A New Resolution of the Judy Benjamin Problem

Igor Douven
Institute of Philosophy, University of Leuven
`igor.douven@hiw.kuleuven.be`

Jan-Willem Romeijn
Faculty of Philosophy, University of Groningen
`j.w.romeijn@rug.nl`

Abstract

Van Fraassen's Judy Benjamin problem has generally been taken to show that not all rational changes of belief can be modelled in a probabilistic framework if the available update rules are restricted to Bayes's rule and Jeffrey's generalization thereof. But alternative rules based on distance functions between probability assignments that allegedly can handle the problem seem to have counterintuitive consequences. Taking our cue from a recent proposal by Bradley, we argue that Jeffrey's rule can solve the Judy Benjamin problem after all. Moreover, we show that the specific instance of Jeffrey's rule that solves the Judy Benjamin problem can be underpinned by a particular distance function. Finally, we extend the set of distance functions to ones that take into account the varying degrees to which propositions may be epistemically entrenched.

Often a learning experience makes us certain of a proposition we were previously uncertain of. But, as Jeffrey famously pointed out, not all learning is like that. Sometimes learning consists in a person's becoming *more* certain of some proposition or propositions without her becoming *entirely* certain of any of them. To cite a well-worn example, a glimpse of a cloth by candlelight may make one more certain that the cloth is blue without making one certain that it is blue (Jeffrey [1983:165 f]). Both kinds of learning can be readily handled by probabilistic means, the first by dint of Bayes's rule (also known as "conditionalization"), the second by dint of Jeffrey's generalization of that rule. Van Fraassen's [1981] Judy Benjamin problem is generally taken to show that there are types of learning experiences that fall outside the scope of both Bayes's and Jeffrey's rule.[1]

While we do not want to contest the general point that there exist cases of learning that cannot be adequately dealt with by either of these rules, we do disagree with van Fraassen that the Judy Benjamin problem forces us to resort to a probabilistic update rule beyond the aforementioned ones. Taking our cue from a recent proposal by Bradley [2005], we will argue not only that the learning event described in van Fraassen's problem poses no special difficulties to Jeffrey's rule, but also that this rule permits a more satisfactory solution to the problem than any of the alternative

---

[1] See also van Fraassen [1989:342 ff] and van Fraassen, Hughes, and Harman [1986].

rules that have been discussed in this context. Indeed, we will point to a specific application of Jeffrey's rule that yields a solution satisfying all the constraints that are pre-theoretically valid in the Judy Benjamin story and that moreover can be underpinned by a variant of one of the aforementioned alternative rules; this variant rule, like the other rules, counsels a particular type of distance minimization between the pre- and post-update probability assignments under the given constraints.

We take this to be a surprising result, given that the case of Judy has already invited comparison between several distance functions, and given that the new distance function to be presented would seem to be a rather obvious one. But we want to go beyond this result and consider a wider range of update rules. For, as will be seen, Jeffrey's rule and the associated distance function provide only a partial answer to the question how one ought to update one's degrees of belief in response to the learning of an indicative conditional, a question that, on our analysis, occupies center stage in the Judy Benjamin problem.[2] While to our mind the partial answer suffices for tackling the designated problem, it is of independent interest to have a more general probabilistic account of what it is to learn an indicative conditional, which is still lacking from the literature. In the final section of the paper, we provide the beginnings of such an account.

1. The Judy Benjamin problem. In the Judy Benjamin problem, a soldier—the fictional character Judy Benjamin—is dropped with her platoon in an area that is divided in two halves, Red territory (R) and Blue territory ($\neg R$), respectively, with each territory in turn being divided in equal parts, Second Company area (S) and Headquarters Company area ($\neg S$). Since the platoon was dropped more or less at the center, Judy Benjamin deems it equally likely that they are in one area as that they are in any of the others, that is, where $\Pr_0$ is her initial degrees-of-belief function,[3] $\Pr_0(R \wedge S) = \Pr_0(R \wedge \neg S) = \Pr_0(\neg R \wedge S) = \Pr_0(\neg R \wedge \neg S) = 1/4$. They then receive the following radio message:

(1)   I can't be sure where you are. If you are in Red territory, the odds are
         $3 : 1$ that you are in Headquarters Company area.

After this, the radio contact breaks down. Supposing Judy accepts this message, how should she adjust her degrees of belief? It is obvious that at least (i) should hold for her new degrees-of-belief function $\Pr_1$:

  (i) her conditional degree of belief for being in Headquarters Company area ($\neg S$) given that she is in Red territory (R) is three times her conditional degree of belief for being in Second Company area (S) given R; as these conditional degrees of belief must sum to 1, it follows that $\Pr_1(\neg S \mid R) = 3/4$ and $\Pr_1(S \mid R) = 1/4$.

But (ii) and (iii) seem intuitively no less compelling as desiderata to be satisfied by her new degrees-of-belief function:

  (ii) none of her conditional degrees of belief given any proposition in $\{R \wedge S, R \wedge \neg S, \neg R\}$ has changed, that is, $\Pr_1(A \mid X) = \Pr_0(A \mid X)$ for all A, with $X \in \{R \wedge$

---

[2]See in the same vein Nayak et al. [1996].

[3]More precisely, $\Pr_0$ is a probability function that represents degrees of belief, but in this paper we will refer to such functions as degrees-of-belief functions.

$S, R \wedge \neg S, \neg R\}$; after all, (1) seems to contain no information that could warrant a change in those conditional degrees of belief;

(iii) $\text{Pr}_1(\neg R) = \text{Pr}_0(\neg R) = \text{Pr}_0(\neg R \wedge S) + \text{Pr}_0(\neg R \wedge \neg S) = 1/2$, for neither does (1) seem to contain any information relevant to whether she is in Red rather than in Blue territory.

The question then is whether there is any probabilistic update rule that respects (i)–(iii).

A preliminary question to be asked is whether there is any probabilistic update rule that applies to (1) *at all*. It is clear that neither Bayes's rule nor Jeffrey's rule was devised for this kind of learning event, and it is generally thought that neither of them is applicable to (1).[4] A rule that does apply, and that van Fraassen discusses—while stopping short of endorsing it—is the rule he calls "INFOMIN." Given, first, a prior degrees-of-belief function Pr_0 defined on an algebra $\mathcal{F}$, second, the partition $\{A_i\}$ of minimal elements (i.e., strongest consistent propositions) of $\mathcal{F}$ such that $\text{Pr}_0(A_i) > 0$ for all i, and third, constraints π to be imposed on the posterior degrees-of-belief function, this rule demands that we adopt the posterior degrees-of-belief function $\text{Pr}_1 \in \{\text{Pr} \mid \text{Pr} \text{ satisfies } \pi\}$ that minimizes the following "relative entropy" function:

$$(2) \qquad \text{RE}(\text{Pr}_0, \text{Pr}_1) \quad = \quad \sum_i \text{Pr}_1(A_i) \log \frac{\text{Pr}_1(A_i)}{\text{Pr}_0(A_i)}.$$

Informally put, suppose we have collected all degrees-of-belief functions defined on a given language in a space of functions, so that changes from one belief state to another can be represented as moves from one point in the space of functions to another. RE then defines something that is in certain respects similar to a distance function on this space, thereby allowing us to regard some belief changes as being more radical than others, depending on the distance that the point representing the new belief state has to the point representing the old belief state. The normative principle that subsequently determines the update is that changes in belief states must respect the constraint or constraints imposed by the receipt of new information, but must otherwise be maximally prudent and conservative. This can be spelled out as the requirement that we minimize RE under a set of constraints. In van Fraassen, Hughes, and Harman [1986], two further rules, similar to INFOMIN, are discussed that also apply unambiguously to (1).

As it turns out, application to the Judy Benjamin case of any of the said rules leads to a violation of desideratum (iii). For instance, according to INFOMIN Judy's new degree of belief for being in Blue territory should be (approximately) .532. But instead of looking for a formal rule that does satisfy all of (i)–(iii), van Fraassen and his coauthors seek to downplay the importance of (iii).[5] Suppose—they argue, giving

---

[4]Though Grove and Halpern [1997] argue that Judy can deal with (1) by means of Bayes's rule. In order to do so, she is supposed to extend the algebra on which her degrees-of-belief function is defined by including the event that the radio officer tells Judy a conditional degree of belief, and perhaps further specifics of the protocol that the radio operator is using to formulate his messages to her. Note, however, that our understanding of a conditional is not normally cast in terms of the exact reasons for stating the conditional, or in terms of the specifics of the protocol the utterer is using. It is often cast in terms of a blanket or unspecific idea of how the antecedent and consequent are related. In other words, it is doubtful that the requisite extension of the algebra is generally available.

[5]Williamson [2009] makes the curious remark that by objecting to a rule that it fails to respect (iii) one begs the question, as this objection "assumes that one should continue to equivocate between red

Peter Williams credit for the argument—that the radio officer had said:

(3)   If you are in Red territory, then the odds are $0 : 1$ that you are in Second
        Company area.

