his. Furthermore, if Chris has different evidence and (rationally) sets p(Dan will keep a secret) = 0.1,

---

[12] Ibid.
[13] Assuming an atomless algebra.

and tells Dan his secret nonetheless, it seems that he may also have faith that Dan will keep his secret. What this shows is that having faith is not entailed by having particular degrees of belief or a particular relationship between ones evidence and one's degrees of belief – nor does it require some particular degree of belief or evidential relationship.

So perhaps the sense in which faith requires going beyond the evidence is this: having faith in *X* expressed by *A* requires *acting as if* X is true. Ann performs the same action she would perform if she *knew* that X were true, and therefore has faith. However, this is too quick: although having faith does require acting as if X is true – that is, performing the action that one would perform if one were certain that X is true – this is not a sufficient condition for faith. Indeed, all rational agents will act as they would if they knew X were true, once their degree of belief reaches a certain threshold (which depends on the utilities of the various outcomes) – and yet not all rational action can be appropriately described as exhibiting faith. For example, a gambler might pay $1 for a chance to win $100 if a card randomly drawn from a deck is the three of hearts (probability 1/52) and thus perform the act that he would perform if he knew for sure that he would draw the three. This act is a risk – it stands to turn out very well if the three comes up, and not so well otherwise. But it does not seem appropriate to say that the poker player has faith that the three of hearts will come.

So while the first three conditions were each too strong to be requirements of faith, this condition is too weak: we need a further sense in which a person who has faith must go beyond the evidence.

Earlier, I mentioned Kierkegaard's "postponement" argument: that anyone who seeks to base his religious faith on objective inquiry postpones his faith forever. For example, let's say a person is deciding whether or not to have faith in the claim that Jesus declared Peter and his successors to be infallible. If she tries to decide by engaging in historical inquiry until she is certain that this claim is true, then she cannot have faith until the inquiry is finished – but it will never be finished. There is something insightful in examples like this: it does seem that while a person is engaged in an inquiry to determine whether X is true, one cannot yet have faith that X. As above, we might try explain this insight by claiming that having faith in a proposition requires taking that proposition to be certainly true, and engaging in such an inquiry will never yield complete certainty. However, another alternative is that one cannot get to faith by engaging in an inquiry until faith results not because an inquiry can never yield the certainty required for faith, but rather because engaging in an inquiry itself *constitutes* a lack of faith. In other words, **faith requires not engaging in an inquiry whose only purpose is to figure out the truth of a proposition one purportedly has faith in**. Faith is incompatible with performing an inquiry

until one arrives at certainty because faith requires *a decision to cease one's inquiry* (and to perform the act one would perform if *x* were true), and if one waits until one is certain to cease one's inquiry, one's inquiry will never cease. So one will never have faith, but not because one failed to arrive at certainty, *but because one set certainty as a precondition for arriving at faith –* that is, for ceasing one's inquiry.

As an example, if a man has faith that his spouse isn't cheating, this seems to rule out his hiring a private investigator, or opening her mail, or even striking up a conversation with her boss to check that she really was working late last night – that is, it rules out conducting an inquiry to verify that his spouse isn't cheating. If he does any of these things, then she can rightfully complain that he didn't have faith in her, *even if* she realizes that, given his evidence, he should not assign degree of belief 1 to her constancy. To use a religious example, when so-called "doubting" Thomas asks to put his hand in Jesus's side to verify that he has been resurrected in the flesh, this is supposed to indicate that he lacks faith.

We can say something even stronger: faith seems to require not looking for further evidence even if the evidence is *cost-free*. For example, consider a case in which the man simply stumbles across an envelope which he knows contains evidence that will either vindicate his wife's constancy or suggest that she has been cheating. He seems to display a lack of faith in her constancy if he opens it, and to display faith in her constancy if he does not. And this seems true even if the evidence has been acquired in a scrupulous way: we might imagine the wife herself presents the envelope to the man, as a test of his faith.[14] Similarly, if I have faith that my friend will keep a secret, this seems to rule out asking a third party whether he thinks that friend is trustworthy.

