For the Wrong Reasons

Remco Heesen
M.Sc. Philosophy of Science

Abstract. This essay critically discusses the Judy Benjamin problem, a problem in Bayesian belief updating, focusing in particular on the proposed solution due to Douven and Romeijn (2010). We examine their rebuttal of traditional treatments of the problem and their threefold characterisation of their own proposal. The first characterisation is in terms of desiderata, the second in terms of Adams conditioning (see Bradley 2005) and the third in terms of distance-minimisation. Correcting some mathematical mistakes in their paper, we show that the three characterisations are indeed equivalent. We then state some worries concerning the distance-based approach, which by extension cast doubt on the whole project.

In a recent paper, Douven and Romeijn (2010) take up the Judy Benjamin problem (Van Fraassen 1981). They argue that contrary to what Van Fraassen and the subsequent literature have claimed, on learning a conditional statement we should not necessarily change our degrees of belief in the antecedent. Van Fraassen had dismissed this intuitively straightforward treatment of the Judy Benjamin case based on an argument using Bayesian conditioning on a conditional, which Douven and Romeijn consider but find wanting. In this essay we will discuss the various characterisations Douven and Romeijn give for their approach. We will point out and correct a few mistakes in their mathematics. After briefly sketching the context of the Judy Benjamin problem and introducing the problem itself, we will first discuss the desiderata Douven and Romeijn give for a solution of the problem. Then we will discuss why Van Fraassen rejects one of these and outline his alternative treatment of the problem. We return to Douven and Romeijn to see how they in turn reject Van Fraassen's reasoning. They then relate a new rule, "Adams conditioning", to their desiderata—wrongly, as we will show. We then turn to a third characterisation that Douven and Romeijn give, via a distance function between probability assignments. After correcting the theorem with which they link this approach to the desiderata we will discuss some problems of this distance-based approach.

In this essay the terms "probability" (Pr) and "degrees of belief" will be used interchangeably to denote a *Bayesian* approach to probabilities. The most important characteristic of this approach is that there is some latitude in the initial assignment of probabilities to propositions (how much is irrelevant for our purposes), which is used to shift the initial assignment when new information is learned. Two central claims of Bayesianism are that a rational person's degrees of belief satisfy the *probability calculus* (which means that mathematical probability theory applies to it) and that she updates her probabilities by (Bayesian) *conditioning*. This means that, given that new information is learned which can be fully characterised by a proposition E, her new probability assignment Pr_1 will be related to her old assignment Pr_0 such that for all propositions C

$$\mathrm{Pr}_1(C) = \mathrm{Pr}_0(C|E).$$

A problem with this type of conditioning is that it covers only a limited number of cases. What if, for instance, information is learned that makes a proposition more likely, but not certain? *Jeffrey conditioning* was invented to deal with this kind of cases. This update rule uses the concept of a *partition*, a set of mutually exclusive and exhaustive propositions.

Definition 1 (Jeffrey conditioning) *Given that some learning event is fully described by a change in the degrees of belief over a partition* $\{A_1, \ldots, A_n\}$*, one updates by Jeffrey conditioning if for all* C

$$\Pr_1(C) = \sum_{i=1}^{n} \Pr_0(C|A_i) \Pr_1(A_i).$$

Again there are types of learning that may not be covered by this rule. The case of Judy Benjamin (first introduced by Van Fraassen 1981) is a putative example of this.

Example 2 Private Judy Benjamin and her platoon have been dropped in a swampy area which is equally divided between the rival Red Army (R) and their own Blue Army ($\neg R$). Each of the two territories is again equally divided between Second Company Area (S) and Headquarters Company Area ($\neg S$) (see figure 1). As the platoon has no way of telling where it has been dropped, Judy initially assigns

	S	$\neg S$
R	$\frac{1}{4}$	$\frac{1}{4}$
$\neg R$	$\frac{1}{4}$	$\frac{1}{4}$

Figure 1

$$\Pr_0(R \wedge S) = \Pr_0(R \wedge \neg S) = \Pr_0(\neg R \wedge S) = \Pr_0(\neg R \wedge \neg S) = \frac{1}{4}.$$

Next, the platoon manages to establish radio contact with its own headquarters, which relays the following message: "If you have strayed into Red Army territory, the odds are 3:1 that you are in their Headquarters Company Area". The radio subsequently gives out, revealing no more information. The question then is how Judy Benjamin should rationally update her degrees of belief over the partition $\{R \wedge S, R \wedge \neg S, \neg R \wedge S, \neg R \wedge \neg S\}$.