Then he would effectively have told the platoon that it is *not* in Red Second Company area. And this, the authors contend, would have allowed Judy to conditionalize on $\neg R \vee \neg S$, with the obvious result that $\mathrm{Pr}_1(\neg R) = 2/3$ and $\mathrm{Pr}_1(R) = 1/3$, where Pr_1 is again her new degrees-of-belief function after the update. The second message seems not essentially different from the first:[6] they both contain information relevant to how one ought to grade the possibilities consistent with one's being in Red territory, but not about how one ought to grade the more general possibility of being in Red territory and that of being in Blue territory relative to one another; differently put, there seems to be nothing special about odds $0 : 1$ as opposed to odds $3 : 1$ (or any other odds). Yet the receipt of (3) leads to a change in Judy's degree of belief for being in Blue territory. But then not too much should be made of the fact that, given the update rules van Fraassen, Hughes, and Harman discuss, Judy's degree of belief for being in Blue territory changes after the receipt of (1) as well.

What van Fraassen and his coauthors consider to be more troubling is that we seem to have no principled grounds for choosing between the update rules they discuss, as these rules all satisfy (i) and (ii) and also satisfy any more formal ("symmetry") requirement that one might plausibly want to impose, and as, moreover, quasi-empirical considerations (comparison with respect to performance by means of computer simulations) give a rather mixed result. This is troubling, given that these rules do not generally agree in their outputs; they yield for instance slightly different answers as to what Judy's degrees of belief in the propositions of interest ought to be after an update on (1). The problem of selecting a unique update rule of the kind at issue is regarded to be still open today.[7]

Here, we want to argue for two related claims on these update rules. First, we maintain that the problem of choosing the correct distance function is to some extent independent of the Judy Benjamin problem, as the latter calls neither for INFOMIN nor for any similar rule. As intimated earlier, we think that Jeffrey's rule, even if not devised for inputs like (1), can handle them perfectly well, and indeed more naturally than the rules van Fraassen and others consider in this context: Jeffrey's rule yields a solution to the Judy Benjamin problem that satisfies not only (i) and (ii) but also (iii). Of course, in view of the above argument against (iii), this may not seem very telling. However, in the next section we will explain why, as we believe, van Fraassen, Hughes, and Harman have been too quick in dismissing (iii).

Our second claim directly concerns the choice of a distance function. If instead of using Jeffrey's rule one wants, for whatever reason, to use an update rule based on a distance function between probability assignments, then a perfectly respectable

---

and blue if the evidence permits" (p. 7 of manuscript). This is curious as it is entirely unclear how one could beg any questions simply by registering one's intuitive verdict (as opposed to giving an argument) that (1) contains no information relevant to whether Judy is in Red rather than in Blue territory, and that therefore her probabilities for being in Red territory and, respectively, Blue territory should not change.

[6]Here we are ignoring the fact that zero probability is not the same as logical truth: events with measure zero may nevertheless happen.

[7]See on this also Joyce [1999:215 ff], who discusses some further rules still. Uffink [1995] contains a systematic critique of various attempts to justify INFOMIN.

such rule is available that satisfies all of (i)–(iii). We state this rule in Section 4, and we prove that it has the desired property.

2. Updating on conditionals: the simple view. Why should the learning of (3) lead Judy to conditionalize on $\neg R \vee \neg S$, as the aforementioned authors suppose she does, and ought to do, in that event? They do not say, but the only at least prima facie plausible answer we can think of on their behalf is this: "Because the indicative conditional (3) has the truth conditions of the material conditional, $R \supset \neg S$, or equivalently, $\neg R \vee \neg S$. So, if we learn the former we should conditionalize on the latter." Whether or not this is what they would really want to answer in response to our question, the view that the truth conditions of the indicative conditional are those of the corresponding material conditional,[8] and learning the indicative conditional amounts to conditionalizing on the material conditional, is interesting in its own right—if for no other reason, then for its alluring simplicity. But is it *correct*? We are raising two questions here, one about the truth conditions of indicative conditionals, the other about how one should update on such conditionals, supposing an indicative conditional to have the truth conditions of the corresponding material conditional. To make the second question more general: How is one to update on an indicative conditional—whether or not it has the truth conditions of the material conditional— and in particular, is there a defensible update rule that agrees with how, according to van Fraassen, Hughes, and Harman, Judy should respond to the learning of (3)?

Famously, there is little consensus about the semantics of indicative conditionals.[9] The major divide among semanticists is between those who hold that such conditionals have truth conditions and those who deny this. The latter typically do think that conditionals have assertability and acceptability conditions. Most popular here is the view according to which a conditional is assertable by/acceptable to someone iff her degree of belief in the consequent given the antecedent is high.[10] But even among those who do think that conditionals have truth conditions, the material conditional account, according to which a conditional has the same truth conditions as the corresponding material conditional, though once popular, is currently very much out of favor. The first part of the answer we gave above on behalf of van Fraassen and his coauthors is thus anything but uncontroversial. Strictly speaking, however, one could hold that updating on a conditional is updating on the corresponding material conditional even though one denies that the former has the truth conditions of the latter; such a position would leave something to be explained, but it is consistent. This makes the second question of how we ought to update on a conditional the more important one.

---

[8]At least when the indicative conditional's antecedent and consequent are both propositions in the algebra on which we take people's degrees-of-belief functions to be defined. Plausibly, in the Judy story this condition is supposed not to hold for the conditional in (1). (Grove and Halpern [1997] is aimed exactly at repairing this.)

[9]From here on, by the unqualified use of "conditional" we refer to indicative conditionals. We should also say that, throughout, we will be concerned exclusively with so-called simple conditionals, that is, conditionals which do not contain embedded conditionals. If, as for instance Jackson [1987:127–137] and Edgington [1995:382 ff] argue, by far the most non-simple conditionals that we encounter in daily life can either be reduced to simple ones or be dismissed as being incomprehensible, it is not much of a restriction that we limit ourselves to simple conditionals.

[10]See, for instance, Adams [1975] and Edgington [1995] for influential defenses of this view.

With respect to this question, which will be central in the following, the situation is in a sense still a bit worse than "lack of consensus": it is a question that, for some reason or other, has received little to no attention in the literature, even though the answer to it would seem far from obvious given any of the better-known views on the semantics of conditionals.[11]

For instance, it might be thought that if conditionals do not have truth conditions, and thus do not express propositions, then probabilities will just have to be defined on sentences—which is possible, of course. However, that does not really solve the problem for those who deny that conditionals have truth conditions, for they cannot maintain that probabilities of conditionals are probabilities of truth. Indeed, they generally recognize that probabilities of conditionals are probabilities in name only and are really to be thought of as degrees of assertability/acceptability. But then in order to use Bayes's rule for updating on conditionals, we would have to be able to mix probabilities of truth and degrees of assertability/acceptability. We are unaware of any proposal of how to do this. In fact, it is not clear whether this idea of mixing makes sense at all.

Of more immediate concern is that even if we grant that a conditional has the same truth conditions as the corresponding material conditional, it is still unclear what updating on a conditional amounts to. It might initially be thought that if there is any position that has a straightforward answer to our question, it is the material conditional account. If, for any A and B, "If A, B" has the truth conditions of $\neg A \vee B$, why is the answer to the question not simply the one that was suggested above on behalf of van Fraassen et al., to wit, that to update on the former is to apply Bayes's rule to the latter? Closer inspection reveals that, even assuming that the material conditional account provides the right semantics for conditionals, the answer to the question about updating cannot be so simple.

First consider

Fact 2.1 *For all A, B such that $\Pr(A) > 0$ and $\Pr(B) > 0$, $\Pr(A \mid A \supset B) \leqslant \Pr(A)$; if, in addition, $\Pr(A) < 1$ and $\Pr(B \mid A) < 1$, then $\Pr(A \mid A \supset B) < \Pr(A)$.*[12]

Thus, supposing that the proper way to update on a conditional is to conditionalize on the corresponding material conditional, it follows that by learning a conditional, one can never become more certain of that conditional's antecedent and that, under very general conditions, one can only become less certain of the antecedent. However, we take the following example to be an incontrovertible case in which the probability of the antecedent of a conditional must remain invariant upon learning that conditional.

---

[11]Cf. Skyrms [1980:169]: "[W]e have no clear conception of what it might be to conditionalize on a conditional." But see also Adams [1994], where a logical model of updating on conditional information is presented. In this paper, Adams employs a metalanguage for probability statements in which conditionals can be formalized, and then derives some intuitive connections between these statements and probability assignments over the object language. As he admits, though, a full integration of levels is problematic (see also Adams [1998:147 n]). Grove and Halpern [1997] can be taken as a further development of Adams's proposal. However, their approach runs into its own problems; see footnote 4.