This way of explaining how faith requires going beyond one's current evidence – that it requires deciding that one's current evidence is enough and one will not gather more – has the benefit of explaining how one can countenance the possibility that one might lose one's faith in the future. Even if a person will not engage in further inquiry, she might be presented with further evidence anyway: for example, even if she will not go out of her way to make further inquiries into her friend's trustworthiness, she might stumble across her friend revealing her secret.[15] This also enables us to make sense of the distinction between actively giving up one's

---

[14] Indeed, my account can easily explain why presenting him with the envelope could be a test of faith: it is a test to see whether he will choose to acquire further evidence.

[15] I've said that having faith does not require having a particular degree of belief. So how can losing faith in *x* consist in lowering one's degree of belief in *x*, as one does when one encounters indisputable evidence against *x* (e.g. the friend telling the secret)? The answer is that lowering one's degree of belief in *x* will usually result in a person failing to meet one of the other conditions of faith – for example, one will no

faith, which involves *deciding* to look for more evidence, and losing one's faith, which involves finding oneself in a state where one fails to meet one of the conditions for faith (i.e. you don't want to do the relevant act any more, you want to look for further evidence).

So we have the following analysis of faith:

A person has faith that X, expressed by *A*, only if that person performs act *A*, and *A* is better if X is true than it is if X is false, and there is some alternative act *B* that is available to the person such that *B* is better than *A* if X is false, **and the person refrains from gathering further evidence to determine the truth or falsity of *x* (or would refrain, if further evidence were available), even if the evidence is cost-free.**

One upshot of this is that it is possible that two people can have the same evidence, the same probability function, and the same utility function, and yet one displays faith and the other doesn't. So, for example, assume Ann and Erin have the same evidence, and both have p(Dan will keep a secret) = 0.9, and have the same utility functions. Now assume that each has a choice to ask a third party what he thinks about Dan's secret-keeping ability before deciding whether to tell Dan her secret. Ann decides to simply tell her secret; Erin decides to ask the third party, but then ends up telling her secret to Dan anyway. Here, it seems clear that Ann displays faith that Dan will keep a secret (relative to the act of revealing her own secret), whereas Erin does not

Having presented my analysis of faith, I now turn to the question of whether it is rational to have faith.

4. Expected Utility and Evidence-Seeking

I've argued that faith in X, expressed by *A*, requires that a person refrain from gathering further evidence that would bear on X. This requirement does not conflict with Bayesian rationality: it puts no constraints on either one's degree of belief in X or the relationship among degrees of belief or their relationship to evidence. But might it conflict with practical rationality? Ordinarily, we think that more information will put us in a better position.

I am going to show that in certain situations, more evidence might actually have a negative value, because of the possibility that the evidence is misleading: that is, more evidence might sometimes put us in a worse position. And, given this possibility, if the agent has a particular attitude towards risk (an attitude not licensed by standard decision theory), then more information is all-things-considered worse from the point of view of practical rationality. In particular, I am going to focus on situations of the following form: an agent is deciding between two acts, *A* and ~*A*. *A* is much better if X is true than *A* is if X is false; ~*A* is the same either way,

---

longer perform act *A* (e.g. sharing one's secret) or endorse a decision to perform *A* (if one has already done *A* and cannot undo it), or one will decide to look for further evidence.

but *A* is better than ~*A* if X is true and worse than ~*A* if *x* is false . Finally, the agent's degree of belief in *X* is fairly high: in particular, it's high enough that *A* is the better act for him to perform. To deal with specific utility and probability values, let us say that:

