On Indeterminate Updating of Credences

Leendert Huisman*

The strategy of updating credences by minimizing the relative entropy has been questioned by many authors, most strongly by means of the Judy Benjamin puzzle. I present a new analysis of Judy Benjamin–like forms of new information and defend the thesis that in general the rational posterior is indeterminate, meaning that a family of posterior credence functions rather than a single one is the rational response when that type of information becomes available. The proposed thesis extends naturally to all cases in which new information is traditionally handled by minimizing the relative entropy.

1. Introductory Overview. In the Bayesian approach to rational deliberation (e.g., Howson and Urbach 1993; Jaynes 2003; Joyce 2009), an agent indicates how strongly she believes a certain statement by giving that statement a credence between 0 and 1. The closer to 1 this credence is, the more strongly the agent considers the statement to be true. In this approach, credences behave like probabilities and are updated when new information becomes available. Consider then the following story (van Fraassen 1981, 376–77):

> [Judy Benjamin] enters the army and during war games, she and her patrol are dropped in a swampy area which they have to patrol. . . . The war games area is divided into the region of the Blue Army, to which Judy Benjamin and her fellow soldiers belong, and that of the Red Army. Each of these regions is further divided into Headquarters Company Area and Second Company Area. The patrol has a map which none of them understands, and they are soon hopelessly lost. Using their radio they are at one point able to contact their own headquarters. After describing whatever they remember of their movements, they are told by the duty officer "I don't know whether or not you have strayed into the Red Army territory. But if

you have, the probability is 3/4 that you are in their Headquarters Company Area." At this point the radio gives out.

How should Judy Benjamin update her credences, in particular, her credence that she has landed in the Blue Army's area?

In classical Bayesianism, the only type of new information for which updating is defined is that which makes the agent become certain of some proposition, say E. If Cr_0 indicates the agent's prior credence, the credence before the new information became available, and Cr her posterior credence—that is, the credence obtained by updating her prior credence—then

$$Cr(.) = Cr_0(.|E). \tag{1}$$

This form of updating is also known as Bayesian Conditionalization. It was extended by Jeffrey (1983) to include new information that would make the agent assign precise credences to the members of a partition, a set of statements of which not more than one can be true and of which exactly one must be true. If we indicate this partition by $\{B_i\}$, where B_i is an arbitrary member of the partition, then, according to Jeffrey,

$$Cr(.) = \sum Cr_0(.|B_i)Cr(B_i), \tag{2}$$

in which $\{Cr(B_i)\}$ is the new information. This form of updating is often referred to as Probability Kinematics and, formally, includes Bayesian Conditionalization. The accompanying information, the set $\{Cr(B_i)\}$, will be referred to as Jeffrey information.

The information that Judy Benjamin receives, however, does not seem to have either of these forms. It is not some proposition that she is made aware of, nor is it a posterior credence function on the members of a partition. Instead, what she is told is the posterior value of a specific conditional credence. No generally accepted rule for updating credences upon the acquisition of that type of information is available. One that has been vigorously defended by some (e.g., Shore and Johnson 1980; Williams 1980) is minimization of the relative entropy:

$$\sum Cr(L) \ln \frac{Cr(L_i)}{Cr_0(L_i)}, \tag{3}$$

subject to one or more constraints representing the new information. The set $\{L_i\}$ is the coarsest relevant partition, and the constraint in Judy Benjamin's case is the value of a posterior conditional credence.

Uffink (1995), however, has shown that the arguments in favor of minimizing the relative entropy are not convincing, and van Fraassen (1981), with his Judy Benjamin story, has shown that minimizing the relative entropy can, in fact, lead to counter-intuitive if not outright implausible results.

Let 'B' indicate 'I am in the Blue Army's area', 'H' 'I am in a headquarters area', and 'S' 'I am in a second company area'. Finally, let '¬B' stand for 'not B'. Judy's credences before she received the radio message were $Cr_0(B) = 0.5$ and $Cr_0(H\&\neg B) = Cr_0(S\&\neg B) = 0.25$. The new information amounts to $Cr(H|\neg B) = 0.75$. What, for her, is now $Cr(B)$? Minimizing the relative entropy gives $Cr(B) = 0.533$, but that is rather counter-intuitive, for why would information about where in the Red Army's area she has landed, assuming that she has landed in that area, give information about whether she has landed in that army's area? And, assuming that it does give some information, why would it make her increase her credence in not having landed in that area? In fact, minimizing the relative entropy will make her raise her credence for having landed in the Blue Army's area, no matter what value of $Cr(H|\neg B)$ she is given. That result is implausible and led van Fraassen to comment: "It is hard not to speculate that the dangerous implications of being in the enemy's headquarters area, are causing Judy Benjamin to indulge in wishful thinking" (379).

Van Fraassen, Hughes, and Hartman (1986) proposed several alternatives to updating by minimizing the relative entropy and demonstrated that all of them have the same counter-intuitive consequence as minimization of the relative entropy and that none of these alternatives is clearly preferable over the others. Grove and Halpern (1997) suggested that the problem be embedded in a larger space in which one can conditionalize on the conditional credence. Even though they do obtain the plausible answer that Judy's credence in having landed in the Red Army region does not change, this approach still has serious problems. It is unclear whether it can easily be applied to more complex Judy Benjamin–like problems, and it introduces the new problem of having to choose a prior credence function in the larger space.

Douven and Romeijn (2011) proposed a minimization procedure that produces the intuitively correct answer in the Judy Benjamin case but which has severe drawbacks in other situations. It corresponds, in fact, to Adams conditioning (Bradley 2005) in which the credence in the antecedent is assumed not to change. This may be a correct assumption in some situations but is clearly not tenable in others. Furthermore, Douven and Romeijn's arguments in favor of Adams conditioning, even in the context of the original Judy Benjamin problem, are not convincing. It seems plausible that, in the Judy Benjamin story as told by van Fraassen, Judy has little reason to alter her credence in being in the Blue Army's area because all she learns is where she is likely to be if she is in the Red Army's area. But in that version her prior credence in being in any of the four possible regions is the same; she is equally convinced that she will land in a headquarters region as that she will land in a second company region. What would happen if her prior beliefs were different?

Consider, for example, the scenario in which Judy's prior credence in having landed in some headquarters region is very low. For example, the training intentionally included a low and known probability of being dropped off in a headquarters region. When she then learns that, if she has landed in the Red Army's area, she is most likely to find herself in that army's headquarters region, it seems more plausible for her to lower her credence for having landed in that army's area. Keeping it the same, after all, would imply that it is more likely than not that she has landed in the headquarters region of some army, and that consequence is contradicted by the training parameters. Contra Adams conditioning, therefore, her credence in having landed in the Red Army's area may very well change, depending on her prior credences and the new information she receives. The second scenario (low probability of landing in a headquarters region) leads to a lowering of her credence that she has landed in the Red Army's area, which is exactly what minimizing the relative entropy predicts.