Douven and Romeijn (2010) list three desiderata that we might want the new probability assignment Pr_1 to satisfy.

1. The conditional probability, given that Judy Benjamin is in Red Army territory, of being in Headquarters Company Area should be three times as large as that of being in Second Company Area. Since they necessarily add to one, this yields $\mathrm{Pr}_1(\neg S|R) = \frac{3}{4}$ and $\mathrm{Pr}_1(S|R) = \frac{1}{4}$.

2. Since the new information concerns *only* the relations between propositions in the partition $\{R \wedge S, R \wedge \neg S, \neg R\}$, conditional degrees of belief given any of these three propositions should be unchanged: $\mathrm{Pr}_1(C|A) = \mathrm{Pr}_0(C|A)$ for all C and for all $A \in \{R \wedge S, R \wedge \neg S, \neg R\}$.

3. The new information also tells us nothing about whether or not the antecedent (being in Red Army territory) is true, so $\mathrm{Pr}_1(R) = \mathrm{Pr}_0(R) = \frac{1}{2}$.

These desiderata seem intuitively highly plausible. But are they correct? Van Fraassen (1981 p.379) and Van Fraassen et al. (1986 p.455) think not, giving the following argument against desideratum 3. Suppose the radio officer had said, "If you have strayed into Red Army territory, the odds are 0:1 that you are in their Headquarters Company Area", or equivalently, "If R, then S". In that case we could simply apply Bayesian conditioning on $R \supset S$, yielding

$$\mathrm{Pr}_1(R) = \mathrm{Pr}_0(R|R \supset S) = \mathrm{Pr}_0(R|\neg R \vee S) = \frac{\mathrm{Pr}_0(R \wedge S)}{\mathrm{Pr}_0(\neg R \vee S)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \neq \frac{1}{2}.$$

So if desideratum 3 is taken to mean that on learning any information of the form "If R, then odds for $\neg S$ versus S are r" we should keep our degrees of belief in R constant, the above constitutes a counterexample.

Van Fraassen et al. (1986) then develop three competing rules for updating probability assignments on the basis of learning new information: INFOMIN, MTP and MUD. Each of them is based on the idea of conservatively using the new information: a distance function is defined and the rule requires us to minimise the distance between the old and the new probability assignment under constraints specified by the new information. For INFOMIN, for example, the "relative entropy" function is defined:

$$\mathsf{RE}(\mathrm{Pr}_0, \mathrm{Pr}_1) = \sum_i \mathrm{Pr}_1(A_i) \log \frac{\mathrm{Pr}_1(A_i)}{\mathrm{Pr}_0(A_i)},$$

where $\{A_i\}$ is the partition of strongest consistent propositions. The rule then states that given Pr_0 and constraint(s) π we ought to update to $\mathrm{Pr}_1 \in \{\mathrm{Pr} \mid \mathrm{Pr} \text{ satisfies } \pi\}$ which minimizes $\mathsf{RE}(\mathrm{Pr}_0, \mathrm{Pr}_1)$. In the case of the Judy Benjamin problem, the constraint is $\dfrac{\mathrm{Pr}_1(R \wedge S)}{\mathrm{Pr}_1(R \wedge \neg S)} = \dfrac{1}{3}$ so that INFOMIN yields (approximately) $\mathrm{Pr}_1(R \wedge S) = 0.117$, $\mathrm{Pr}_1(R \wedge \neg S) = 0.351$ and $\mathrm{Pr}_1(\neg R) = 0.532$ (Van Fraassen 1981). Note that this indeed violates 3. Indeed, it is consistent with the objection that was posed: as the value of the constraint approaches 0 or ∞, $\mathrm{Pr}_1(R)$ goes to $\frac{1}{3}$. The problem posed by Van Fraassen et al. (1986) is how to choose between or argue for any of the three rules they define, given that all of them satisfy the desiderata the authors propose but yield different results.