$u(A\&x) = 10\ u(\sim A\&x) = 1$

$u(A\&\sim x) = 0\ u(\sim A\&\sim x) = 1$

$p(x) = 0.9$

These are meant to have the form of typical instances of faith: note that they meet the criteria for faith mentioned in the first section. For example, let's say an agent is deciding whether to become a monk, and assigns degree of belief 0.9 to the proposition that God exists. If God exists, then becoming a monk is very good – the agent will experience all the goods of the religious life – but if God does not exist, then becoming a monk will result in the agent living a life that ultimately serves no purpose. On the other hand, failing to become a monk has no real costs (even from a religious point of view, if we assume becoming a monk is supererogatory): the agent lives roughly the same life as a non-monk whether or not God exists. Or an agent is deciding whether to reveal a secret to someone else, in whose trustworthiness he has degree of belief 0.9. If he reveals the secret, it is very good if the friend keeps it, and very bad if he doesn't. On the other hand, if he doesn't reveal the secret, then his life goes roughly the same regardless of whether the friend would have told the secret (he doesn't confide in anyone, and no one learns his secret). Finally, an agent might be deciding whether to marry a particular person, and is 0.9 certain that this person will make a good spouse. Again, if he does marry her, it is really good if she turns out to be a good spouse, and really bad otherwise. If he does not marry her, he will marry someone else, and so his life will be the same whether she would have been a good spouse or not. In each case, performing the act could turn out very well or poorly, whereas not performing the act is the same either way. We could think of the act as an opportunity for something great, that one might take or pass up.

Now we want to know whether the agent should gather more information before she makes her decision, assuming the information is cost free (including the cost of postponing the decision): conversely, we want to know whether it is rational for the agent to make the decision on faith.

On the standard theory of decision-theoretic rationality, expected utility theory, it is true that more information is always better: as a theorem due to I.J. Good shows, it always maximizes expected utility to gather new information (in Good's terminology, make a new observation) and then use that information to update one's beliefs, provided the cost of gathering that information

is negligible.[16] Therefore, the requirement that one refrain from further inquiry into the truth of x, even if the inquiry is cost-free or cost-negligible, does conflict with expected utility theory. So if expected utility theory is the correct theory of practical rationality, then having faith (in the sense I've spelled out) is irrational.

I am going to present the details of Good's theorem and provide some examples here; however, if the reader is less mathematically inclined, he can skip this section without missing the important philosophical upshots of the mathematics.

It is helpful to say something about Good's setup. An agent's expected utility before performing an experiment or making an observation is simply that of the act with the highest expected utility, given his current degrees of belief. For example, if the agent is choosing between three acts, A_1, A_2, and A_3, we can calculate the expected utility of each act: the expected utility of A_j is the sum of the utility values of performing A_j in each possible state, each value weighted by the probability of that state obtaining (that is, $\sum_i p(S_i)u(A_j \mid S_i)$). A rational agent will choose whichever of the three acts has the highest expected utility, so his expected utility in this position is just the expected utility of the act he chooses, or the maximum expected utility among the acts (that is, $\max_j \sum_i p(S_i)u(A_j \mid S_i)$). Now we want to calculate the expected utility of performing the experiment to see how it compares to this value. If he performs the experiment and then updates his degrees of belief based on the result of the experiment (which we might think of as a particular piece of evidence), he will then have one of several potential probability distributions. For each of these probability distributions, we can calculate the expected utility of A_1, A_2, and A_3, and we can determine which act has the highest expected utility on each of the resulting probability distributions; in other words, we can determine which act the rational agent will perform given each possible result (or piece of evidence he might receive).[17] For each possible result of the experiment, the utility of getting that result and using it will be the expected utility of the act the agent will perform if he gets that result (for each result E_k, the expected utility of using evidence E_k is $\max_j \sum_i p(S_i \mid E_k)u(A_j \mid S_i)$). We can then calculate the expected utility of performing the experiment, which is just the utility of getting each result, weighted by the probability of getting that result, summed:

---

[16] Good, I.J. (1967) "On the Principle of Total Evidence." *British Journal for the Philosophy of Science* 17 (4). Thanks to Branden Fitelson for bringing this paper to my attention.

[17] These may be different acts in different cases: for example, it might be that the experiment has four possible results, and that if the first result occurs, A_1 maximizes his (resulting) expected utility function; if the second result occurs, A_2 maximizes his expected utility function; and if one of the other results occurs, A_3 maximizes his expected utility function.