But it is not just changing prior credences that can lead us away from Adams conditioning or minimizing the relative entropy. It is easy to construct stories that are structurally similar to that of Judy Benjamin and that have identical prior credences but in which the credence in the antecedent clearly changes upon receiving information that is analogous to that which Judy receives. Consider the following story:

> Harry has learned that his friend Tom has been offered a new job. Harry knows little about the job other than that it will require relocation to either the West Coast or the East Coast. Harry knows that Tom is also interested in moving west (they now live in Cleveland), and getting a new job may interfere with those plans. In fact, he thinks that Tom, being young and adventurous, may very well give up his present job even without finding a new one, move west, and hope to find a job there. Harry has a credence of 0.5 that Tom will move west soon (within a year) and a credence of 0.5 that Tom will accept the new job. He then learns from Sue, a mutual friend, that the odds are 3 to 1 that Tom will have to move to San Francisco if he accepts the job offer. Harry's credence that Tom will accept the offer now immediately increases.

I think it is uncontroversial that Harry's credence that Tom will accept the job offer increases when he learns from Sue that Tom might have to move west if he did accept the offer, for moving west is something he was thinking of doing anyhow. So Adams conditioning will not do. But minimizing the relative entropy will not do either because Harry's credence in the antecedent, that Tom will accept the job offer, increases rather than decreases as minimizing relative entropy would predict. Not that the latter update technique is always wrong. If Sue had told Harry that the job offer comes with a 3-year commitment in New Jersey, Harry's credence that Tom will accept

the offer would have gone down, which is in agreement with minimization of the relative entropy.

The job story and the Judy Benjamin story are structurally the same because we can map "being in the Red Army's" territory on "accepting the job offer" and "being in a headquarters area" on "moving west," which will maintain all the (conditional) credences. They are not the same in all respects, however, and the question may arise whether the difference in the change in the credence of the antecedent can be explained by these non-structural differences. For one thing, the Judy Benjamin story is indexical: Judy does not know where she is. That difference cannot explain the difference in doxastic behavior, however, because the story could have been told from the viewpoint of an observer in the Blue Army's area who is able to listen to the radio communications between Judy and the duty officer. His credences concerning Judy's location are not indexical, but the same analysis can be given for his posterior credence that she is in the Blue Army's area as for Judy's credence that she is in that area, with the same implausible result if minimization of the relative entropy is used. A more important difference is that Judy's posterior credence concerns where she is, which is an already established fact: she is where she is, but she does not know where that is. In the job offer story, however, the relevant posterior credence concerns whether Harry will accept the job offer and when he will move west, the answers to which lie in the future and depend on Harry's desires. Even though this difference is real, the simplistic rule to use Adams conditioning in the first case and something else in the second case cannot be valid because Adams conditioning is wrong in the Judy Benjamin case after a minor change in her prior credences, while Tom's posterior credence that Harry will accept the job offer can go both up and down, depending on the details of the job offer. It does not seem from these two examples and the minor variations in the two examples, therefore, that the choice of which mechanism to use to determine posterior credences is tied in any simple way to the details of the story.

The Judy Benjamin story by itself has, of course, only limited value, but it can be generalized to a new type of information that is not of the Bayesian type (one that is suitable for Bayesian Conditionalization) or of the Jeffrey type (suitable for Probability Kinematics). I refer to this new type of information as Judy Benjamin information. It consists of a set $\{Cr(B_i|A)\}$ of conditional credences, with $\{B_i\}$ a partition and A some statement for which $Cr_0(A)$ does not vanish. The essential difference between Judy Benjamin type of information and Jeffrey type of information is that the former requires the values of conditional credences pertaining to some partition rather than absolute credences. The central lesson coming out of the Judy Benjamin example is that minimizing the relative entropy is not likely to be the correct way of updating credences in the general case involving Judy Ben-

jamin type of information. This point has been confirmed subsequently by various other authors (most recently by Douven and Romeijn 2011).

I should point out, however, that van Fraassen (1989, chap. 13.6) does not seem to see it that way. In his opinion, maximizing the entropy or minimizing the relative entropy is generally valid, at least when the new information consists of expectation values of random variables, which class includes that of the Judy Benjamin variety. That that update method gives strange results in the Judy Benjamin case is, if I understand him correctly, not sufficient reason for him to abandon it. He does allow, however, for a nonunique answer to the Judy Benjamin problem. Moreover, the result that Judy's credence of being in the Red Army's region decreases has been argued to be correct by Lukits (2014).

Minimizing relative entropies is not appropriate in the original Judy Benjamin story, and using some other update mechanism, such as Adams conditioning, is inappropriate when some of the details of the Judy Benjamin story are changed. A single update mechanism when Judy Benjamin type of information becomes available will, therefore, not do. But, Adams conditioning seems, intuitively, to be the correct way of updating credences in the original Judy Benjamin story, and updating by minimizing the relative entropy undoubtedly gives the intuitively correct results in some other scenarios, showing that both seem to be appropriate in at least some scenarios. What seems to be more promising, therefore, is a function that maps each specific scenario with Judy Benjamin type of information to a specific update mechanism. In other words, what might be required is a family of update mechanisms and a function that maps the set of Judy Benjamin type of scenarios into this family of update mechanisms.

I present a plausible family of update mechanisms in section 2, but finding an associated function that maps Judy Benjamin–type problems into this family is less straightforward. In fact, I argue in section 3 that no such function is likely to be found, at least when we demand that it maps each specific problem onto a single update mechanism. I argue, instead, that the best we can do is to have such a function map Judy Benjamin–type problems onto subsets of update mechanisms—even, in some cases, onto the whole family of mechanisms. We may consider a subset having more than one member a consequence of our limited semantic sophistication when confronted with Judy Benjamin–type problems, but, as I further argue, it may also be an inevitable consequence of the nature of such problems in the sense that no increase in semantic sophistication will reduce the subsets to singletons—in other words, that Judy Benjamin–type problems may lead by their very nature to indeterminate posteriors.