Douven and Romeijn, however, reject all three of these rules. They want to reinstall desideratum 3 and hence have to deal with the objection against it. The objection says that in the case where the new information is "If R, then S":

(i) The truth conditions of the new information (an indicative conditional) are the same as those of the material conditional $R \supset S$; and

(ii) This justifies updating by conditioning on the latter in response to learning the former.

Douven and Romeijn say little about (i), but they observe that it is far from uncontroversial. However, extra justification for (ii) than (i) might be given, so the focus should be on (ii) itself. Douven and Romeijn use the following mathematical fact in their construction of a counterexample.

Fact 3 *For all* A *and* B *such that* $0 < \Pr(A) < 1$, $\Pr(B) > 0$ *and* $\Pr(B|A) < 1$, *it holds that* $\Pr(A|A \supset B) < \Pr(A)$.

They then describe the following case.

Example 4 Sarah and Marian want to go for sundowners (S) at the Westcliff hotel. They plan to sit inside if it rains (R) or outside if it does not ($\neg R$). They are unsure whether or not it will be raining, so $0 < \Pr_0(R) < 1$. Something unexpected may come up, so $\Pr_0(\neg S) > 0$, but as indicated they have taken rain into account, so $\Pr_0(\neg S|R) < 1$. But then they learn that a wedding party will take place at the hotel, so sitting inside is not an option. This means that they will not be having sundowners if it rains: they learn "If R, then $\neg S$". As a result they set $\Pr_1(R \wedge \neg S) = 0$. Because the new information clearly has no influence on their knowledge of the weather, $\Pr_1(R) = \Pr_0(R)$. However, if they had conditioned on $R \supset \neg S$ they would have obtained $\Pr_1(R) = \Pr_0(R|R \supset \neg S) < \Pr(R)$ by fact 3.

Note the similarity between this example and the Judy Benjamin case. In both cases conditioning on $R \supset \neg S$ leads to decreasing confidence in R ($\Pr_1(R) < \Pr_0(R)$) while intuitively we think there should be no such change ($\Pr_1(R) = \Pr_0(R)$). The key difference between the two examples is that in the latter, the intuition that our confidence in R is unchanged by the information is strengthened because supposing otherwise would assume that having a wedding can influence the weather. The conclusion is that conditioning on the corresponding material conditional is in general not the rational response on learning an indicative conditional.

The next thing Douven and Romeijn note is that the solution based on desiderata 1–3 is a case of *Adams conditioning*, introduced by Bradley

(2005). We state a generalization of the definition he gives and the theorem he proves.

Definition 5 (Adams conditioning) *Given are a partition* $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ *and a probability assignment* Pr_0 *with* $\mathrm{Pr}_0(B_i|A) > 0$ *for all* i*. Let the new information be such that the conditional degrees of belief for* B_i *given* A *change to* $\mathrm{Pr}_1(B_i|A)$ *for all* i*. Then the update is a case of Adams conditioning if for all* C*:*

$$\mathrm{Pr}_1(C) = \mathrm{Pr}_0(C|\neg A)\,\mathrm{Pr}_0(\neg A) + \sum_{i=1}^{n} \mathrm{Pr}_0(C|A \wedge B_i)\,\mathrm{Pr}_1(B_i|A)\,\mathrm{Pr}_0(A). \tag{1}$$

It turns out that updating by Adams conditioning is equivalent to the three desiderata, as shown by the following theorem.

Theorem 6 (Bradley) *Given a partition* $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ *with* $\mathrm{Pr}_0(B_i|A) > 0$ *for all* i*. A probability assignment* Pr_1 *is obtained from* Pr_0 *by Adams conditioning on a change in the conditional degrees of belief in* B_i *given* A *(for all* i*) iff for all* C *and* i*,*

- $\mathrm{Pr}_1(A) = \mathrm{Pr}_0(A)$,
- $\mathrm{Pr}_1(C|A \wedge B_i) = \mathrm{Pr}_0(C|A \wedge B_i)$,
- $\mathrm{Pr}_1(C|\neg A) = \mathrm{Pr}_0(C|\neg A)$.