$\sum_k p(E_k) \max_j \sum_i p(S_i \mid E_k) u(A_j \mid S_i)$. Good proves that unless the act recommended is the same no matter what the result of the experiment is – that is, unless the same act maximizes expected utility for each E_k – then this value is always higher than the agent's expected utility before making the observation. In other words, it is always rational to make a new observation and use it.

For example, consider the person who has degree of belief 0.9 that God exists and is deciding whether to become a monk. He is considering whether to first perform a particular experiment: pray for something trivial (e.g. that he'll find a dollar bill on the ground, or his favorite song will come on the radio) and see if God responds. Let us assume that his degrees of belief obey the probability calculus and that he has the following degrees of belief in the propositions "God exists" (X) and "When I pray for something trivial in this circumstance I receive it" (R), and negations and conditional combinations of these:

p(X) = 0.9   p(X | R) = 0.99   p(X | ~R) = 0.05   p(R) ≈ 0.904

p(~X) = 0.1   p(~X | R) = 0.01   p(~X | ~R) = 0.95   p(~R) ≈ 0.096

Let us also assume that his utility payoffs for becoming a monk (A) and not becoming a monk (~A) are as above:

u(A&X) = 10   u(~A&X) = 1

u(A&~X) = 0   u(~A&~X) = 1

We want to know whether the agent should make the decision now, or perform the experiment before making the decision, assuming that performing the experiment has no cost, including any cost from postponing the decision. If he makes the decision now, his expected utility values are:

EU(A) = (0.9)(10) + (0.02)(0) = 9

EU(~A) = (0.9)(1) + (0.02)(1) = 1

So he will do A, with expected utility 9. On the other hand, if he carried out the experiment and found that R was true, his expected utility values would be:

EU(A) = (0.99)(10) + (0.01)(0) = 9.9

EU(~A) = 1

And if he carried out the experiment and found that ~R was true, his expected utility values would be:

EU(A) = (0.05)(10) + (0.95)(0) = 0.5

EU(~A) = 1

So, if he carries out the experiment and sees result R, then he will do A, with expected utility 9.9, and if he carries out the experiment and sees result ~R, then he will do ~A, with expected utility

1. Since he will get R with probability 0.904, and ~R with probability 0.096, his expected utility from performing the experiment is:

$\quad$ EU(E) = (0.904)(9.9) + (0.096)(1) ≈ 9.05

Since this is higher than the expected utility of doing A without performing the experiment (9), he should perform the experiment. As the theorem predicts, it is better to perform the experiment, provided the cost is negligible.

$\qquad$ What if the experiment is such that it won't influence what the agent will do? For example, consider a different experiment, that might yield result Q or result ~Q, about which he has the following degrees of belief:

$\quad$ p(X) = 0.9 $\qquad\qquad$ p(X | Q) = 0.99 $\qquad\qquad$ p(X | ~Q) = 0.2 $\qquad\qquad$ p(Q) ≈ 0.886

$\quad$ p(~X) = 0.1 $\qquad\qquad$ p(~X | Q) = 0.01 $\quad$ p(~X | ~Q) = 0.8 $\quad$ p(~Q) ≈ 0.114

This experiment differs from the first experiment in that a "negative" result will tell less strongly against X. Again, if he makes the decision now, he will do A with expected utility 9. If he carries out the experiment and finds that Q is true, his expected utility values will again be:

$\quad$ EU(A) = (0.99)(10) + (0.01)(0) = 9.9

$\quad$ EU(~A) = 1

And if he carries out the experiment and finds that ~Q is true, his expected utility values will be:

$\quad$ EU(A) = (0.2)(10) + (0.8)(0) = 2

$\quad$ EU(~A) = 1

So, no matter which result he sees, he will do A; A will have an expected utility of 9.9 if he gets result Q, and A will have an expected utility of 2 if he gets result ~Q. Since he will get Q with probability 0.886, the expected utility of performing the experiment is:

$\quad$ EU(E) = (0.886)(9.9) + (0.114)(2) = 9

This is equivalent to the expected utility of doing A without performing the experiment (and unsurprisingly, since the expected utility of performing the experiment if he is going to do A either way is simply a rearrangement of the terms of the expected utility of doing A without performing the experiment). If the agent will do the same act either way, then (as long as the experiment has no cost), the agent will be indifferent between performing the experiment and not performing it.