This conclusion has obvious relevance for the general problem of how to update credences when new information, of any type, becomes available. That is, it addresses the core of Bayesianism, the idea that the strengths of

our beliefs can be modeled by probability-like credences and that the effect new information has on those credences is a determinate update from one credence function to another. The argument alluded to in the previous paragraph indicates that, unlike the severely constrained case of Jeffrey type of information, more general forms of new information lead to posterior credence functions that are, practically or constitutionally, indeterminate. I return to this point in section 4. It answers more fully van Fraassen's original concern: whether minimizing relative entropies is always the correct way of updating credences upon the acquisition of new information, and, if not, what should take its place. Van Fraassen had already indicated that the answer to the first question is negative (but not in his later work [van Fraassen 1989]; see previous comments). The conclusion of this article is that no other single update mechanism will do either and that, for general forms of new information, including Judy Benjamin type of information, updating credences may produce indeterminate posteriors.

2. A Universal Family of Update Mechanisms. In this section, I present a family of possible mechanisms for updating credences when confronted with Judy Benjamin type of information. I focus on the posterior credence of the (negation of the) conditioning statement.

Let $\{B_i\}$ be a partition, A such that $Cr(A) > 0$, and $\{Cr(B_i|A)\}$ the set of newly acquired posterior credences of the members of the partition. Unless otherwise mentioned, the partition has n members, and I assume throughout this article that the prior credences of the members of the partition are nonzero. The posterior conditional credences are nonnegative and sum to 1. How $\{Cr(B_i|A)\}$ is obtained is not part of the update process. It is presumably conveyed to the agent by means of conditional statements and may involve odds ratios, conditional credences, and the like. Interpreting conditional statements is notoriously difficult, and I assume that the information, in whatever form it was given originally, has been translated correctly to the set $\{Cr(B_i|A)\}$. New information of the Judy Benjamin type is similar to Jeffrey information, be it that the former is based on the partition $\{\neg A, B_1 A, \ldots, B_n A\}$ rather than the partition $\{B_i\}$. The new credences of the members of the former partition are given by $Cr(\neg A)$, and $Cr(B_i A) = Cr(B_i|A)$ $(1 - Cr(\neg A))$, with $Cr(\neg A)$ unknown.

A powerful way of determining posterior credences is by minimizing some functional of the prior and posterior credences under appropriate constraints, with this functional having the property that it is minimal when prior and posterior credences are the same. Shore and Johnson (1980) have studied plausible properties that any such functional should have, and Uffink (1995) showed that their axioms are met by the members of the family of α-divergences $U_\alpha(Cr, Cr_0)$ (also referred to as Rényi relative entropies), defined as

$$U_\alpha(Cr, Cr_0) =_D \frac{sgn(\alpha)}{\alpha - 1} \ln\left(\sum Cr_0(L_i)r(L_i)^\alpha\right), \qquad (4)$$

in which the sum is taken over the members of the coarsest relevant partition ($\{L_i\}$), and

$$r(.) =_D \frac{Cr(.)}{Cr_0(.)}. \qquad (5)$$

If $Cr_0(L_i) = 0$ for some i, U_α is infinite (and, therefore, not a minimum) unless (a) $Cr(L_i)$ vanishes as well and (b) $r(L_i)$ is assumed to be finite. The family of α-divergences will be indicated by **F**. Clearly $U_0(Cr, Cr_0) = 0$ and does not correspond to a usefully minimizable functional. We will see that it can be taken to correspond to Adams conditioning. The case of $\alpha = 1$ is more interesting. It is not defined directly by equation (4), but, by taking the limit of α going to 1, we find the relative entropy (eq. [3]).

   Divergence $U_\alpha(Cr, Cr_0)$ has a number of attractive features, in addition to meeting the Shore and Johnson axioms. First, it includes the relative entropy, as we already saw. Second, it is nonnegative and equal to zero if and only if $Cr(B_i) = Cr_0(B_i)$ for all members of the partition because $x^{1-\alpha}$ is a concave function of x for $0 < \alpha < 1$ and a convex one otherwise. Third, U_α (Cr, Cr_0) reproduces Probability Kinematics when Jeffrey type of information becomes available. This crucial property can be demonstrated easily and completely generally. Consider an arbitrary statement E for which we want to determined the posterior credence: E gives rise to the partition $\{EB_1, \ldots, EB_n, \neg EB_i, \ldots, \neg EB_n\}$. We proceed by minimizing $U_\alpha(Cr, Cr_0)$ for this partition with the Jeffrey constraint

$$Cr(EB_i) + Cr(\neg EB_i) = Cr(B_i) \qquad (6)$$

and known $\{Cr(B_i)\}$. Equation (4) becomes

$$U_\alpha(Cr, Cr_0) = \frac{sgn(\alpha)}{\alpha - 1} \ln\left(\sum Cr_0(EB_i)r(EB_i)^\alpha + \sum Cr_0(\neg EB_i)r(\neg EB_i)^\alpha\right). \qquad (7)$$

Note that $r(EB_i) = Cr(B_i)Cr(E|B_i)/Cr_0(EB_i)$ and $r(\neg EB_i) = Cr(B_i)(1 - Cr(E|B_i))/Cr_0(\neg EB_i)$ and that the members of $\{Cr(E|B_i)\}$ can be varied independently in [0, 1]. Since both Cr_0 and $\{Cr(B_i)\}$ are known and fixed, minimizing U_α is accomplished by minimizing it with respect to each $Cr(E|B_i)$. This leads to $r(EB_i) = r(\neg EB_i)$, or $Cr(E|B_i) = Cr_0(E|B_i)$. The matrix of second derivatives is easily shown to be positive and definite. The equality between $Cr(E|B_i)$ and $Cr_0(E|B_i)$ is called rigidity and is a consequence of the particular form of U_α. Equation (2) follows immediately because Jeffrey Condi-

tionalization can be proven when rigidity is assumed (Diaconis and Zabell 1982).

With a Judy Benjamin constraint, the appropriate partition is $\{\neg A, B_1A, \ldots, B_nA\}$, and $\{Cr(B_i|A)\}$ is given. I restrict the analysis to the case that $Cr_0(B_iA) > 0$ for all i. To simplify subsequent expressions, I abbreviate $Cr_0(B_i|A)$ by p_i and $Cr(B_i|A)$ by q_i: q_i is known and positive for each i, and $q_i = 1$. The Judy Benjamin constraint becomes $Cr(B_iA) = q_i(1 - Cr(\neg A))$ for $i = 1, \ldots, n$, with $Cr(\neg A)$ unknown. We proceed as follows. We first determine $Cr(\neg A)$ by minimizing U_α with the underlying partition $\{\neg A, B_1A, \ldots, B_nA\}$. Equation (4) becomes

$$U_\alpha(Cr, Cr_0) = \frac{\text{sgn}(\alpha)}{\alpha - 1} \ln\Big(\sum Cr_0(\neg A)r(\neg A)^\alpha + \sum Cr_0(AB_i)r(AB_i)^\alpha \Big). \quad (8)$$

Minimizing with respect to $Cr(\neg A)$ leads to the equation $r(\neg A) = r(A)D$ with

$$D^{\alpha-1} =_D \sum q_i \left(\frac{q_i}{p_i}\right)^{\alpha-1}. \quad (9)$$

The final result is

$$Cr(\neg A) = \frac{Cr_0(\neg A)D}{1 - Cr_0(\neg A) + Cr_0(\neg A)D}. \quad (10)$$

Equation (10) will be referred to as the Judy Benjamin equation. Clearly, when $\alpha = 0$, $D = 1$, and $Cr(\neg A) = Cr_0(\neg A)$. As we saw before, U_0 is constant and, therefore, not suitable for minimization, but we can replace that member of the family of α-divergences with Adams conditioning. When $\alpha = 1$, we need to take the limit of α going to 1, and find $\ln D = \sum q_i \ln(q_i/p_i)$.