[12]The following is a slight modification and extension of a proof given in Williamson [2007:232 n] (see also Popper and Miller [1983]): $\Pr(A \supset B) = 1 - \Pr(A \wedge \neg B) = 1 - \Pr(A)\Pr(\neg B \mid A) = 1 - \Pr(A)\Pr(A \wedge \neg B \mid A) \geqslant 1 - \Pr(A \wedge \neg B \mid A) = \Pr(A \supset B \mid A)$. If $\Pr(B \mid A) < 1$, then $\Pr(\neg B \mid A) > 0$ and hence also $\Pr(A \wedge \neg B \mid A) > 0$, so if also $\Pr(A) < 1$, then "$\geqslant$" can be replaced by "$>$" in the foregoing. Fact 2.1 follows immediately from this, as, by probability theory, $\Pr(A \mid B) \leqslant (<) \Pr(A)$ iff $\Pr(B \mid A) \leqslant (<) \Pr(B)$ for all A and B such that $\Pr(A)$ and $\Pr(B)$ are positive. (Note that if $\Pr(B) > 0$, then automatically $\Pr(A \supset B) > 0$.)

Example 1 Sarah and Marian have arranged to go for sundowners at the Westcliff hotel tomorrow. Sarah feels there is some chance that it will rain, but thinks they can always enjoy the view from inside. To make sure, Marian consults the staff at the Westcliff hotel and finds out that in the event of rain, the inside area will be occupied by a wedding party. So she tells Sarah: "If it rains tomorrow, we cannot have sundowners at the Westcliff." Upon learning this conditional, Sarah sets her probability for sundowners *and* rain to zero, but she does not adapt her probability for rain.

For Sarah, neither her initial degree of belief for rain tomorrow, $\Pr_0(R)$, nor her initial degree of belief for having sundowners at the Westcliff, $\Pr_0(S)$, has an extreme value; nor is she initially certain that they will not have sundowners in the event of rain, that is, $\Pr_0(\neg S \mid R) < 1$. Hence, by Fact 2.1, and assuming conditionalization, we must have that $\Pr_1(R) = \Pr_0(R \mid A \supset B) < \Pr_0(R)$. But surely we do not want to say that Sarah should optimistically reduce her probability for rain after learning the conditional! No matter how keen she is on the sundowners, she should simply set the probability for having sundowners while it is raining (i.e., $R \wedge S$) to zero, meanwhile leaving her probability for rain as it is, precisely as (we said) she does.

We would like to point to another problem one faces if one wants to maintain that updating on "If R, $\neg S$" is updating on $\neg R \vee \neg S$. For this, we return to the story of Judy Benjamin. Suppose, after the platoon was dropped, the radio officer had said:

(4)   It is as likely that you are in Red territory as that you are in Blue territory, and if you are in Red territory, then you're in Headquarters Company area.

We venture that, whatever effect this message should have on Judy's belief system, the order of the conjuncts of the message should not matter. That is to say, if the officer had said:

(5)   If you are in Red territory, then you're in Headquarters Company area, and it is as likely that you are in Red territory as that you are in Blue territory,

then that should have had the same effect on Judy's belief system. Indeed, according to standard semantics, and assuming that conditionals have truth conditions (whether or not they are those of the corresponding material conditionals), the two messages have exactly the same truth conditions, and hence the same meaning. So how could learning one have a different effect on one's belief system than learning the other? One might try to argue that there is some pragmatic difference between the messages which could explain possible differences in the result of updating on the one rather than on the other, but we fail to see how any of the known pragmatic mechanisms could be successfully invoked here to show that there is a difference between (4) and (5). It should for instance be clear that the Gricean idea that the order of the conjuncts should reflect temporal order does not apply to these sentences: the conjuncts do not report events that have occurred in some definite order.

Now consider what happens if we identify updating on a conditional with updating on the corresponding material conditional and then update on (4) and, respectively, (5). For generality, just suppose that for the initial probability function $\Pr_0$, $\Pr_0(R) = \rho$ and thus $\Pr_0(\neg R) = 1 - \rho$, and say that $\Pr_0(R \wedge S) = \rho\sigma$ and $\Pr_0(R \wedge \neg S) = \rho(1 - \sigma)$,

7

with $0 < \sigma < 1$. Now Judy first updates—by means of Jeffrey conditionalization—on the first conjunct of (4). This yields $\Pr_1(R) = \Pr_1(\neg R) = 1/2$ and $\Pr_1(R \wedge \neg S) = (1 - \sigma)/2$. She then updates on the second part of that message, which—as per the targeted assumption—means that she conditionalizes on the proposition $\neg R \vee \neg S$. One easily calculates that this yields $\Pr_2(\neg R) = 1/(2 - \sigma)$ and $\Pr_2(R) = \Pr_2(R \wedge \neg S) = (1 - \sigma)/(2 - \sigma)$. Since $\sigma \neq 1$, the probability of being in Red territory does not equal that of being in Blue territory. Next suppose that she updates on the first conjunct of (5) first. Then we get $\Pr_1(\neg R) = (1 - \rho)/(1 - \rho\sigma)$ and $\Pr_1(R) = \Pr_0(R \wedge \neg S) = \rho(1 - \sigma)/(1 - \rho\sigma)$. Finally, she updates on the second conjunct of (5), again by means of Jeffrey conditionalization. This yields, of course, $\Pr_2(R) = \Pr_2(\neg R) = 1/2$. So now her degree of belief for being in Red territory does equal that for being in Blue territory!

Here it might be retorted that, once we admit Jeffrey conditionalization, we are committed to order dependence in any case. While this is true, it does not mean that we should not take measures to counteract such dependence whenever that dependence is intuitively unpalatable and we *can* take such measures (without abandoning Jeffrey's rule altogether); and, as we shall see later on, we can take such measures indeed. In addition to this, the case we are considering is not one in which Judy gets, more or less by coincidence, one piece of information before the other. The information is provided to her by someone in a particular order. If the order were supposed to matter, one would expect the radio officer to be careful in determining that order. But it seems that in asserting sentences of the sort we are considering here—that is, the radio messages (4) and (5)—we tend to be indifferent between asserting the conjuncts in one order and asserting them in the other order. And this is so because we expect that the order will not matter to what the addressee ends up believing.

Hence, while attractively simple, the thought that learning a conditional consists of conditionalizing on the corresponding material conditional appears doubtful.[13] Recall now what led us to these considerations. Van Fraassen, Hughes, and Harman sought to discredit, or at least mitigate the force of, the intuition that the learning of (1) should leave Judy's degree of belief for her being in Blue territory unchanged, by arguing that it is violated anyhow if we learn (3), which seems not relevantly different from (1). As we saw, however, the obvious argument underlying the claim about the learning of (3) is predicated on a semantics of conditionals as well as on a view on how we are to update on conditionals both of which are contentious. Of course, van Fraassen and his coauthors may have had some other argument in mind when claiming that the learning of (3) would allow Judy to conditionalize on $\neg R \vee \neg S$. And, quite apart from what argument they may have had in mind, there might, as a matter of fact, simply be *no* acceptable account of conditionals or updating on conditionals on which the learning of (1) or (3) leaves Judy's degree of belief for being in Blue territory unchanged. In that case, we should perhaps not give a lot of weight to the designated intuition indeed. In the following, however, we propose a probabilistic account of updating on conditionals that does allow us to solve the Judy Benjamin problem in an intuitively perfectly satisfactory way.

---

[13]What to say, from a Bayesian perspective, about the other major truth-conditional view on conditionals—basically, Stalnaker's possible worlds semantics—is quite unclear. Nolan [2003] hints, in note 41 of the paper, that given this semantics the probability of a conditional is determined by an imaging function in the sense of Lewis [1976]. However, he relegates elaboration of that hint to future work. What can already now be said is that the example he gives on p. 261, in which he assesses the probability of a given conditional, is utterly unconvincing.

3. Conditionals and Jeffrey conditioning. The bulk of our proposal, to be presented in this section, concerns the kind of case in which, intuitively, the learning of a conditional is or would be irrelevant to one's degree of belief for the conditional's antecedent. While this may be the normal case, the rule for updating on conditionals to be presented in this section does not apply universally, as will be seen in Section 5. However, it does apply to the Judy Benjamin case as well as to the case from Example 1, as in both the learning of the relevant conditional should intuitively leave the probability of the antecedent unaltered.

First consider updates on conditionals of the general form "If A, then the odds for $B_1, \ldots, B_n$ are $c_1 : \cdots : c_n$," where the set $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ partitions the space on which we suppose people's degrees-of-belief functions to be defined. For basically the same reasons we wanted (i)–(iii) to hold in the Judy Benjamin story, we will, for the general case, want to adopt the following desiderata:

(i*) after learning the conditional, a person's conditional degree of belief for B_i given A should equal $c_i / \sum_{j=1}^n c_j$, for all i;

(ii*) learning the conditional should leave a person's degrees of belief conditional on each of the propositions $\neg A$ and $A \wedge B_i$ ($i \leqslant n$) unchanged;

(iii*) a person's degree of belief for A after she learns the conditional should equal her degree of belief for A immediately *before* she learns the conditional.

The rule for updating on conditionals of the designated kind that we want to propose consists of two parts. The first part says that, after the learning of the conditional, a person's degrees of belief should be in accordance with (i*) indeed; that is, it dictates that

(6)   After learning "If A, then the odds for $B_1, \ldots, B_n$ are $c_1 : \cdots : c_n$," where $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ is a partition, a person should set her degree of belief for B_i conditional on A equal to $c_i / \sum_{j=1}^n c_j$, for all i.