$\qquad$ So if we accept standard decision theory, then rational agents cannot have faith.

5. Risk and Evidence-Seeking

$\qquad$ Does this result hold for agents who take risk into account in the way I suggest; that is, does it always maximize risk-weighted expected utility to perform an additional (cost-free) experiment before making a decision? It turns out the answer is no: sometimes is it rational to

refrain from performing an experiment. I will first show why mathematically, and then explain why, intuitively, a person who is risk-avoidant will not benefit from performing an additional experiment in certain situations – situations that we think of as paradigm examples of faith, indeed, of well-placed faith. Consequently, if it is rational to be risk-avoidant, it is rational to have faith. Again, the less mathematically inclined reader can skip this section, since the upshots are described in a more intuitive way in the next section.

Again, the agent's RWEU before performing an experiment is the RWEU of the act with the highest RWEU, given his current degrees of belief: $\max_j RWEU(A_j | p)$[18] = (in our example) $\max \{0 + r(p(X))(10), 1\}$. And the agent's RWEU if he gets result E_i is going to be $\max_j RWEU(A_j | p_{Ei}) =$ (in our example) $\max \{0 + r(p(X | R))(10), 1\}$ if he gets result R and similarly for result ~R. In our discussion above, once we knew the EU of each result, we could treat the decision to perform the experiment as a gamble that yields EU_{Ek} with probability E_k, and calculate the expected utility of this gamble. However, we cannot do that here: whereas the expected utility of a gamble that yields, with probability p, an outcome of utility a is equivalent to the expected utility of a gamble that yields, with probability p, another *gamble* with *expected* utility a (assuming this gamble is identical elsewhere), it is not the case that the RWEU of a gamble that yields, with probability p, an outcome of utility a is equivalent to the RWEU of a gamble that yields, with probability p, another *gamble* with *risk-weighted expected* utility a (and is identical elsewhere). This is because the same subgambles might make different contributions to different gambles, because they might contribute different global properties.

So, instead, we consider the decision to perform the experiment as a gamble with n*m states, where m is the number of possible states S_i relevant to the decision, and n is the number of possible results of the experiment E_i.[19] For example, if there are two relevant states (X and ~X), and two possible results (R and ~R), then there are four possible states: R&X, R&~X, ~R&X, ~R&~X. For each experiemental result, we know the act the agent would do if he got that result, and we know the value of that act in each state. So performing the experiment is equivalent to taking the following gamble:

$u((\text{argmax}_{Ai} RWEU(A_i | p_E)) \text{ if } E\&X))$ with probability $p(E\&X)$

$u((\text{argmax}_{Ai} RWEU(A_i | p_E)) \text{ if } E\&\sim X)$ with probability $p(E\&\sim X)$

$u((\text{argmax}_{Ai} RWEU(A_i | p_{\sim E})) \text{ if } \sim E\&X)$ with probability $p(\sim E\&X)$

$u((\text{argmax}_{Ai} RWEU(A_i | p_{\sim E})) \text{ if } \sim E\&\sim X)$ with probability $p(\sim E\&\sim X)$

---

[18] "RWEU(A_j | p)" stands for the risk-weighted expected utility of act A_i assuming the agent has degree of belief function p.
[19] The EU of an experiment can be treated in this same way; it just happens to be equivalent to the other formulation because we can replace, in a gamble, an outcome by a gamble yielding that same outcome without changing the value of the overall gamble.

where RWEU $(A \mid p_E)$ is the risk-weighted expected utility of doing A when the probabilities of the states are given by updating p by conditionalizing on E.