Once $Cr(\neg A)$ has been determined, the credences of all the members of the partition are known, but then we are back at a standard Jeffrey problem, for which the solution is

$$Cr(.) = Cr(\neg A)Cr_0\Big(.|\neg A\Big) + Cr(A) \sum Cr_0\Big(.|\neg AB_i\Big)q_i, \quad (11)$$

for the calculated value of $Cr(\neg A)$. Note the extreme simplicity of the result: when determining a posterior distribution for Judy Benjamin type of information by minimizing an α-divergence, we find Jeffrey Conditionalization again. The input that drives the update is $\{Cr(B_i|A)\}$, and $Cr(\neg A)$ is obtained from the Judy Benjamin equation. All α-divergences produce a Jeffrey Conditionalization result; they differ only in their values of $Cr(\neg A)$.

In general, the Judy Benjamin equation requires a simple calculation once D is known. The question now is how $Cr(\neg A)$ varies when α is varied. First, note that D is nonnegative and equal to 1 when $\alpha = 0$. Second, $Cr(\neg A)$ is

a monotonically increasing function of D, and, therefore, we need to determine only how D varies with α. Third, D is a continuous monotonically increasing function of α. For $\alpha > 1$, this follows immediately from the Hölder inequality. For $\alpha < 1$, define $C = 1/D$ and write $C^{1-\alpha} = \Sigma q_i (p_i/q_i)^{1-\alpha}$. Finally, the derivative with respect to α at $\alpha = 1$ exists and is positive. Consequently, $Cr(\neg A)$ is a monotonically increasing function of α as well: it is less than $Cr_0(\neg A)$ when α is negative and larger than $Cr_0(\neg A)$ when α is positive.

Because of the monotonicity of $Cr(\neg A)$ as a function of α, the possible values of $Cr(\neg A)$ are bounded by their values for $\alpha = -\infty$ and $\alpha = \infty$. In equation (9), the ratios q_i/p_i can take all values in $(0, \infty)$. Let s be the index of the smallest ratio and l that of the largest one. Note that $q_s/p_s < 1$ and that $q_l/p_l > 1$ because $q_i = p_i = 1$. Let $D^{-\infty}$ and D^{∞} be the values of D for $\alpha = -\infty$ and $\alpha = \infty$, respectively. We find that $D^{-\infty} = q_s/p_s$ and $D^{\infty} = q_l/p_l$. Finally, we have

> **Theorem 1.**   For any α-divergence and s and l such that $q_s/p_s \leq q_i/p_i \leq q_l/p_l$ for all i,
>
> $$\frac{Cr_0(\neg A)q_s}{Cr_0(A)p_s + Cr_0(\neg A)q_s} \leq Cr(\neg A) \leq \frac{Cr_0(\neg A)q_l}{Cr_0(A)p_l + Cr_0(\neg A)q_l}.$$

Both bounds are in [0, 1]. The lower bound is larger than 0 unless $q_s = 0$, and the upper bound is less than 1 unless $p_l = 0$ (which is excluded by assumption). When q_s/p_s and q_l/p_l both go to 1, that is, when q_i goes to p_i for every i, the bounds become increasingly tight and, in the limit, become equal to $Cr_0(\neg A)$.

These bounds naturally extend to bounds on $Cr(AB_i) = q_i Cr(A)$. To simplify the notation, define $r_i = q_i/p_i$. We find that

$$\frac{Cr_0(AB_i)r_i}{Cr_0(A) + Cr_0(\neg A)r_l} \leq Cr(AB_i) \leq \frac{Cr_0(AB_i)r_i}{Cr_0(A) + Cr_0(\neg A)r_s}. \tag{12}$$

In contrast with the bounds on $Cr(\neg A)$, which straddle the prior credence $Cr_0(\neg A)$, for some indices, the bounds on $Cr(AB_i)$ exclude the corresponding prior credence. In particular, the lower bound exceeds $Cr_0(AB_i)$ when $r_i - r_l > Cr_0(A)(1 - r_l)$, and the upper bound is less than $Cr_0(AB_i)$ when $r_i - r_s < Cr_0(A)(1 - r_s)$. Since $r_s < 1 < r_l$, we find that $Cr(AB_s)$ is less than $Cr_0(AB_s)$ and $Cr(AB_l)$ is larger than $Cr_0(AB_l)$ for any α.

For the original Judy Benjamin problem, the range of possible credence values of $Cr(B)$ is [1/3, 3/5], as was already found by Uffink (1995). That for $Cr(\neg BH)$ is [3/10, 1/2], and that for $Cr(\neg B\neg H)$ is [1/10, 1/6], neither one of which contains the prior value 1/4. In other words, even though Judy's

credence that she is in the Red Army's area might increase or decrease when she gets the information from the duty officer, her credence that she is now in the Red Army's headquarters region definitely increases, and her credence that she is now in the Red Army's second company region definitely decreases.

3. Indeterminate Updating. Different members of the family of α-divergences may lead to different posteriors in the Judy Benjamin case, but the correct update rule for that problem should be found among the members of that family. It would clearly be desirable to find the assumed single correct posterior, but I argue in this section that such a result is not likely to emerge. Instead, the correct solution is (a subset of) the set of all posteriors produced by the family of rules. I now present and defend this new solution of the Judy Benjamin problem and address some possible objections.

3.1. Vague Posterior Credences. In the preceding section, I presented the purely mechanical part of updating by minimizing α-divergences. I showed that this minimization produces a particularly simple result for Judy Benjamin type of information. Upon receiving this information, the credence in $\neg A$ increases or decreases, depending on which particular functional was used. Unfortunately, there is little guidance in the literature for how to choose the optimal functional. The relative entropy has many theoretical advantages (Williams 1980) but gives counter-intuitive results when applied to the original Judy Benjamin problem (van Fraassen 1981; Seidenfeld 1986; van Fraassen et al. 1986); the inverse relative entropy (Joyce 2009; Douven and Romeijn 2011) leads to Adams conditioning (Bradley 2005) and may give a more plausible answer for that particular problem but keeps the credence in $\neg A$ unaltered, regardless of the prior credence function or $\{Cr(B_i|A)\}$.