So, for instance, after learning (1), Judy should set her conditional degree of belief for being in Headquarters Company area given that she is in Red territory equal to $3/4$ and her conditional degree of belief for being in Second Company area given that she is in Red territory equal to $1/4$.

The second part of the rule answers the question how one is to accommodate shifts in one's conditional degrees of belief such as are brought about by following (6). For the kind of case at issue—the kind in which learning a conditional gives one no reason to revise one's degree of belief for the conditional's antecedent—an answer has already been proposed by Bradley [2005]. Specifically, he proposes the following general update rule—which he dubbed "Adams conditioning"—for adjusting one's degrees of belief in response to an induced change in one or more of one's conditional degrees of belief:[14]

Definition 3.1 *Let a partition* $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ *be given such that, on one's initial degrees-of-belief function* $\Pr_0$*, it holds that* $\Pr_0(B_i \mid A) > 0$ *for all i, and suppose that one is caused to change some or all of one's conditional degrees of belief* $\Pr_0(B_i \mid A)$

---

[14]Bradley [2005:351] states a slightly less general version of this definition, but it is clear from his considerations that he subscribes to the version given here as well.

to $\Pr_1(B_i \mid A)$. *Then one updates by* Adams conditioning *on this change iff, for all propositions C,*

$$\Pr_1(C) \ = \ \Pr_0(C \mid \neg A)\Pr_0(\neg A) \ + \ \sum_{i=1}^{n} \Pr_0(C \mid A \wedge B_i)\Pr_1(B_i \mid A)\Pr_0(A).$$

Combined, the two parts yield the following formula for accommodating conditionals of the form and kind considered above:

$$(7) \qquad \Pr_1(C) \ = \ \Pr_0(C \mid \neg A)\Pr_0(\neg A) \ + \ \sum_{i=1}^{n} \Pr_0(C \mid A \wedge B_i)\frac{c_i}{\sum_{j=1}^{n} c_j}\Pr_0(A)$$

for all propositions C. Clearly, conditionals of the form "If A, B" can be thought of as being special instances of the more general form considered so far: "If A, then the odds for $B, \neg B$ are $1 : 0$." As one readily verifies (because then by (6) we must have $\Pr_1(B \mid A) = 1$ and $\Pr_1(\neg B \mid A) = 0$), instead of (7) we can use the simpler

$$(8) \qquad \Pr_1(C) \ = \ \Pr_0(C \mid \neg A)\Pr_0(\neg A) \ + \ \Pr_0(C \mid A \wedge B)\Pr_0(A).$$

Now notice that updating by (7) (or by (8), in the simplest case) satisfies desideratum (i*) by construction. That it satisfies (ii*) and (iii*) as well follows from[15]

Theorem 3.1 (Bradley) *Given a partition* $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ *such that, for all i,* $\Pr_0(B_i \mid A) > 0$. *Then a person's new degrees-of-belief function* $\Pr_1$ *comes from her old degrees-of-belief function* $\Pr_0$ *by Adams conditioning on a change in the conditional degrees of belief for* B_i *given A, for all i, iff for all C*

- $\Pr_1(A) \ = \ \Pr_0(A)$;
- $\Pr_1(C \mid A \wedge B_i) \ = \ \Pr_0(C \mid A \wedge B_i)$;
- $\Pr_1(C \mid A \wedge \neg B_i) \ = \ \Pr_0(C \mid A \wedge \neg B_i)$;
- $\Pr_1(C \mid \neg A) \ = \ \Pr_0(C \mid \neg A)$.

As desiderata (i)–(iii) are special instances of (i*)–(iii*), we see immediately that updating on (1) by dint of (7) yields a fully satisfactory outcome in the Judy Benjamin case. Moreover, if (7) is accepted, then updating on (3), too, leaves Judy's degree of belief for being in Red territory unchanged, and similarly in the case of Example 1. So, contrary to what van Fraassen claims, (3) certainly does not show that the intuition underlying the third desideratum cannot be generally respected anyway. Finally, as the conditional sentence that is part of (4) and (5) would, in the given context, seem of the sort that makes (7) applicable—neither in (4) nor in (5) does it seem to provide additional information relevant to Judy's degree of belief for being in Red territory—accepting (7) has the additional virtue that updating on (4) yields exactly the same outcome as updating on (5).

The for present purposes perhaps most crucial observation of Bradley's paper is that, in view of Theorem 3.1, "Adams conditioning is just a special case of Jeffrey conditioning" (p. 352). As (7) is just a special case of Adams conditioning, updating on a conditional "If A, then the odds for $B_1, \ldots, B_n$ are $c_1 : \cdots : c_n$" on our account is also really just a Jeffrey update. To be more exact, it is a Jeffrey update

---

[15]See the Appendix of Bradley [2005] for a proof (the theorem proved there is a bit less general than the one stated here, but it is obvious that Bradley's proof generalizes swiftly to one for Theorem 3.1).

on the partition $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$, with constraints $\Pr_1(\neg A) = \Pr_0(\neg A)$ and $\Pr_1(B_i \mid A) = c_i / \sum_{j=1}^{n} c_j$. This means that solving the Judy Benjamin problem does not require any exotic update rule: Jeffrey's rule suffices for that, and it solves the problem in a fully satisfactory way, unlike the rules that have been more commonly called upon in dealing with this problem.

Some might object that we cannot simply use Jeffrey's rule for the kind of update required in the case of Judy. After all—they might say—Jeffrey's rule takes a probability assignment over an entire partition as input, and in the case of Judy only part of that assignment is given. But this objection overlooks the fact that the context of the Judy Benjamin case provides us with the additional probabilistic information that is needed for applying Jeffrey's rule, as expressed in (iii). Jeffrey's rule does not place requirements on how we obtain the probability assignment over the partition. Indeed, in the prototypical cases for which Jeffrey proposed his rule, a clear separation between probabilities that derive from explicit information and those that derive from context does not seem feasible. Moreover, to suggest that Adams conditioning has an edge over Jeffrey's rule because it does not rely on such contextual elements would be wrongheaded. Whether we decide to apply Adams conditioning on the basis of the context or rather supply the additional input probability to Jeffrey's rule on the same basis is neither here nor there.

4. A distance function for Adams conditioning. While, as we argued, no exotic rule is needed to solve the Judy Benjamin problem, it is still interesting to note that a distance function very much like RE also does the trick, in particular, that it solves the problem in a way that does justice to all of (i)–(iii). To see this, first note that RE is not a proper distance measure, for it is not symmetric: the distance from $\Pr_0$ to $\Pr_1$, as measured by this function, is not equal to the distance from $\Pr_1$ to $\Pr_0$. So we can obtain an alternative function, which we call "Inverse Relative Entropy," by simply swapping the roles of the two probability functions:

$$(9) \qquad \mathsf{IRE}(\Pr_0, \Pr_1) \quad = \quad \sum_i \Pr_0(A_i) \log \frac{\Pr_0(A_i)}{\Pr_1(A_i)}.$$

Here, the A_i are again the minimal elements of the algebra $\mathcal{F}$. Like RE, the function IRE reaches the minimal value 0 iff the old and the new probability assignment are equal. Moreover, if we impose a constraint pertaining to all propositions of a partition of the minimal elements of $\mathcal{F}$, with $\Pr_0(A_i) > 0$ for all i, then updating by IRE minimization also concurs with updating by Jeffrey's rule.[16]

But, importantly, the results of minimizing IRE differ from the results achieved by means of minimizing RE in cases in which the constraint does not involve all propositions of the designated partition. In the Appendix, we prove the following theorem:

Theorem 4.1 *Let $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ be a partition such that $\Pr_0(B_i \mid A) > 0$, for all $i = 1, \ldots, n$. Say that on this partition we impose the constraint that $\Pr_1(A \wedge B_1) : \cdots : \Pr_1(A \wedge B_n) = c_1 : \cdots : c_n$, and we find the new assignment $\Pr_1$ by minimizing the IRE distance between $\Pr_0$ and $\Pr_1$. Then we have for all C and i that*

---

[16]We do not provide a proof for this, but it can be seen to follow quite easily from Lemma A.1 in the Appendix.

- $\Pr_1(A) = \Pr_0(A)$;
- $\Pr_1(C \mid A \wedge B_i) = \Pr_0(C \mid A \wedge B_i)$;
- $\Pr_1(C \mid A \wedge \neg B_i) = \Pr_0(C \mid A \wedge \neg B_i)$;
- $\Pr_1(C \mid \neg A) = \Pr_0(C \mid \neg A)$.