So, let us assume the agent has the same probability and utility functions as before, but now let us assume he is risk avoidant, with $r(p) = p^2$. Recall that his utilities and probabilities are:

u(A&X) = 10	u(~A&X) = 1	p(X) = 0.9
u(A&~X) = 0	u(~A&~X) = 1	p(~X) = 0.1

p(X | R) = 0.99	p(X | ~R) = 0.05	p(R) ≈ 0.904
p(~X | R) = 0.01	p(~X | ~R) = 0.95	p(~R) ≈ 0.096

p(X | Q) = 0.99	p(X | ~Q) = 0.2	p(Q) ≈ 0.886
p(~X | Q) = 0.01	p(~X | ~Q) = 0.8	p(~Q) ≈ 0.114

Given these values, the risk-weighted expected utility of doing A and of doing ~A without performing the experiment are:

$$\text{RWEU}(A) = 0 + (0.9)^2(10 - 0) = 8.1$$
$$\text{RWEU}(\sim A) = 1 + (0.9)^2(1 - 1) = 1$$

So he will do A, with RWEU 8.1. If he performs the first experiment and gets result R, then the RWEU of doing A will be 9.8, and the RWEU of doing ~A will be 1, so he will do A. If he gets result R, then the RWEU of doing A will be 0.025, and the RWEU of doing ~A will be 1, so he will do ~A. Therefore, performing the experiment is equivalent to taking a gamble that yields 10 utiles if R&X, 0 utiles if R&~X, and 1 utile if ~R. The RWEU of this gamble is:

$$\text{RWEU}(E) = 0 + r(p(\sim R) + p(R\&X))(1 - 0) + r(p(R\&X))(10 - 1)$$
$$= (0.096 + 0.895)^2(1) + (0.895)^2(9) = 8.2$$

This is higher than the RWEU of doing A without performing the experiment; therefore it is rational for the agent to perform the experiment.

On the other hand, if he performs the second experiment and gets result Q, then the RWEU of doing A will again be 9.8, so he will again do A; and if he gets result Q, the RWEU of doing A will be 0.4, and he will again do ~A. Performing the second experiment is equivalent to taking a gamble that yields 10 utiles if Q&X, 0 utiles if Q&~X, and 1 utile if ~Q. The RWEU of *this* gamble is:

$$\text{RWEU}(E') = 0 + r(p(\sim Q) + p(Q\&X))(1 - 0) + r(p(Q\&X))(10 - 1)$$
$$= (0.114 + 0.877)^2(1) + (0.877)^2(9) = 7.9$$

This is lower than the RWEU of doing A without performing the experiment, so it is not rational to perform this experiment.

6. The Possibility of Misleading Evidence

What is going on in the second experiment? Before the agent looks at the evidence, he is taking a gamble (A) that has a p(X) chance of yielding 10, and a p(~X) chance of yielding 0. Looking at the evidence spreads out the possibilities (assuming that if he gets the "negative" result ~R, he will switch to ~A): he now has a p(R&~X) chance of getting 0 (he gets the positive result and so does A, but ~X is true nonetheless), a p(~R) chance of getting 1 (he gets the negative result and so does ~A), and a p(R&X) chance of getting 10 (he gets the positive result and so does A, and X is true). In effect, doing the experiment distributes some probability from the worst option (0) to the middle option (1), namely the probability associated with a "correct" negative result (~R&~X); it also distributes some probability from the best option (10) to the middle option (1), namely the probability associated with an "incorrect" negative result (~R&X). So we have an improvement in what happens in one state (the ~R&~X state) and a decline in what happens in another (the ~R&X state). For an expected utility maximizer, the improvement always makes up for the decline (again, provided the negative result makes ~A the better bet). But for a risk-weighted expected utility maximizer, this will not always be the case: if an agent is risk-avoidant, then improvements need to be greater in order to offset declines: part of what it means to be risk avoidant is to not be willing to accept the possibility of a loss in exchange for an equivalently-sized possibility of a gain. And improvements will not always be greater. It is difficult to give a simple description of what these situations are,[20] but two characteristics that they have in common are that (1) there is already a fairly high chance of X and (2) the negative result is not conclusive enough in favor of ~X.