But perhaps each individual case in which Judy Benjamin type of information becomes available requires its own functional. Additional information that is present in each particular story (being dropped in overlapping geographical regions in Judy's story, Tom wanting to move west in the job offer case) may suffice to determine which functional (different ones for different stories) to minimize. To use that additional information, however, would require it to be formalized in such a way that the determination becomes feasible. If such a capturing of additional information were possible, one might imagine a catalog of different types of Judy Benjamin–like scenarios accompanied by a list of functionals, one functional for each item in the catalog.

Such capturing of additional information is possible in many situations, as, for example, when we receive the information that the match will be canceled if it rains. Adams conditioning is now clearly the correct way to

update one's credences because our credence that it will rain should not be changed by the information that a match will be canceled if it does. Nevertheless, I do not consider the construction of such a catalog and the corresponding list of functionals feasible in general, for a number of reasons. First, there is little in the literature that suggests how to formally describe these different types of additional details or, for that matter, whether such a formal description even is possible. The canceled match story is special because of the obvious lack of a link from the decision to cancel a match to the occurrence of rain. Second, if such a formal description could be formulated, the resulting catalog would be very large, to say the least, because of the almost endless variety of added details that might influence the agent's posterior credences. The job offer stories provide just two examples of such added details. In both stories, these details provide Harry with some information about how likely it is that Tom will accept the job offer. That type of additional detail can be varied indefinitely, with potentially different reactions on Harry's part.

Third, even if such a catalog could be constructed, that still does not tell us what functionals to choose for the different items in the catalog. That choice is already obscure for the job offer story, which has a fairly rich background of details. What functional would have to be minimized when no details are provided? For example, how would one handle the following scenario?

> Carla is an FBI agent assigned to monitoring a group of right-wing white supremacists called the "White Vipers." Membership of this group is only partially known. Intercepted messages between members of the group oftentimes mention a person only known as "The Crusher," but Carla's team does not know whether he is a member of the White Vipers, considered to be a friend, or on their hit list. Recently, the team started to intercept messages alluding to the bombing of some government building. They do not know which building is being targeted or when this bombing is supposed to take place, and they are anxious to get more information. Then they intercept a message stating that, if the bombing proceeds as planned, the odds that The Crusher will be killed in the blast are 3 to 1.

What should now be Carla's credence that the bombing will take place? Let "A" stand for "the bombing will take place as planned" and "B" for "The Crusher will be killed in the blast." Carla really has nothing to go on. As far as she knows, $Cr(\neg A)$ could take any value after her team's find. This scenario is very different from the one in which Jeffrey type of information is supplied and in which equation (2) can be used without any knowledge of the contents of any of the B_i because, in that case, minimizing any α-divergence would lead to the same result. In Carla's case, however, different α-divergences might give different results, and it is, therefore, important for her to know which one she should

use. But how should she choose the appropriate functional? The relative entropy is the preferred classical choice, but it would lead to an increase in $Cr(\neg A)$ that might be completely inappropriate if the White Vipers consider The Crusher to be an enemy. Adams conditioning might be a good choice too, but it would leave $Cr(\neg A)$ unaltered, which might be inappropriate as well. It might be easy to classify this particular story and find its place in the hypothetical catalog, but assigning it the appropriate functional to minimize will not be.

Fourth, we might consider a simpler catalog, one that has a small number of categories such as "$Cr(\neg A)$ is larger than $Cr_0(\neg A)$" or "$Cr(\neg A)$ is definitely not larger than $Cr_0(\neg A)$," with "A" referring to the antecedent. A handful would suffice to describe whether there is a difference between $Cr(\neg A)$ and $Cr_0(\neg A)$ and, if so, whether that difference is positive or negative. There might even be a category appropriate for Carla the FBI agent when nothing is known about the difference between $Cr(\neg A)$ and $Cr_0(\neg A)$. But being able to extract the direction of the change does nothing for determining the size of the change. For the latter, a specific functional needs to be selected, which reintroduces the problem discussed in the preceding paragraphs.

In the absence of a well-defined functional to minimize, the agent might step back, consider **F** as a whole, and use theorem 1 to at least bound her posterior credence in $\neg A$. These bounds are independent, after all, of what particular functional was used. They produce a vague posterior credence, of course, but the range determined by the bounds will contain the correct credence appropriate to whatever details and nuances are contained in the scenario the agent is confronted with. This response to a Judy Benjamin problem assumes that there is a correct posterior credence, even though, at present, we lack the semantic sophistication to fully determine it. Bounds are then the best we can do, at least until we devise better tools. We could even improve on the bounds if the story is such that an increase or a decrease in $Cr(\neg A)$ can be excluded. In that case, the range of possible credences is bounded on one side by $Cr_0(\neg A)$, thereby improving the accuracy of the estimate.

3.2. Indeterminate Posterior Credences.   I would like to suggest, however, that this vagueness is constitutional of Judy Benjamin type of information rather than due to present limitations of our logic tool set. The Judy Benjamin problem is almost a Jeffrey problem because only one posterior credence, that of $\neg A$, is missing. If it were available, the posterior credences of all the members of the partition $\{\neg A, AB_1, \ldots, AB_n\}$ would be known, and we could use Probability Kinematics. In that case, there would be no vagueness, even when considering all possible α-divergences, because they all give the same answer. This suggests that the Judy Benjamin problem has incompletely specified information—for the purpose of up-

dating credences, that is—and that we cannot expect a single well-defined posterior credence—in other words, that the Judy Benjamin problem suffers from being underdetermined rather than from vagueness. This being underdetermined is fundamental and is in fact suggested by the definition of a conditional credence: $Cr(B_i|A) =_D Cr(B_iA)/Cr(A)$. If $Cr(B_i|A)$ is not equal to $Cr_0(B_i|A)$, the numerator or the denominator or both need to be changed. Merely providing a new value for the conditional credence, however, does not tell us which one of those possibilities is most appropriate. Any value for $Cr(\neg A)$ in [0, 1) can be assumed, with $Cr(B_iA) = q_i(1 - Cr(\neg A))$ completing the posterior credences of the members of the partition $\{\neg A, AB_1, \ldots, AB_n\}$.