Comparing clauses 1–4 of Theorem 4.1 with clauses 1–4 of Theorem 3.1 shows immediately that Adams conditioning on a conditional of the form "If A, then the odds for $B_1, \ldots, B_n$ are $c_1 : \cdots : c_n$" leads to exactly the same result as minimizing IRE under the (sole) constraint that the odds for $A \wedge B_1, \ldots, A \wedge B_n$ are to be $c_1 : \cdots : c_n$. For the Judy Benjamin problem, this is a happy result. It means that we can have a perfectly adequate solution to this problem even if we favor an approach that has Judy determine her new probability function by minimizing the distance (according to some distance function) between that and her old probability function. What makes the result not just a happy one but also a surprising one is that it does not require straying very far from the rules van Fraassen and his coauthors have investigated. To the contrary, IRE minimization is arguably a very close cousin to INFOMIN, which requires minimizing RE, because the notion of closeness under the distance RE, or "RE-closeness" for short, is very similar to what may be called IRE-closeness. In fact, they are not just formally very similar—which is clear from comparing (2) with (9)— but also conceptually: where INFOMIN has you select the probability function that is RE-closest to your present probability function *as seen from your current perspective*, IRE minimization has you select the probability function that is RE-closest to your present probability function *as seen from the perspective you will have after adopting the probability function to be selected.*

5. Updating on conditionals continued: epistemic entrenchment. We are inclined to think that Adams conditioning, or, equivalently, Jeffrey conditioning with the explicit constraint of keeping the antecedent's probability fixed in the update, or, again equivalently, IRE minimization, covers most of the cases of learning a conditional. Unfortunately, however, it would be wrong to think that it covers all of them, as Example 2 already shows.

Example 2 A jeweller has been shot in his store and robbed of a golden watch. However, it is not clear at this point what the relation between these two events is; perhaps someone shot the jeweller and then someone else saw an opportunity to steal the watch. Kate thinks there is some chance that Henry is the robber (R). On the other hand, she strongly doubts that he is capable of shooting someone, and thus, that he is the shooter (S). Now the inspector, after hearing the testimonies of several witnesses, tells Kate: "If Henry robbed the jeweller, then he also shot him." As a result, Kate becomes more confident that Henry is *not* the robber, while her probability for Henry having shot the jeweller does not change.

As far as this description of the case goes, Kate's response seems pre-theoretically perfectly in order. If it is, then the learning of a conditional can decrease our degree of belief for the antecedent. Clearly, such learning events cannot be based on Adams conditioning, or on the corresponding kind of Jeffrey update, or on IRE minimization, all of which necessarily leave the probability of the conditional's antecedent unchanged upon learning the conditional. In fact, if it had not been for the problems advanced

earlier, conditioning on the material conditional would have seemed more suitable, since that operation accommodates our intuition that the probability of Henry being the robber decreases.

To accommodate the case of Kate, we can proceed much like we did in the cases of Judy and Sarah. From the conditional itself and the context we can determine the constraints that the new probability assignment must satisfy, such that we can derive a new probability assignment over the entire partition, and then we can apply Jeffrey's rule. In particular, the conditional determines that the probability of Henry's being the robber but not having shot the jeweller is zero. We might further stipulate that the probability of Henry's having shot the jeweller does not change as an effect of learning the said conditional. This effectively determines a new probability assignment to the propositions of the partition $\{R \wedge S, R \wedge \neg S, \neg R \wedge S, \neg R \wedge \neg S\}$, which is all we need to run Jeffrey's rule. In a sense, Example 2 is the mirror image of Example 1: where Sarah held onto her probability for the antecedent, Kate wants to leave the probability of the consequent unaffected. And indeed, by suitable transformations of the variables we can devise a variant of Adams conditioning for this case, and a distance function to go with it.

It seems, though, that we can tell a different story about Kate on the basis of the example. Naturally, Kate's belief that Henry is up to robbery may vary in strength. In addition to the constraints of the foregoing, she might also wish to impose the constraint that learning the conditional must not affect the probability of the antecedent either. However, she cannot simultaneously satisfy these constraints; her beliefs are subject to forces pulling in opposite directions. On the one hand, she cherishes the idea that Henry is not a murderer, but on the other hand she realizes full well that he was in need of some fast cash and might therefore well be the robber. Hence, she must try to find a trade-off between maintaining a high probability of Henry's being the robber and maintaining a low probability of his having shot the jeweller, and she must do so under the constraint that he cannot have done the former without having done the latter.

Even in this case, Kate may use Jeffrey's rule. She can determine a new probability assignment to the propositions of the above-mentioned partition and simply take this as input to the rule. However, the probabilities that Kate assigns to these propositions upon accepting "If Henry is the robber, then he also shot the jeweller" depend on her initial probabilities, and further on the firmness, or, as we shall say in the following, *epistemic entrenchments*, of her beliefs about Henry's being the robber and his having shot the jeweller.[17] Setting the probability of the propositions to some new value by hand seems to ignore these dependencies, or rather leaves the impact of these dependencies to be worked out independently of the update mechanism.

In the remainder of this section, we would like to present, though somewhat tentatively, a distance function that incorporates these dependencies directly. The attractive feature of the resulting update mechanism is that it does not require us to fill in a new probability distribution over the partition, as is the case in a Jeffrey update, nor to pin part of it down to the old value by default, as in Adams conditioning. Instead, the update mechanism presented in the following requires us to bring to the table the epistemic entrenchments of the propositions under scrutiny. The new probabilities

---

[17]Nayak et al. [1996] also draw upon the notion of epistemic entrenchment in their account of updating on conditionals, which, however, is a strictly qualitative extension of AGM theory and does not consider probabilistic change.

for the antecedent and consequent then follow from the old probabilities together with those epistemic entrenchments.

There are many different ways of incorporating epistemic entrenchments in a distance function. Outside the relatively well-known territory of relative entropy, we find a plethora of such functions, all with their own peculiarities. For present purposes, we will use a variant of the so-called Hellinger distance function.

The Hellinger distance reads as follows:

$$(10) \qquad \mathsf{HEL}(\mathrm{Pr}_0, \mathrm{Pr}_1) \quad = \quad \sum_i \left(\sqrt{\mathrm{Pr}_1(Q_i)} - \sqrt{\mathrm{Pr}_0(Q_i)} \right)^2.$$

Recall that the propositions Q_i represent the strongest consistent propositions in the algebra. For instance, in the murder case we might set $Q_1 = R \wedge S$, $Q_2 = R \wedge \neg S$, $Q_3 = \neg R \wedge S$, and $Q_4 = \neg R \wedge \neg S$. As with INFOMIN, or the rule of minimizing IRE, we can minimize the Hellinger distance between the old and new probability assignments under any constraint. Moreover, as with the other rules, this update procedure generalizes Jeffrey's rule: if we fix the probability assignment over a whole partition, the result of minimizing HEL is equal to the result of a Jeffrey update. But there is also an important difference between, on the one hand, distance functions like RE and IRE and, on the other, HEL. For RE, for instance, negative deviations from the old probability assignment of some Q_i, that is, $\mathrm{Pr}_1(Q_i) - \mathrm{Pr}_0(Q_i) < 0$, lead to negative contributions to the total distance; but these negative contributions are always offset by positive contributions due to deviations from the old probabilities of other propositions Q_j, with $i \neq j$, in such a way that the net contribution of all deviations is always positive. For HEL, we have that all deviations, positive and negative ones, lead to positive contributions to the total distance. We will use this property of HEL to construct a distance function that also takes into account the various epistemic entrenchments of the propositions Q_i.

The crucial property of this new distance function is that we factor in the deviations from different propositions in the partition differently. Whereas HEL weighs the contributions from deviations for the propositions Q_i all equally, the "epistemic entrenchment" function

$$(11) \qquad \mathsf{EE}(\mathrm{Pr}_0, \mathrm{Pr}_1) \quad = \quad \sum_i w_i \left(\sqrt{\mathrm{Pr}_1(Q_i)} - \sqrt{\mathrm{Pr}_0(Q_i)} \right)^2,$$

with the w_i being elements of $\mathbb{R}^+ \cup \{\omega\}$, allows us to give more weight to deviations in probability of some of the Q_i—for instance, those entailing the consequent of a given conditional—than to deviations in probability of others, for instance, those entailing the negation of the consequent. It thereby allows us to regulate whether, and to what extent, learning a conditional reflects back on the probability of the antecedent or rather influences the probability of the consequent. Clearly, this opens up a whole range of trade-offs between adapting the probability of the antecedent and that of the consequent, within which Jeffrey's rule and the procedure of minimizing HEL are just two special cases.