Intuitively, if someone believes a proposition X to a high degree, then evidence in favor of that proposition (e.g. Q) won't drastically alter her degree of belief in it (it will only raise it slightly), so it won't make much of a difference to the value of taking gamble A. Therefore, the slight additional benefit that would come from being slightly more certain about X will not make much of a difference to her evaluation of whether she should perform the experiment. So the real question lies in determining the value of evidence against X (e.g., ~Q). Because A is a risky endeavor, evidence against X will often put the risk-avoidant agent in a situation in which is it rational for her to perform ~A. But from her (new) point of view, there's still a fairly high chance

---

[20] [[Note: I am still working on some necessary and sufficient conditions here, and hope to include them in the next draft.]]

(p(X | ~Q)) that she's gotten <u>misleading</u> evidence: evidence that makes it prudent to do ~A even though, as it turns out, X is true – that is, evidence that leads her to do the action that is rational but in the end gives her a lower payoff. This isn't the standard way the term "misleading evidence" is used, but it seems to me to capture an important way in which evidence can be misleading: it can lead a perfectly rational person to do something that happens to turn out worse for him.

If there is a fairly high chance that a negative result will present misleading evidence, then there's a fairly high chance that the negative result could lead the agent to miss out of the possibility of something great: doing A when X is true. And, for the risk avoidant agent, this potential detriment might outweigh the benefit of a negative result that is not misleading (because ~X is in fact true).

I emphasize that being "misled" needn't involve any irrationality on the agent's part, because the action that is *rational*, given what one *believes*, is not always the action that is best, given what is *true*. Our case is very different from a case in which the agent does not want to discover new evidence lest she find out something depressing – in our terms, it is different from a case in which R&~X is preferable to ~R&~X or to ~X.[21] I am certainly not claiming that this preference is rational. Instead, ours is a case in which looking for more evidence comes with a significant chance that one will perform an action that is rational but, given the way the world turns out, wrong – it comes with a significant chance that one will make a mistake, so to speak. And this seems to accord with the examples I gave of faith, above: if a private investigator turns up evidence that leads one to believe one's significant other will be unfaithful, it would indeed be rational not to marry that person, but, given your beliefs, *there is a significant chance that by doing the rational thing you are missing out on some great good*, namely a relationship with a spouse who is *in fact faithful*. And if God doesn't respond to a believer's heartfelt prayer in a "testing" situation, it might indeed be rational for the believer not to become a monk, but even then, there will be a significant chance that he is missing out on a great good, namely becoming a monk when God in fact exists.

If one's credence in X is antecedently very high, then carrying out an experiment that turns out to confirm X is not particularly helpful: it is very unlikely to change one's course of action, and it increases the value of this action only slightly. On the other hand, carrying out an experiment that turns out to disconfirm X might lead one to rationally adopt a different action:

---

[21] Note that our agent would rather be in state ~R&~X than in state R&~X (or state ~X): in the former state, she would do ~A and get a payoff that has utility 1, and in the latter two states she would do A and get a payoff that has utility 0.

and it might lead one to adopt a different again that yields a better result, but it might lead one to adopt a different action that yields a worse result. In other words, carrying out an experiment is a risky endeavor. And if one's credence in X is antecedently very high, the experiment that disconfirms X is not conclusive enough, and one is risk-avoidant, then the potential bad that comes from the experiment outweighs the potential good.

So faith and risk avoidance rise or fall together, with one caveat that I will mention shortly. Of course, this could be taken in two ways: one could think that a variety of attitudes towards risk are rationally permissible, and conclude that it is sometimes rationally permissible to fulfill the requirement of faith that I have spelled out. Or one might think that we are always rationally required to look for all the available cost-free evidence before making decisions, and conclude that it is not rational to adopt any attitude towards risk except the one recommended by expected utility theory. As should be obvious, I favor the former approach, though I have not here provided an argument for my preferred theory of rational decision making. I have, however, undermined one significant reason we might have had for thinking that we are always required to look for evidence: whereas we might have thought more evidence always puts us in a better position, this is not the case with misleading evidence.