Rather than insist that we find a unique update procedure for this less than Jeffrey type of information so we can construct a determinate posterior credence for $\neg A$, we should accept that the Judy Benjamin information is incomplete. The posterior credence $Cr(\neg A)$ is missing and the appropriate response is to assume that it can take any value within whatever bounds we can establish. We should determine, therefore, all possible posteriors that are compatible with the information that has been provided, but with the additional requirement that they be obtained through rational means of updating.

To be precise, let K be the set of possible values of $Cr(\neg A)$, given the prior credence Cr_0 and the Judy Benjamin information $\{Cr(B_i|A)\}$. The set $K = [l, u]$, in which l and u are the bounds established in theorem 1. It contains $Cr_0(\neg A)$, as we have seen. If we accept, as I propose we should, that the Judy Benjamin problem has no determinate posterior credence function, then it behooves us to accept

> **Postulate 1.**   The posterior for Judy Benjamin type of information is the set of all credence functions defined by equation (11) with $Cr(\neg A)$ in (a subset of) K.

The posterior for a Judy Benjamin problem as a set of credence functions is very similar to the indeterminate credences proposed by, for example, Levi (1974) and defended more recently by Joyce (2010) and Hájek and Smithson (2012). There is an important difference between the two proposals, however, Levi and Joyce advocate the set of credence functions that are compatible with whatever constraints are available, augmented perhaps with some other principles that are deemed to be appropriate. That set would be considerably larger than the one proposed in postulate 1 because any value of $Cr(\neg A)$ in [0, 1] would lead to an acceptable credence function if compatibility with the Judy Benjamin constraints was the only standard. I propose instead that we limit ourselves to those credence functions that can be obtained by a rational update procedure from the prior credence function and the avail-

able information. If a rational update procedure corresponds to minimizing an α-divergence, then that means choosing $Cr(\neg A)$ from K, rather than from $[0, 1)$.

The proposal has a variety of arguments in its favor. First, it is certainly the epistemologically most prudent answer in the case of Carla, the FBI agent. Carla knows nothing about the identity of The Crusher or the date and target of the bombing. When she learns that $Cr(B|A) = 3/4$, she has no reason to suspect that her credence in $\neg A$ will increase rather than decrease. The best she can do is consider all possible ways of determining the posterior credence function for all α-divergences. She will not arrive at a unique value for $Cr(\neg A)$, but, at least, her posterior credence will not out-infer her available evidence.

Adding details to the story will not diminish the strength of this argument. Such additional details may reduce the set of possible values of $Cr(\neg A)$ from K to some subset of K and may even lead to a subset with a single member (as in the canceled match story). Harry's story, however, is not in that category. It seems plausible that Harry's credence that Tom will accept the job offer increases, but that does not give Harry sufficient information to choose a unique divergence to minimize. Choosing any specific divergence would out-infer Harry's evidence, even though the details in the story are specific enough to exclude divergences that can only decrease $Cr(A)$ (such as the relative entropy).

Second, the Judy Benjamin posterior set will, in general, have more than one credence function, but this set will still be useful because, indeterminate as it is, it has a number of nontrivial properties similar to those that are associated with Probability Kinematics. To make such a comparison meaningful, we first have to establish what is meant by a property of a posterior that happens to be a nontrivial set of credence functions. I indicate the posterior set by **Cr** and the prior set (which consists of just the prior credence Cr_0) by **Cr$_0$**. The set **Cr** is then said to have a certain property if all of the credence functions it contains have that property (Levi 1974; Joyce 2010). All the properties listed below follow immediately from the construction of **Cr**, that is, from applying Jeffrey Conditionalization with $Cr(\neg A)$ ranging over the members of K.

1. **Cr**$(B_i|A)$ is as stipulated in the Judy Benjamin information; that is, the posterior set meets the goals set by the Judy Benjamin problem.
2. If $Cr(B_i|A) = \mathbf{Cr_0}(B_i|A)$ for all i, then **Cr** = **Cr$_0$**.
3. If $\mathbf{Cr_0}(E) = 1$, then **Cr**$(E) = 1$; if $Cr(\neg A) > 0$ and $q_i > 0$ for all i, then **Cr**$(E) > 0$ if $\mathbf{Cr_0}(E) > 0$.
4. If $\mathbf{Cr_0}(C|D) = 1$, then **Cr**$(C|D) = 1$ for all C and D such that $\mathbf{Cr_0}(D) > 0$.

5. If E is independent of the basic partition under $\mathbf{Cr}_0$ (i.e., if $\mathbf{Cr}_0(E|\neg A) = \mathbf{Cr}_0(E|AB_i) = \mathbf{Cr}_0(E)$ for each i), then E is also independent of that partition under $\mathbf{Cr}$, and $\mathbf{Cr}(E) = \mathbf{Cr}_0(E)$.

The first, third, and fourth properties follow immediately from equation (11). The fifth one is an immediate consequence of rigidity. The second property implies that, if the Judy Benjamin information is such that nothing new is learned, the posterior equals the prior, which was already mentioned in section 2. The third property guarantees that the Judy Benjamin posterior will have all the certainties that were present in the prior and that it will not have additional certainties unless they were forced on it by the Judy Benjamin information. The final two properties imply that doxastic implications and independence are preserved. Note, however, that there is a limit to the extent to which independence is preserved. It need not be true that, if D and E are independent under $\mathbf{Cr}_0$, they are also independent under $\mathbf{Cr}$.

3.3. Some Possible Objections. I next consider some possible objections to this account. First, it is unclear why rational ways of determining solutions to the Judy Benjamin equation should be restricted to minimizing α-divergences or, for that matter, minimizing functionals at all. Minimizing functionals is plausible because the new information comes in the form of constraints on the posterior credence function, and it is then reasonable to look for such functions that (*a*) meet the constraints and (*b*) are otherwise not too far away, in some sense, from the prior credence function. I defended the use of the family of α-divergences when I introduced them, but there might be other suitable families of update mechanisms. Uffink (1995) has shown that only α-divergences will meet the Shore and Johnson axioms, but, of course, that argument is only as strong as the justification of those axioms. But even if another family would turn out to be more appropriate, it is not implausible that the same results would emerge. After all, the sole role of the members of the family is to determine $\mathrm{Cr}(\neg A)$. Once that value has been determined, calculating the posterior credence functions just uses Probability Kinematics. The sole role of the family as a whole is then only to given bounds on the possible values of $\mathrm{Cr}(\neg A)$.