To illustrate the new rule, we show how it facilitates the type of update that Kate was supposed to make in the foregoing, trading off deviations in antecedent and consequent. A more general formulation of Kate's case involves an update on conditional odds, with the constraint that Q_1 (i.e., Henry is the robber and also the shooter, $R \wedge S$)

is to be r times as likely as Q_2 (= $R \wedge \neg S$), and the further constraint that the new probability function must somehow conserve the low probability for S and the high probability for R, to varying degrees. Imagine that we choose $w_1 = w_3 \gg w_2 = w_4$, so that deviations $\mathrm{Pr}_1(Q_2) - \mathrm{Pr}_0(Q_2)$ and $\mathrm{Pr}_1(Q_4) - \mathrm{Pr}_0(Q_4)$ make (much) smaller contributions to the total distance than deviations $\mathrm{Pr}_1(Q_1) - \mathrm{Pr}_0(Q_1)$ and $\mathrm{Pr}_1(Q_3) - \mathrm{Pr}_0(Q_3)$. Then the probability assignment Pr that is closest to the old assignment Pr_0 is one in which the probabilities for Q_1 and Q_3 have deviated less than if all weights had been equal. Therefore, setting the weights w_i to these values makes the deviations in the probability for Henry's having shot the jeweler, S, smaller than if the propositions had all had the same weight. In other words, these values of the w_i are associated with an epistemic state of Kate in which she is strongly inclined to stick to her low probability for S, but much less inclined to stick to her high probability for R.[18]

Some further illustrations of how the weights w_i determine which of the probabilities will remain close to their initial value and which will diverge more widely with the addition of the constraint are given in the table below, which shows a number of new probability assignments arrived at by minimizing EE from the initial probability assignment

$$\mathrm{Pr}_0(Q_1) \;=\; .1 \qquad\qquad \mathrm{Pr}_0(Q_3) \;=\; .1$$
$$\mathrm{Pr}_0(Q_2) \;=\; .7 \qquad\qquad \mathrm{Pr}_0(Q_4) \;=\; .1$$

under the constraint that $\mathrm{Pr}_1(Q_1)/\mathrm{Pr}_1(Q_2) = r$, as a function of r and the weights w_i attached to the various Q_i. For the latter, we set $w_1 = w_3$ and $w_2 = w_4 = 1$, and we vary the value of $w_1 (= w_3)$. The results show that we can regulate the trade-off that Kate makes between adapting her probabilities for R and, respectively, S by changing the weights (numbers have been rounded to two decimal places).[19]

r	w_1	$\mathrm{Pr}_1(Q_1)$	$\mathrm{Pr}_1(Q_2)$	$\mathrm{Pr}_1(Q_3)$	$\mathrm{Pr}_1(Q_4)$
3	1	.53	.18	.15	.15
	5	.21	.07	.13	.60
	100	.10	.03	.10	.76
50	1	.47	.01	.26	.26
	5	.15	.00	.13	.72
	100	.10	.00	.10	.79

[18]If we set $w_3 = w_4 \gg w_1 = w_2$, then the probability for R stays close to its original value. In particular, if we set $w_3 = w_4 = \omega$ and $w_1 = w_2 \in \mathbb{R}^+$, deviations in the probability of R lead to infinitely large contributions in the distance between the old and the new probability assignment. As a result, the new probability for R stays equal to the old value, just as in the application of IRE or Adams conditioning. In fact, it is not difficult to show that these update rules are limiting cases of EE.

[19]The table may give the impression that the application of EE can lead to a collapse of one of the probabilities to zero even while the constraint does not entail a probability of zero. Fortunately, this cannot happen. Clearly, if the constraint forces a particular ratio between the probability of two propositions in a partition, neither proposition can obtain a probability of zero because a ratio always requires two nonzero probabilities. If, on the other hand, the constraint forces a particular absolute difference between the two probabilities, neither probability will be forced to zero either. At zero, the derivative of the distance function EE is minus infinity, whereas nowhere on the domain is it positive and infinite. Therefore, decreasing the probability of one of the propositions in the partition to zero, in favor of an increase of the probabilities of any number of other propositions, can never lead to a positive contribution in the total distance function.

Recall that the above is intended as a tentative proposal only. One reason for being a little hesitant about it is that the weights a person is, on the proposal, supposed to assign to the relevant propositions will not come out of thin air but may be assumed to be interconnected with (even if presumably not fully determined by) things she believes; nor will these weights remain fixed once and for all but will, plausibly, themselves change in response to things the person learns. And, as it stands, our proposal is silent on both of these issues. To properly address them, we may well have to go beyond our current representation of epistemic states in terms of degrees of belief plus weights. For instance, it is conceivable, and it might be helpful for present concerns, to incorporate into the epistemic state a metric on possible worlds next to probability and weight functions. Or one could represent epistemic states as sets of probability assignments instead of single ones—as various authors have suggested as a way of dealing with vague probabilities—so that learning a constraint on these probability assignments could be accommodated by conditionalizing on the set of assignments. Finally, it may well be that learning other types of information, which might involve more complicated constraints on the belief state of the agent, necessitates the use of entirely different update techniques; the various types of constraint that may be associated with learning an indicative conditional by no means exhaust the spectrum.

On the other hand, when it comes to updating on conditionals that are probabilistically dependent on their antecedents, we should not too readily discount the possibility that all has been said about it once the rule of minimizing EE, or a kindred rule, has been pointed out. It may be that we cannot resort to any further rules for determining or adapting the weights representing the degrees of epistemic entrenchment, but that we have to rely on our own judgment for those purposes. Consider that something very similar is already the case for the kind of uncertain learning events that Jeffrey's rule was devised for: there is no rule telling us how a glimpse of a tablecloth in a poorly lit room is to change our assignment of probabilities to the various relevant propositions concerning the cloth's color. Basically the same point also applies to Adams conditioning (or, equivalently, Jeffrey's rule with the explicit constraint that the probability of the antecedent is to be kept fixed, or, again equivalently, the rule of IRE minimization) in that it may fall entirely upon us to decide, on the basis of contextual information, whether or not this rule applies to the learning of a given conditional. In fact, the point may be more general still. As Bradley [2005:362] stresses, even Bayes's rule "should not be thought of as a universal and mechanical rule of updating, but as a technique to be applied in the right circumstances, as a tool in what Jeffrey terms the 'art of judgment'." In the same way, determining and adapting the weights EE supposes, or deciding when Adams conditioning applies, may be an art, or a skill, rather than a matter of calculation or derivation from more fundamental epistemic principles.

6. Summary. We take our main results to be the following. First, we have shown that van Fraassen has been too quick in dismissing the intuitive verdict that Judy's degree of belief for her being in Red territory should not change as a result of updating on (1). That verdict might be hard to uphold if we were committed to the view that learning a conditional occasions conditionalizing on the corresponding material conditional, but this view was seen to run into several difficulties. Second, it was argued that, contrary to what has been generally supposed, Jeffrey's well-known version of con-

ditionalization is enough to solve satisfactorily the Judy Benjamin problem. Third, this proposed application of Jeffrey's rule appeared to relate naturally to Adams conditioning, which in turn was shown to have a natural underpinning in terms of an inverse relative entropy distance minimization rule. And finally, as Adams conditioning and the corresponding distance minimization rule apply only to conditionals the learning of which does not affect the probability of their antecedent, we provided a further distance function especially for the remaining class of conditionals. However, our commitment to this rule was more provisional, and we explicitly left open the possibility of other rules applying to the said class.

Appendix: proof of Theorem 4.1

We first prove two lemmas. The first of these establishes that when minimizing IRE, a change to the probability of a compound proposition leads to changes in the probabilities of the propositions it consists of that are proportional to the original probabilities of these latter propositions. Note that this is the same as saying that updating by minimizing IRE respects the rigidity condition imposed by Jeffrey. As Diaconis and Zabell [1982] show, several distance functions can replicate Jeffrey's rule; because the function IRE satisfies the rigidity condition, as made explicit in Lemma A.1, it also coincides with Jeffrey updating. That is, if the constraints on the probabilities concern a whole partition of propositions, the result of updating by IRE minimization is the same as the result of updating by Jeffrey's rule. The second lemma, on the other hand, is less naturally connected to the extant literature. It concerns the behavior of the distance IRE under the type of constraints associated with conditionals.

For the proof, assume a partition $\{Q_1, Q_2, \ldots, Q_N\}$ of strongest consistent propositions of a given algebra. Define U and V as disjoint sets composed of elements of this partition, $U = \bigcup_{j=1}^{m} Q_j$ and $V = \bigcup_{j=m+1}^{n} Q_j$; thus, $U \cap V = \varnothing$. Let $W = \bigcup_{j=1}^{n} Q_j = U \cup V$. Let $\Pr_0$ be a probability assignment over the partition. We are looking for the probability assignment $\Pr_1$ over the same partition that minimizes IRE under a given constraint. For convenience we write $\Pr_k(U) =: u_k$, $\Pr_k(V) =: v_k$, $\Pr_k(W) =: w_k = u_k + v_k$, and $\Pr_k(Q_j) =: q_{kj}$, for $k = 0, 1$.