Proof If Pr_1 is obtained by Adams conditioning, then by (1):

$$\mathrm{Pr}_1(A) = \sum_{i=1}^{n} \mathrm{Pr}_1(B_i|A)\,\mathrm{Pr}_0(A) = \mathrm{Pr}_0(A).$$

Similarly, for an arbitrary proposition C and index i

$$\begin{aligned}\Pr_1(C|A \wedge B_i) &= \frac{\Pr_1(C \wedge A \wedge B_i)}{\Pr_1(A \wedge B_i)} = \frac{\Pr_0(C|A \wedge B_i)\Pr_1(B_i|A)\Pr_0(A)}{\Pr_1(B_i|A)\Pr_0(A)} \\ &= \Pr_0(C|A \wedge B_i), \\ \Pr_1(C|\neg A) &= \frac{\Pr_1(C \wedge \neg A)}{\Pr_1(\neg A)} = \frac{\Pr_0(C|\neg A)\Pr_0(\neg A)}{\Pr_0(\neg A)} = \Pr_0(C|\neg A).\end{aligned}$$

This proves the first part of the theorem. Next we suppose that the conditions listed above hold and prove that (1) is satisfied for all C, implying that Adams conditioning has taken place. Let C be an arbitrary proposition. Then by the law of total probability

$$\Pr_1(C) = \Pr_1(C|\neg A)\Pr_1(\neg A) + \sum_{i=1}^{n} \Pr_1(C|A \wedge B_i)\Pr_1(A \wedge B_i),$$

and by the conditions listed above, this equals

$$\begin{aligned}&= \Pr_0(C|\neg A)\Pr_0(\neg A) + \sum_{i=1}^{n} \Pr_0(C|A \wedge B_i)\Pr_1(A \wedge B_i) \\ &= \Pr_0(C|\neg A)\Pr_0(\neg A) + \sum_{i=1}^{n} \Pr_0(C|A \wedge B_i)\Pr_1(B_i|A)\Pr_0(A). \quad \square\end{aligned}$$

Douven and Romeijn (2010) add a fourth criterion, $\Pr_1(C|A \wedge \neg B_i) = \Pr_0(C|A \wedge \neg B_i)$, claiming that the full set of four criteria is equivalent to Adams conditioning, but this is erroneous. Bradley's proof for $\Pr_1(C|A \wedge \neg B) = \Pr_0(C|A \wedge \neg B)$ depends on $A \wedge \neg B$ being a single element of the strongest relevant partition. The following example shows that if this does not hold, $\Pr_1(C|A \wedge \neg B_i)$ obtained by Adams conditioning may be different from $\Pr_0(C|A \wedge \neg B_i)$.

Example 7 Jane goes to the market to buy a gift for her sick friend Mike. She sees a nice little fruit stall and tentatively decides to buy him some fruit (F). She thinks he might like apples (A), oranges (O) or bananas (B). She is undecided which of these Mike will prefer, and considers the additional (not too likely) option that the fruit is no good and she has to go somewhere else: $\mathrm{Pr}_0(\neg F) = \mathrm{Pr}_0(F \wedge A) = \mathrm{Pr}_0(F \wedge O) = \mathrm{Pr}_0(F \wedge B) = \frac{1}{4}$. Without letting it influence her decision, she also knows that apples are grown nearby (N), so $\mathrm{Pr}_0(N|F \wedge A) = 1$, whereas bananas are grown far away: $\mathrm{Pr}_0(N|F \wedge B) = 0$. She is uncertain where oranges are grown: $\mathrm{Pr}_0(N|F \wedge O) = \frac{1}{2}$. Note that

$$\begin{aligned}
\mathrm{Pr}_0(N|F \wedge \neg A) &= \frac{\mathrm{Pr}_0(N \wedge F \wedge \neg A)}{\mathrm{Pr}_0(F \wedge \neg A)} \\
&= \frac{\mathrm{Pr}_0(N \wedge F \wedge O) + \mathrm{Pr}_0(N \wedge F \wedge B)}{\mathrm{Pr}_0(F \wedge O) + \mathrm{Pr}_0(F \wedge B)} \\
&= \frac{\frac{1}{8} + 0}{\frac{1}{4} + \frac{1}{4}} = \frac{1}{4}.
\end{aligned}$$