That this consideration is not mere speculation is demonstrated by the family of f-divergences (Ali and Silvey 1966; Csiszár and Shields 2004):

$$U_f(Cr,\ Cr_0) =_D \sum Cr_0(L_i) f\left(\frac{Cr(L_i)}{Cr_0(L_i)}\right), \qquad (13)$$

in which the kernel f is a nonnegative function on $(0, \infty)$ with an everywhere-defined, continuous, and monotonically increasing derivative g and is such that $f(1) = g(1) = 0$. Minimizing members of this family is a plausible way of

obtaining posterior credence functions because many f-divergences have been proposed and studied in the literature as measures of distance (Liese and Vajda 2006) between different probability functions. Some examples are the relative entropy $(f(x) = x\ln(x) - x + 1)$ and the inverse relative entropy $(f(x) = -\ln(x) + x - 1)$ used by Douven and Romeijn (2011) to effect Adams conditioning.[1] Others are the Hellinger distance $(f(x) = (\sqrt{x} - 1)^2)$ and the Kagan distance $(f(x) = (x - 1)^2/x)$. Furthermore, since g is monotonically increasing and equal to 0 when its argument is equal to 1, f has a minimum in $(0, \infty)$ only when its argument is equal to 1. This minimum is set to 0 by requiring that $f(1) = 0$. Consequently, U_f is nonnegative and equal to 0 only when $Cr(B_i) = Cr_0(B_i)$ for all i. Finally, U_f gets penalized, so to speak, increasingly severely for increasing deviations of Cr from Cr_0 because f is a convex function of x.

This family is different from that of the α-divergences. Nevertheless, the conclusions that can be drawn from both families are practically the same. If the new information is of the Jeffrey type, Probability Kinematics results, regardless of which f-divergence was used. The Judy Benjamin equation for the calculation of $Cr(\neg A)$ is replaced by a nonlinear equation whose solution is unique if it exists (app. A). Given $Cr(\neg A)$, the posterior credence function is provided by equation (11) for both α-divergences and f-divergences. The most important result, however, is that f-divergences, too, have nontrivial bounds on the possible values of $Cr(\neg A)$ and that those bounds are the same as those for α-divergences (app. B).

Second, it is disconcerting, to say the least, that precise information and a determinate prior can lead to an indeterminate posterior. I share this unease, but an indeterminate posterior does not imply that the agent is now licensed to hold her beliefs with whatever strengths she desires. Notice, for example, that the posterior is not indeterminate for all beliefs. If some belief is independent of the members of the Judy Benjamin partition, its posterior credence remains determinate and equal to the prior credence. Furthermore, even if the posterior credence becomes indeterminate, it may still contain nontrivial information. In the original Judy Benjamin case, for example, Judy's credence that she is now in the headquarters area of the Red Army does increase in the sense that the lower bound on her posterior credence is larger than her prior credence that she is in that area.

The phenomenon of a determinate credence function becoming (partially) indeterminate is analogous to that of dilation (Seidenfeld and Wasserman 1993). In the latter case, an already indeterminate prior becomes even more indeterminate, in a precisely defined manner, upon the acquisi-

---

1. The added term x − 1 ensures that f has the correct properties. It disappears when the sum in eq. (13) is taken.

tion of new information. In the coin toss experiment (Seidenfeld and Wasserman 1993), for example, two coins are fair but the conjunction of both coins showing heads is indeterminate (between 0 and 0.5). Updating on Judy Benjamin type of information is more interesting because the prior credence is not indeterminate (or, at least, need not be), while the posterior may be, depending on the details of the background story. But just as classical dilation has a cause—indeterminate priors—the Judy Benjamin analogue to dilation has a cause as well: indeterminate updating. In that sense, the present phenomenon of Judy Benjamin type of information giving rise to an indeterminate posterior is simply another form of dilation.

4. Beyond Judy Benjamin. The standard credential update process is one in which the prior credence function is updated to a posterior credence function upon the acquisition of some new information. The various examples presented in this article make it clear that a single update mechanism cannot cover all subtleties and varieties contained in Judy Benjamin type of information. The only appropriate way then of dealing with that type of information is to consider all possible update mechanisms or, at least, all members of some large family of mechanisms that can plausibly be used to determine posterior credences. The interpretation defended in this article is to consider the indeterminateness fundamental: a given Judy Benjamin problem may have additional information that can be used to restrict the family of mechanisms, but, in general, it does not. On this view, the correct posterior in the presence of Judy Benjamin information is the set of posteriors obtained by employing (a subset of) all mechanisms in the family of plausible update mechanisms. The suggestion made in this article is that the family of update mechanisms is that of minimizing α-divergences, even though, as I mentioned in the preceding section, other families, such as that of f-divergences, might also be used.

We might ask why Judy Benjamin type of information is special as compared to Jeffrey type of information in that the former requires minimizing all divergences, leading to a set of posterior credence functions, while Probability Kinematics (or Bayesian Conditionalization for that matter) produces a single such function. But there really is nothing special. We could, and I propose we should, insist that modifying credences upon the delivery of new information is always done (at least, when the new information is a convex constraint on the posterior credence function) by minimizing all members of the family of divergences under the constraints appropriate to the new information—in other words, that updating always leads to a posterior set of credence functions. Minimizing all members of the family would constitute a standard credential update process if they all produced the same posterior credence function. It does so for Jeffrey information but not for Judy Benjamin information.

The Judy Benjamin problem as given by van Fraassen has served as the prime example of why minimizing the relative entropy cannot always be considered to be the correct way of determining a posterior credence function. That minimizing the relative entropy in the presence of Jeffrey information leads to Probability Kinematics is not a sufficient rational for continuing to use that functional exclusively, because any α-divergence (or f-divergence, for that matter) produces the same result. It is to be expected that minimizing the relative entropy for even more general types of information than the Judy Benjamin one may also produce implausible posterior credences in suitably chosen examples and that the general problem of constructing a posterior credence when new information of any type becomes available might better be done by constructing a set of posterior credence functions.

As a very general type of new information, we might, for example, consider posterior expectations of simple random variables, where "simple" means that the random variable has only a finite number of possible values. Jeffrey and Judy Benjamin types of information are a special case of that general form. The general solution proceeds by first determining $\mathbf{Cr}(.)$ for all members of the partition (as defined by the finite set of possible values of the random variable) by minimizing each member of the family of functionals and then using Probability Kinematics to obtain the general posterior credence set. As in the Judy Benjamin case, additional information contained in the accompanying story may reduce the set of functionals to be minimized, but, in general, the result will be a range of possible credence values for each member of the partition. Such a construction replaces updating by minimizing the relative entropy and does not suffer the limitations of the latter because it uses all possible members of the family and not just the single one whose choice has been motivated largely by past successes in restricted areas of inquiry.