Lemma A.1 *Let the constraint be such that $w_1 = w_0 + s$, for $s \in \mathbb{R}$. Now write $u_1 = u_0 + ts$ and $v_1 = v_0 + (1-t)s$. Then the minimum distance* IRE$(\Pr_0, \Pr_1)$ *is reached at $t = \frac{u_0}{w_0}$.*

Proof. Note that both U and V are composed of elements Q_j, and that the distance function IRE must be evaluated on the level of these strongest consistent propositions. First, we consider single pairs Q_j and $Q_{j'}$, that is, we look at changes in their

proportion if their total probability increases. Using the conventions above, we write the derivative of the distance as

$$
\begin{aligned}
\frac{d}{dt}\mathrm{IRE}(\Pr_0, \Pr_j) &= \frac{d}{dt}\left[c + q_{0j}\log\frac{q_{0j}}{q_{0j}+ts} + q_{0j'}\log\frac{q_{0j'}}{q_{0j'}+(1-t)s} \right] \\
&= \frac{d}{dt}\Big[c + q_{0j}\log q_{0j} + q_{0j'}\log q_{0j'} \\
&\qquad - q_{0j}\log(q_{0j}+ts) - q_{0j'}\log(q_{0j'}+(1-t)s) \Big] \\
&= \frac{q_{0j'}s}{q_{0j'}+(1-t)s} - \frac{q_{0j}s}{q_{0j}+ts}.
\end{aligned}
$$

Here c represents the contributions from deviations in the probability of all the propositions other than Q_j and $Q_{j'}$, none of which depend on t. Setting $\frac{d}{dt}\mathrm{IRE} = 0$ we can solve for t, and with some algebra we find that $t = \frac{q_{0j}}{q_{0j}+q_{0j'}}$. That is to say, the proportions of Q_j and $Q_{j'}$ do not change. Moreover, since

$$
\frac{d^2}{dt^2}\mathrm{IRE}(\Pr_0, \Pr_1) = \frac{q_{0j'}s^2}{\left(q_{0j'}+(1-t)s\right)^2} + \frac{q_{0j}s^2}{(q_{0j}+ts)^2} \quad > \quad 0
$$

for all t, this minimum is unique.

Now consider U and V, which are aggregates of these elementary propositions, and the constraint that imposes the shift $w_1 = w_0 + s$. Whatever the details of this shift, for every Q_j within W there is some value for s_j such that $q_{1j} = q_{0j} + s_j$. By the above fact we have that, for each pair Q_j and Q_{j+1}, the increases s_j and s_{j+1} will be proportional to the probabilities q_{0j} and $q_{0(j+1)}$, so $\frac{s_j}{s_{j+1}} = \frac{q_{0j}}{q_{0(j+1)}}$ for every $j < n$. Finally, we have that $\sum_{j=1}^{n} s_j = s$. We now have n constraints on the same number of increases s_j. We solve for the s_j and find

$$
s_j = \frac{q_{0j}}{\sum_{j'=1}^{n} q_{0j'}} s \qquad \text{for } j \leq n,
$$

and hence that

$$
\sum_{j=1}^{m} q_{1j} = \sum_{j=1}^{m} q_{0j} + \frac{\sum_{j=1}^{m} q_{0j}}{\sum_{j=1}^{n} q_{0j}} s.
$$

This comes down to $u_1 = u_0 + \frac{u_0}{w_0}s$, and thus $v_1 = v_0 + \frac{v_0}{w_0}s$. ⊠

Lemma A.2 *Let the constraint be $\frac{v_1}{u_1} = r$, and let $w_1 = w_0 + t$, for $t \in \mathbb{R}$. Then the minimum distance $\mathrm{IRE}(\Pr_0, \Pr_1)$ is reached at $t = 0$, so $w_1 = w_0$.*

Proof. To arrive at a probability assignment that fulfills to the constraint, we first adapt the initial probabilities u_0 and v_0 by a specific deviation s:

$$
\frac{v_1}{u_1} = \frac{v_0 + s}{u_0 - s} = r.
$$

We solve for s and obtain the solution

$$
s = \frac{ru_0 - v_0}{1 + r}.
$$

18

Note that $r > 0$, and that s can take any value depending on r; the resulting probabilities u_1 and v_1 must have the required proportion.

Now consider additional deviations $(1 + r)t$ from propositions outside W, writing $w_1 = w_0 + (1 + r)t$. Accordingly, we have $u_1 = u_0 - s + t$ and $v_1 = v_0 + s + rt$, thereby making sure that the required proportion of u_1 and v_1 is maintained. The parameter t thus labels all the possible probability functions fulfilling the constraint. We determine where in the domain of t the total distance $\mathsf{IRE}(\Pr_0, \Pr_1)$ is minimal. Note that by Lemma A.1 the deviation s is distributed proportionally among the elements Q_j within both U and V. Hence we can write

$$
\begin{aligned}
\mathsf{IRE}(\Pr_0, \Pr_1) \quad = \quad & \sum_{j=1}^{m} q_{0j} \log \frac{q_{0j}}{q_{0j} + \frac{q_{0j}}{u_0}\left(t - \frac{ru_0 - v_0}{1+r}\right)} \\
+ \quad & \sum_{j=m+1}^{n} q_{0j} \log \frac{q_{0j}}{q_{0j} + \frac{q_{0j}}{v_0}\left(rt + \frac{ru_0 - v_0}{1+r}\right)} \\
+ \quad & \sum_{j=n+1}^{N} q_{0j} \log \frac{q_{0j}}{q_{0j} - \frac{q_{0j}}{(1-w_0)}(1 + r)t}.
\end{aligned}
$$

We now use that

$$
q_{0j} - \frac{q_{0j}}{u_0}\frac{ru_0 - v_0}{1 + r} = \frac{q_{0j}(u_0 + v_0)}{u_0(1 + r)}, \qquad q_{0j} + \frac{q_{0j}}{v_0}\frac{ru_0 - v_0}{1 + r} = \frac{rq_{0j}(u_0 + v_0)}{v_0(1 + r)}.
$$

By some further algebra we arrive at

$$
\begin{aligned}
\mathsf{IRE}(\Pr_0, \Pr_1) \quad = \quad & c - \sum_{j=1}^{n} q_{0j} \log(w_0 + (1 + r)t) - \sum_{j=n+1}^{N} q_{0j} \log(1 - w_0 - (1 + r)t) \\
= \quad & c - w_0 \log(w_0 + (1 + r)t) - (1 - w_0) \log(1 - w_0 - (1 + r)t),
\end{aligned}
$$

where c does not depend on t. We take the derivative with respect to t, and find

$$
\frac{d}{dt}\mathsf{IRE}(\Pr_0, \Pr_1) = \frac{(1 - w_0)(1 + r)}{(1 - w_0) - (1 + r)t} - \frac{w_0(1 + r)}{w_0 + (1 + r)t}.
$$

Setting $\frac{d}{dt}\mathsf{IRE}(\Pr_0, \Pr_1) = 0$, we can solve for t, and find that $t = 0$. Moreover, since

$$
\frac{d^2}{d^2t}\mathsf{IRE}(\Pr_0, \Pr_1) = \frac{w_0(1 + r)^2}{(w_0 + (1 + r)t)^2} + \frac{(1 - w_0)(1 + r)^2}{((1 - w_0) - (1 + r)t)^2} > 0
$$

for all t, this minimum is unique. Hence we have that $w_1 = w_0$. $\boxtimes$

With these lemmas in place, we prove the four equalities in Theorem 4.1. Recall that the constraint is that $\Pr_1(A \wedge B_1) : \cdots : \Pr_1(A \wedge B_n) = c_1 : \cdots : c_n$. Now identify, in Lemma A.1, the variable U with the proposition $\neg A \wedge C$ and V with $\neg A \wedge \neg C$, and assume that the constraint leads to some deviation $w_1 = w_0 + s$, as above. We can then apply the lemma and derive that

$$
\frac{u_1}{w_1} = \frac{u_0 + \frac{u_0}{w_0}s}{w_0 + s} = \frac{u_0}{w_0},
$$

and hence that $\Pr_1(C\mid\neg A) = \Pr_0(C\mid\neg A)$. In the same vein identify, for every $i \leqslant n$, U with $A \wedge B_i \wedge C$ and V with $A \wedge B_i \wedge \neg C$, and assume that the constraint leads to $w_1 = w_0 + s$. We can then again apply Lemma A.1, and by analogous reasoning we arrive at $\Pr_1(C\mid A \wedge B_i) = \Pr_0(C\mid A \wedge B_i)$ for each $i \leqslant n$. A similar argument, for each $i \leqslant n$, yields $\Pr_1(C\mid A \wedge \neg B_i) = \Pr_0(C\mid A \wedge \neg B_i)$. Thus we have proved the last three clauses of Theorem 4.1.

To complete the proof, we need to show that $\Pr_1(A) = \Pr_0(A)$, meaning that the probability of the antecedent does not change when we update by IRE minimization on the constraint at issue. We prove this by a repeated application of both lemmas in a mathematical induction over the index i of the propositions B_i. The induction base is that the invariance holds for a constraint of the form $\Pr(A \wedge B_1) : \Pr(A \wedge B_2) = u_1 : v_1 = 1 : r$. This can be proved by a direct application of Lemma A.2, followed by an application of Lemma A.1. First, we identify U with $A \wedge B_1$, V with $A \wedge B_2$, and then derive from Lemma A.2 that $w_1 = w_0$, so that $\Pr_1(A \wedge (B_1 \vee B_2)) = \Pr_0(A \wedge (B_1 \vee B_2))$. Then we identify U with $\neg A$, V with $A \wedge \bigvee_{i>2} B_i$, so that by Lemma A.1, using $s = 0$, we can derive that $\Pr_1(\neg A) = \Pr_0(\neg A)$, and hence that $\Pr_1(A) = \Pr_0(A)$.