She walks up to the market salesman and tells him that she wants to buy some fruit for Mike. He tells her, "Of my customers, half prefer oranges, a third prefers bananas and only one in six prefers apples." This causes Jane to update her degrees of belief in what Mike will prefer: $\mathrm{Pr}_1(O|F) = \frac{1}{2}$, $\mathrm{Pr}_1(B|F) = \frac{1}{3}$ and $\mathrm{Pr}_1(A|F) = \frac{1}{6}$. Since she has learned nothing else, she updates by Adams conditioning with respect to the partition $\{\neg F, F \wedge A, F \wedge O, F \wedge B\}$. We have

$$\begin{aligned}
\mathrm{Pr}_1(N|F \wedge \neg A) &= \frac{\mathrm{Pr}_1(N \wedge F \wedge \neg A)}{\mathrm{Pr}_1(F \wedge \neg A)} \\
&= \frac{\mathrm{Pr}_1(N \wedge F \wedge O) + \mathrm{Pr}_1(N \wedge F \wedge B)}{\mathrm{Pr}_1(F \wedge O) + \mathrm{Pr}_1(F \wedge B)},
\end{aligned}$$

and using (1), this equals

$$= \frac{\Pr_0(N|F \wedge O)\Pr_1(O|F)\Pr_0(F) + \Pr_0(N|F \wedge B)\Pr_1(B|F)\Pr_0(F)}{\Pr_1(O|F)\Pr_0(F) + \Pr_1(B|F)\Pr_0(F)}$$

$$= \frac{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{3}{4}+0}{\frac{1}{2}\cdot\frac{3}{4}+\frac{1}{3}\cdot\frac{3}{4}} = \frac{3}{10} \neq \frac{1}{4} = \Pr_0(N|F \wedge \neg A).$$

This mathematical problem in Douven and Romeijn's paper notwithstanding, Adams conditioning seems a reasonable thing to do in the kind of situations to which it applies. Bradley (2005 p.352) strengthens this point by observing that, "Adams conditioning is just a special case of Jeffrey conditioning". Theorem 6 shows that Adams conditioning is exactly Jeffrey conditioning over the partition $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ with the new degrees of belief in $\neg A$ equal to the old degrees of belief in it, and the new degrees of belief in $A \wedge B_i$ given by the new information. New degrees of belief in other propositions are obtained by conditioning on the elements in this partition. This shows that cases like the Judy Benjamin problem can be handled by Jeffrey conditioning. That Adams conditioning supplies $\Pr_1(\neg A) = \Pr_0(\neg A)$ as part of the rule whereas Jeffrey conditioning requires it as input is merely a semantical difference, as in either case we will need the context of the problem to tell us whether it is reasonable to require this. Hence there will be no difference in practice between Adams conditioning and Jeffrey conditioning in the particular context where Adams conditioning is applicable (Douven and Romeijn 2010).

Having reinstated desideratum 3 and shown that Adams (Jeffrey) conditioning can satisfy all three desiderata, Douven and Romeijn have all the reasons they need to reject INFOMIN, MTP and MUD, the distance-minimisation rules proposed by Van Fraassen et al. (1986). It turns out, however, that a distance-minimisation rule can still be defended, based on its equivalence to the conditioning approach. This rule requires us to minimise the "inverse relative entropy" between the old and the new probability assignments under the constraint of the new information, with the inverse relative entropy between two assignments $\Pr_0$ and $\Pr_1$ given by

$$\mathsf{IRE}(\mathrm{Pr}_0, \mathrm{Pr}_1) = \mathsf{RE}(\mathrm{Pr}_1, \mathrm{Pr}_0) = \sum_i \mathrm{Pr}_0(A_i) \log \frac{\mathrm{Pr}_0(A_i)}{\mathrm{Pr}_1(A_i)},$$

with $\{A_i\}$ again the partition of strongest relevant propositions. Douven and Romeijn (2010, theorem 4.1) claim to prove that minimising IRE distance under the sole constraint that $\mathrm{Pr}_1(A \wedge B_1) : \cdots : \mathrm{Pr}_1(A \wedge B_n) = c_1 : \cdots : c_n$ entails for all C and i