Appendix A

Existence of Solutions of the Judy Benjamin Equation for f-Divergences

When minimizing f-divergences, the appropriate Judy Benjamin equation is

$$g(r(\neg A)) = \sum q_i g\left(r(A)\frac{q_i}{p_i}\right). \tag{A1}$$

The left-hand side is a continuous and monotonically increasing function of $Cr(\neg A)$, and the right-hand side is a continuous and monotonically decreasing function of the same variable. Therefore, there will be a solution if

the right-hand side is larger than the left-hand side when $\mathrm{Cr}(\neg A) = 0$ and vice versa when $\mathrm{Cr}(\neg A) = 1$. If $g(r(\neg A))$ becomes infinitely negative when $\mathrm{Cr}(\neg A)$ goes to 0, as it does for many f-divergences, define $g(0) = -\infty$. The first part of the requirement is met when

$$g(0) < \sum q_i g\left(\frac{q_i}{\mathrm{Cr}_0(AB_i)}\right). \tag{A2}$$

The second part is met when $g(1/\mathrm{Cr}_0(\neg A)) > g(0)$, but that is always the case because $g(0)$ is negative (possibly $-\infty$), while $1/\mathrm{Cr}_0(\neg A)$ is larger than 1 so $g(1/\mathrm{Cr}_0(\neg A))$ is positive. There is no solution when equation (A2) does not hold.

Appendix B

Bounds on the Solutions of the Judy Benjamin Equation for f-Divergences

First, we rewrite the Judy Benjamin equation as

$$\sum q_i g\left(r(A)\frac{q_i}{p_i}\right) - g(r(\neg A)) = 0. \tag{B1}$$

The ratios q_i/p_i can take all values in $(0, \infty)$. As in the main text, let s be the index of the smallest ratio and l that of the largest one. At least one term in equation (B1) is negative and at least one is positive. And g is monotonically increasing, so $g(r(A)q_s/p_s) - g(r(\neg A))$ is the most negative term and $g(r(A)q_l/p_l) - g(r(\neg A))$ the most positive one, or

$$\frac{q_s}{q_s} \le \frac{r(\neg A)}{r(A)} \le \frac{q_l}{p_l}. \tag{B2}$$

Equation (B2) is a string of implicit inequalities for $\mathrm{Cr}(\neg A)$. The inequalities can be made explicit by using the abbreviations $\mathrm{Cr}(A) = \mathrm{Cr}_0(A)(1 + \varepsilon)$ and $q_i = p_i(1 + \delta_i)$. Equation (B2) then becomes

$$1 + \delta_s \le \frac{(1 - Q_\varepsilon)}{(1 + \varepsilon)} \le 1 + \delta_1, \tag{B3}$$

with $Q = \mathrm{Cr}_0(A)/\mathrm{Cr}_0(\neg A)$. After some straightforward algebra, equation (B3) becomes

$$\frac{\mathrm{Cr}_0(\neg A)\delta_s}{1 + \mathrm{Cr}_0(\neg A)\delta_s} \le -\varepsilon \le \frac{\mathrm{Cr}_0(\neg A)\delta_1}{1 + \mathrm{Cr}_0(\neg A)\delta_1}, \tag{B4}$$

and, after some more algebra and undefining δ_s, δ_1, and ε, we finally find that

$$\frac{Cr_0(\neg A)q_s}{Cr_0(A)p_s + Cr_0(\neg A)q_s} \leq Cr(\neg A) \leq \frac{Cr_0(\neg A)q_1}{Cr_0(A)p_1 + Cr_0(\neg A)q_1}. \quad (B5)$$

REFERENCES

Ali, S. M., and S. D. Silvey. 1966. "A General Class of Coefficients of Divergence of One Distribution from Another." *Journal of the Royal Statistical Society* B 28:131–42.

Bradley, Richard. 2005. "Radical Probabilism and Bayesian Conditioning." *Philosophy of Science* 72:342–64.

Csiszár, Imre, and Paul C. Shields. 2004. "Information Theory and Statistics: A Tutorial." *Foundations and Trends in Communications and Information Theory* 1:417–528.

Diaconis, Persi, and Sandy L. Zabell. 1982. "Updating Subjective Probability." *Journal of the American Statistical Association* 77:822–30.

Douven, Igor, and Jan-Willem Romeijn. 2011. "A New Resolution of the Judy Benjamin Problem." *Mind* 120:637–70.

Grove, Adam J., and Joseph Y. Halpern. 1997. "Probability Update: Conditioning vs. Cross-Entropy." In *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence*, ed. Dan Geiger and Prakash Shenoy. San Francisco: Kaufmann.

Hájek, Allen, and Michael Smithson. 2012. "Rationality and Indeterminate Probabilities." *Synthese* 187:33–48.

Howson, Colin, and Peter Urbach. 1993. *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Chicago: Open Court.

Jaynes, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Jeffrey, Richard C. 1983. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.

Joyce, James M. 2009. "The Development of Subjective Bayesianism." In *Handbook of the History of Logic*, vol. 10, *Inductive Logic*, ed. Dov Gabbay, Stephan Hartmann, and John Woods, 415–76. Amsterdam: North-Holland.

———. 2010. "A Defense of Imprecise Credences in Inference and Decision Making." *Philosophical Perspectives* 24:281–323.

Levi, Isaac. 1974. "On Indeterminate Probabilities." *Journal of Philosophy* 71:391–418.

Liese, Friedrich, and Igor Vajda. 2006. "On Divergences and Informations in Statistics and Information Theory." *IEEE Transactions on Information Theory* 52:4394–4412.

Lukits, Stefan. 2014. "The Principle of Maximum Entropy and a Problem in Probability Kinematics." *Synthese* 191:1409–31.

Seidenfeld, Teddy. 1986. "Entropy and Uncertainty." *Philosophy of Science* 53:467–91.

Seidenfeld, Teddy, and Larry Wasserman. 1993. "Dilation for Sets of Probabilities." *Annals of Statistics* 21:1139–54.

Shore, John E., and Rodney W. Johnson. 1980. "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy." *IEEE Transactions on Information Theory* 26:26–37.

Uffink, Jos. 1995. "Can the Maximum Entropy Principle Be Explained as a Consistency Requirement?" *Studies in History and Philosophy of Modern Physics* B 26:233–61.

van Fraassen, Bas C. 1981. "A Problem for Relative Information Minimizers in Probability Kinematics." *British Journal for the Philosophy of Science* 32:375–79.

———. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.

van Fraassen, Bas C., R. I. G. Hughes, and Gilbert Hartman. 1986. "A Problem for Relative Information Minimizers, Continued." *British Journal for the Philosophy of Science* 37:453–63.

Williams, P. M. 1980. "Bayesian Conditionalisation and the Principle of Minimum Information." *British Journal for the Philosophy of Science* 31:131–44.