For the induction step, assume the induction hypothesis that $\Pr_1(A) = \Pr_0(A)$, and hence that $\Pr_1(\neg A) = \Pr_0(\neg A)$, after having imposed the constraint that $\Pr_1(A \wedge B_1) : \cdots : \Pr_1(A \wedge B_i) = c_1 : \cdots : c_i$. We are now imposing a further constraint on the proposition $A \wedge B_{i+1}$, to wit, $\Pr'_1(A \wedge B_1) : \cdots : \Pr'_1(A \wedge B_{i+1}) = c_1 : \cdots : c_{i+1}$. First, we identify U with $A \wedge \bigvee_{k<i+1} B_k$, and V with $A \wedge B_{i+1}$, and we apply Lemma A.2 to find that $w_1 = w_0$. Note that the probability function $\Pr'_1$ that we arrive at by applying the lemma in this way is closest to the probability function $\Pr_1$, and not necessarily closest to $\Pr_0$. In any case, by application of Lemma A.1, we can derive that the probability function $\Pr'_1$ has the same proportions among the probabilities of propositions outside W as do $\Pr_1$ and $\Pr_0$. These propositions include $\neg A$, so we have that $\Pr'_1(\neg A) = \Pr_0(\neg A)$ and $\Pr'_1(\neg A) = \Pr_1(\neg A) = \Pr_0(\neg A)$.

To complete the proof of the inductive step, we must show that the newly found probability function $\Pr'_1$ is not just closest to the probability function $\Pr_1$, but that, of all those functions $\Pr$ satisfying the constraint $\Pr(A \wedge B_1) : \cdots : \Pr(A \wedge B_{i+1}) = c_1 : \cdots : c_{i+1}$, it is also closest to the probability function $\Pr_0$. First note that the space of probability functions satisfying this constraint can be parameterized by t, according to

$$\forall k \leqslant i+1: \quad \Pr(A \wedge B_k) \;=\; \Pr'_1(A \wedge B_k) + tc_k,$$

$$\Pr\left(\neg A \vee \bigvee_{k>i+1} B_k\right) \;=\; \Pr'_1\left(\neg A \vee \bigvee_{k>i+1} B_k\right) - t\sum_{k\leqslant i+1} c_k.$$

We can write the deviations t in terms of the contributions from $A \wedge \bigvee_{k<i+1} B_k$ and $A \wedge B_{i+1}$ separately as $t\sum_{k<i+1} c_k$ and tc_{i+1}. By the induction hypothesis we know that over the propositions $A \wedge \bigvee_{k<i+1} B_k$ all nonzero t give a positive contribution to the total distance $\mathsf{IRE}(\Pr_0,\Pr_1)$. We must now check whether this positive contribution is offset by the contribution from the proposition $A \wedge B_{i+1}$. We write $A \wedge B_{i+1} = U$ and $\neg A \vee \bigvee_{k>i+1} B_k = V$, and assume that both consist of elements Q_j, as specified in the foregoing. Then, again using Lemma A.1, the net contribution of deviations

20

associated with the change tc_{i+1} in U to the total distance becomes

$$\Delta\mathrm{IRE}(\mathrm{Pr}_0,\mathrm{Pr}_1) \;=\; \sum_{j=1}^{m} q_{1j}\log\frac{q_{1j}}{q_{1j}+\frac{q_{1j}}{u_1}c_{i+1}t} - \sum_{j=1}^{m} q_{1j}\log\frac{q_{1j}}{q_{1j}}$$

$$+ \sum_{j=m+1}^{n} q_{1j}\log\frac{q_{1j}}{q_{1j}-\frac{q_{1j}}{v_1}t\sum_{k\leqslant i+1}c_k}$$

$$- \sum_{j=m+1}^{n} q_{1j}\log\frac{q_{1j}}{q_{1j}-\frac{q_{1j}}{v_1}t\sum_{j<i+1}c_j}.$$

Note that this expression concerns, not the minimal distance to Pr_1, but rather the change in the distance to Pr_0 effected by varying t. By some algebra, we obtain

$$\Delta\mathrm{IRE}(\mathrm{Pr}_0,\mathrm{Pr}_1) =$$

$$c \;-\; u_1\log(u_1 + tc_{i+1}) \;-\; v_1\log\left(v_1 - t\sum_{k\leqslant i+1}c_k\right) \;+\; v_1\log\left(v_1 - t\sum_{k<i+1}c_k\right),$$

where c is a constant not depending on t. We take the derivative with respect to t and find

$$\frac{d}{dt}\Delta\mathrm{IRE}(\mathrm{Pr}_0,\mathrm{Pr}_1) \;=\; \frac{u_1 c_{i+1}}{u_1 + tc_{i+1}} - \frac{v_1\sum_{k\leqslant i+1}c_k}{v_1 - t\sum_{k\leqslant i+1}c_k} + \frac{v_1\sum_{k<i+1}c_k}{v_1 - t\sum_{k<i+1}c_k}.$$

At $t = 0$ this derivative is zero. Moreover, the second derivative is

$$\frac{d^2}{dt^2}\Delta\mathrm{IRE}(\mathrm{Pr}_0,\mathrm{Pr}_1) \;=\; \frac{u_1 c_{i+1}^2}{(u_1 + tc_{i+1})^2} + \frac{v_1\left(\sum_{k\leqslant i+1}c_k\right)^2}{\left(v_1 - t\sum_{k\leqslant i+1}c_k\right)^2} - \frac{v_1\left(\sum_{k<i+1}c_k\right)^2}{\left(v_1 - t\sum_{k<i+1}c_k\right)^2}.$$

The middle term on the right is strictly larger than the rightmost term, and so the second derivative is positive everywhere. Hence, $t = 0$ is a global minimum for the distance IRE.

Consequently, we cannot offset the positive contribution to the distance from the propositions $A \wedge \bigvee_{k<i+1} B_k$ by adapting the probability of the proposition $A \wedge B_{i+1}$ by any amount t, since any further deviation from the probability of that proposition leads to a further increase in the distance. Therefore, the probability Pr'_1 of the foregoing is indeed the probability closest to the probability Pr_0, according to the distance IRE. And because $\mathrm{Pr}'_1(A) = \mathrm{Pr}_0(A)$, we have thereby proved the inductive step, which completes the proof. ⊠

References

Adams, E. [1975] *The Logic of Conditionals*, Dordrecht: Reidel.

Adams, E. [1994] "Updating on Conditional Information," *IEEE Transactions on Systems, Man and Cybernetics* 24:1708–1713.

Adams, E. [1998] *A Primer of Probability Logic*, Stanford: CSLI Publications.

Bradley, R. [2005] "Radical Probabilism and Bayesian Conditioning," *Philosophy of Science* 72:342–364.

Diaconis, P. and Zabell, S. L. [1982] "Updating Subjective Probability," *Journal of the American Statistical Association* 77:822–830.

Edgington, D. [1995] "On Conditionals," *Mind* 104:235–329.

Grove, A. and Halpern, J. [1997] "Probability Update: Conditioning vs. Cross-Entropy," in *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence*, San Francisco: Morgan Kaufmann, pp. 208–214.

Jackson, F. [1979] "On Assertion and Indicative Conditionals," *Philosophical Review* 88:565–589.

Jackson, F. [1987] *Conditionals*, Oxford: Blackwell.

Jeffrey, R. [1983] *The Logic of Decision* (2nd ed.), Chicago: University of Chicago Press.

Lewis, D. [1976] "Probabilities of Conditionals and Conditional Probabilities," *Philosophical Review* 85:297–315.

Nayak, A., Pagnucco, M., Foo, N., and Peppas, P. [1996] "Learning from Conditionals: Judy Benjamin's Other Problems," in: W. Wahlster (ed.) *ECAI96: 12th European Conference on Artificial Intelligence*, New York: Wiley, pp. 75–79.

Nolan, D. [2003] "Defending a Possible-Worlds Account of Indicative Conditionals," *Philosophical Studies* 116:215–269.

Popper, K. R. and Miller, D. [1983] "A Proof of the Impossibility of Inductive Probability," *Nature* 302:687–688.

Skyrms, B. [1980] *Causal Necessity*, New Haven, Conn.: Yale University Press.

Uffink, J. [1995] "Can the Maximum Entropy Principle be Explained as a Consistency Requirement?" *Studies in History and Philosophy of Modern Physics* 26:223–261.

van Fraassen, B. C. [1981] "A Problem for Relative Information Minimizers in Probability Kinematics," *British Journal for the Philosophy of Science* 32:375–379.

van Fraassen, B. C. [1989] *Laws and Symmetry*, Oxford: Clarendon Press.

van Fraassen, B. C., Hughes, R. I. G., and Harman, G. [1986] "A Problem for Relative Information Minimizers, Continued," *British Journal for the Philosophy of Science* 37:453–463.

Williamson, J. [2009] "Objective Bayesianism, Bayesian Conditionalisation and Voluntarism," *Synthese*, in press.

Williamson, T. [2007] *The Philosophy of Philosophy*, Oxford: Blackwell.