- $\mathrm{Pr}_1(A) = \mathrm{Pr}_0(A)$,
- $\mathrm{Pr}_1(C|A \wedge B_i) = \mathrm{Pr}_0(C|A \wedge B_i)$,
- $\mathrm{Pr}_1(C|\neg A) = \mathrm{Pr}_0(C|\neg A)$,
- $\mathrm{Pr}_1(C|A \wedge \neg B_i) = \mathrm{Pr}_0(C|A \wedge \neg B_i)$.

Note, however, that the first three conditions already fully fix a probability assignment for all propositions C (as the second part of our proof for theorem 6 shows) and that a probability assignment so fixed may, in general, have $\mathrm{Pr}_1(C|A \wedge \neg B_i) \neq \mathrm{Pr}_0(C|A \wedge \neg B_i)$ (as example 7 shows). So their theorem is in fact a conjecture that is now disproved. We replace this conjecture with our own theorem (a proof is appended).

Theorem 8 *Let* $\{\neg A, A \wedge B_1, \ldots, A \wedge B_n\}$ *be a partition with* $\mathrm{Pr}_0(A \wedge B_i) > 0$ *for all* i*. Let the new probability assignment* Pr_1 *be obtained by minimising the* IRE *distance between* Pr_0 *and* Pr_1 *under the constraint that* $\mathrm{Pr}_1(A \wedge B_1) : \cdots : \mathrm{Pr}_1(A \wedge B_n) = c_1 : \cdots : c_n$ *(where* $c_1, \ldots, c_n > 0$*). Then it holds for all* C *and* i *that*

- $\mathrm{Pr}_1(A) = \mathrm{Pr}_0(A)$,
- $\mathrm{Pr}_1(C|A \wedge B_i) = \mathrm{Pr}_0(C|A \wedge B_i)$,
- $\mathrm{Pr}_1(C|\neg A) = \mathrm{Pr}_0(C|\neg A)$.

Douven and Romeijn interpret $\mathsf{RE}(\Pr_0, \Pr_1)$ as the distance from $\Pr_0$ to $\Pr_1$, so that INFOMIN tells us to choose $\Pr_1$ that is closest to $\Pr_0$ from the perspective of $\Pr_0$. $\mathsf{IRE}(\Pr_0, \Pr_1)$, by the same interpretation, is the distance from $\Pr_1$ to $\Pr_0$, so that IRE-minimisation tells us to choose $\Pr_1$ that will be closest to $\Pr_0$ as seen from the perspective that we are about to take. This makes a difference, because relative entropy is not symmetric (and is therefore, in the mathematical sense, not a distance metric). This yields the strange result that IRE-minimisation asks us to compare numbers obtained from different perspectives. This may be possible mathematically, but leads to a new philosophical problem that INFOMIN did not have, namely: what guarantees the commensurability of relative entropies obtained from different perspectives? Can we say that we are measuring distance by the same standards if we take a different perspective for each measurement?

	S	$\neg S$
R	?	0
$\neg R$	?	?

Figure 2

More problems arise if we consider again the limiting case that Van Fraassen et al. (1986) used to criticise desideratum 3 (see figure 2). We used example 4 to motivate the Adams conditioning annex IRE-minimisation approach. Note that IRE-minimisation strictly does not apply here because one of the c_i's is zero (which would lead to division by zero problems in the IRE function because it would constrain $\Pr_1(R \wedge S)$ to be zero). But the Adams conditioning approach can be characterised as a limiting case of what happens when we minimise IRE with respect to the partition $\{\neg R, R \wedge S, R \wedge \neg S\}$ and let the c_i associated with $R \wedge S$ go to zero. Now consider the following example (due to Douven and Romeijn 2010).