Faith in X expressed by A is rational when (1) A is a risky act, relative to the available alternatives; (2) there is already a fairly high chance of X; and (3) the negative result of the experiment one is considering performing is not conclusive enough in favor of ~X. The second condition coincides with what we naturally think of as instances of "good" faith, as mentioned above: one already has a fairly high degree of belief in X. This means that our result and our pre-theoretic intuitions coincide, which is promising. However, the third condition introduces the need for a significant caveat about our conclusion. Note that the thing to be analyzed – faith in a proposition expressed by an action – is not relative to any particular experiment, whereas the third condition is: whether it is rational to perform an experiment does depend on the experiment in question. In particular, it depends on the degree to which a negative result would be conclusive evidence for ~X (i.e. the conditional probability of ~X given ~R).[22] And for any degree of belief in X and alternatives A and ~A of the same form as in our example,[23] there will be some possible experiment that it would be rational to perform before choosing A or ~A. For example, an experiment R such that $p(\sim X \mid \sim R) \approx 1$ (and thus $p(X \mid \sim R) \approx 0$) will always be rational to perform, because there is no possibility of misleading evidence in this experiment. But from

---

[22] The threshold that $p(\sim X \mid \sim R)$ needs to meet in order to make it irrational for a risk-avoidant agent to perform the experiment will depend on the possible utilities of doing A and doing ~A, among other values.
[23] I.e. $u(A\&X) > u(\sim A\&X) = u(\sim A\&\sim X) > u(A\&\sim X)$.

what I've said above, it seems that faith requires a refusal to perform *any* experiment whose sole purpose is to determine the truth or falsity of X. The defender of faith as rational absolutely must resolve this dilemma.

There are two possible avenues that the defender of the rationality of faith could take: one is to amend the definition of faith so that it is relative to a particular search one is considering conducting. The downside of this approach is that it does not seem to square with how we think about faith: while faith may be relative to a particular act, it is not obviously relative to a particular evidential search. The other alternative is to leave the definition of faith unrestricted, but limit it to actual acts, and argue that in many cases, there are no actual searches that meet the threshold for providing conclusive enough evidence. On this picture, faith will only be rational in cases in which such searches are not available. So while faith itself won't be relative to a particular search, whether faith is rational will depend on which searches are available. Of course, these do not exhaust the options for resolving the dilemma.

7. Conclusion

My goal was to analyze faith claims in a way that unifies faith claims in religious contexts and in secular contexts, and to determine whether and when it is rational to have faith in something. I focused on how to make sense of the idea that faith requires going beyond one's evidence in some way. Initially plausible attempts to explain this requirement interpreted it as a requirement about one's degree of belief function or its relationship to the available evidence: however, these attempts did not succeed. (This is a good thing from the point of view of someone who wishes to defend faith as a rational attitude, since these attempts all spell out a requirement that is epistemically irrational.) I argued that the best way to explain this requirement is that having faith in some proposition X expressed by A requires that one refrain from searching for further evidence for the truth or falsity of X before deciding to perform A. As it turns out, whether and when it is rational to stop one's search for evidence hinges on which attitudes towards risk it is rationally permissible to adopt when making decisions: more specifically, it hinges on how one should respond to the risk involved in the possibility of misleading evidence. So the rational permissibility of faith and the rational permissibility of risk aversion are inextricably linked. I have argued elsewhere that risk aversion (in the sense of global sensitivity) is in fact rationally permissible; indeed, one upshot of my theory is that there are many acceptable attitudes towards risk, and so a rational agent can and must choose his own level of risk aversion. If I am correct about the rationality of risk aversion, then the choice of whether to have faith is a choice that can and must be up to an individual.