Example 9 A jeweller has been killed and robbed of a golden watch, but not necessarily by the same person. Kate knows that Henry needs money and hence might well be the robber (R), but she is relatively certain he is uncapable of shooting someone (S). After some investigation, the inspector tells her: "If Henry robbed the jeweller, he also shot him", so Kate learns and updates on "If R, then S". Since she is more convinced that he didn't shoot someone than that he robbed someone, she leaves her degrees of belief for the consequent S unchanged and updates by lowering her probability for the antecedent R.

The information received in example 9 is of the same form as in example 4. Yet the way Kate updates does not conform to the principle of leaving the probability for the antecedent unchanged (desideratum 3). Interestingly, it does conform to the update we would get if we updated using IRE-minimisation with respect to the partition $\{S, R \wedge \neg S, \neg R \wedge \neg S\}$ (note the difference!) and took the limit where the c_i associated with $R \wedge \neg S$ goes to zero.

To make matters worse, Bovens (2010) invokes a symmetry argument to show that we should in fact leave neither our probability for R nor our probability for S unchanged. Rather, we should favour a solution that yields $\Pr_1(R \wedge S) = \Pr_1(\neg R \wedge S) = \Pr_1(\neg R \wedge \neg S) = \frac{1}{3}$. It turns out that this solution is obtained when we do IRE-minimisation without using odds ratios. Instead, we just set the constraint $\Pr_1(R \wedge \neg S) = c$ and let c go to zero. This yields Bovens' solution independent of what partition is used.

These considerations lead us to ask the following two questions. First, what exactly are the properties of the different functions that are used in these distance-based approaches (in particular RE and IRE)? When do they agree and when do they differ? To what extent are their recommendations dependent on how we present the problem (e.g. how we partition the outcome space)? Without an answer to these questions no distance-based approach can be defended.

Second, do we think there should be a unique rational answer to this type of updating problems in the first place? They are clearly beyond the scope of ordinary Bayesianism, and perhaps Bayesianism should just be agnostic about the rule an agent uses to update in Judy Benjamin-type cases. If the answer to this question is negative, the door is open to allowing Sarah and Marian to leave their probability for the antecedent unchanged while at the same time allowing Kate to leave her probability for the consequent unchanged (both of which seem at least *prima facie* rational).

We have considered a number of points due to Douven and Romeijn (2010) in relation to the Judy Benjamin problem, which is more generally the problem of how a Bayesian should update her belief function on information of the form "If A, then odds for $B_1, \ldots, B_n$ are as $c_1 : \cdots : c_n$". They give a set of desiderata and we have followed them in rejecting Van Fraassen (1981) and Van Fraassen et al.'s (1986) objection to one of them. We have shown that these desiderata fix the updated probability assignment completely, and do so in the same way that Adams conditioning (coined in Bradley 2005) would. To do this we have corrected Douven and Romeijn's generalization of a theorem due to Bradley. We have also considered an attempt to relate the desiderata to a distance-minimising approach and have corrected the mathematics there as well. Finally, we have pointed at some problems and ambiguities with Douven and Romeijn's approach, in particular for the case where the information received is a simple (non-probabilistic) conditional. We conclude that for either Douven and Romeijn's or any other solution to work, we will first need a convincing argument that this type of problem has a unique rational solution.

References

Bovens, L. (2010) "Judy Benjamin is a Sleeping Beauty". In *Analysis* 70 pp.23-26.

Bradley, R. (2005) "Radical Probabilism and Bayesian Conditioning". In *Philosophy of Science* 72 pp.342-364.

Douven, I. and J-W. Romeijn (2010) "A New Resolution of the Judy Benjamin Problem". In LSE Choice Group Working Papers 5.

van Fraassen, B. C. (1981) "A Problem for Relative Information Minimizers in Probability Kinematics". In *The British Journal for the Philosophy of Science* 32 pp.375-379.

van Fraassen, B. C., R.I.G. Hughes, and G. Harman (1986) "A Problem for Relative Information Minimizers, Continued". In *The British Journal for the Philosophy of Science* 37 pp.453-463.

Appendix: Proof of Theorem 8

Let C be an arbitrary proposition. For our purposes, the partition of strongest relevant propositions is $\{\neg A \wedge C, \neg A \wedge \neg C, A \wedge B_1 \wedge C, A \wedge B_1 \wedge \neg C, \ldots, A \wedge B_n \wedge C, A \wedge B_n \wedge \neg C\}$. For convenience, we abbreviate the names of the relevant parameters and variables as follows:

- $a_0 := \Pr_0(\neg A \wedge C)$ and $a_i := \Pr_0(A \wedge B_i \wedge C), i = 1, \ldots, n$,
- $b_0 := \Pr_0(\neg A \wedge \neg C)$ and $b_i := \Pr_0(A \wedge B_i \wedge \neg C), i = 1, \ldots, n$,
- $x_0 := \Pr_1(\neg A \wedge C)$ and $x_i := \Pr_1(A \wedge B_i \wedge C), i = 1, \ldots, n$,
- $y_0 := \Pr_1(\neg A \wedge \neg C)$ and $y_i := \Pr_1(A \wedge B_i \wedge \neg C), i = 1, \ldots, n$.

This leads to the following optimisation problem:

$$\min \quad f(\mathbf{x}, \mathbf{y}) := \sum_{i=0}^{n} a_i \log \frac{a_i}{x_i} + b_i \log \frac{b_i}{y_i}, \tag{2}$$

$$\text{subject to} \quad x_i + y_i = \frac{c_i}{c_1}(x_1 + y_1), \qquad i = 2, \ldots, n, \tag{3}$$

$$\sum_{i=0}^{n} x_i + y_i = 1, \tag{4}$$

$$x_i, y_i \geq 0, \qquad i = 0, \ldots, n. \tag{5}$$

Here, (2) gives the IRE function between $\Pr_0$ and $\Pr_1$ for the relevant partition, (3) gives the constraints imposed by the new information using the odds c_i, relating $x_i + y_i$ ($\Pr_1(A \wedge B_i)$) to $x_1 + y_1$ for $i \geq 2$ and (4) and (5) impose that the x_i and y_i are probabilities. We proceed by substituting (3) and (4) away, removing n variables from the problem. In particular, we substitute:

$$y_i = \frac{c_i}{c_1}(x_1 + y_1) - x_i, \qquad i = 2, \ldots, n, \tag{6}$$

$$y_0 = 1 - x_0 - \frac{x_1 + y_1}{c_1} \sum_{i=1}^{n} c_i. \tag{7}$$

This yields a new optimisation problem that is equivalent to the previous one.

$$\min g(\mathbf{x}, y_1) := \sum_{i=0}^{n} a_i \log \frac{a_i}{x_i} + b_0 \log \frac{b_0}{1 - x_0 - \frac{x_1+y_1}{c_1} \sum_{i=1}^{n} c_i} + \sum_{i=1}^{n} b_i \log \frac{b_i}{\frac{c_i}{c_1}(x_1 + y_1) - x_i}, \tag{8}$$

subject to

$$x_i, y_1 \geq 0, \qquad i = 0, \ldots, n, \tag{9}$$

$$\frac{c_i}{c_1}(x_1 + y_1) - x_i \geq 0, \qquad i = 2, \ldots, n, \tag{10}$$

$$1 - x_0 - \frac{x_1 + y_1}{c_1} \sum_{i=1}^{n} c_i \geq 0. \tag{11}$$

Solving this optimisation problem amounts to finding the new probability assignment $\Pr_1$. In order to solve the problem we first prove the following lemma.

Lemma 10 *The optimisation problem "*$\min g(\mathbf{x}, y_1)$*, subject to* (9), (10) *and* (11)*" has a stationary point* $(\mathbf{x}^*, y_1^*)$ *with* $y_1^* = \frac{b_1}{a_1+b_1} \frac{c_1}{\sum_{i=1}^{n} c_i}(1 - a_0 - b_0)$, $x_0^* = a_0$ *and* $x_i^* = \frac{a_i}{a_i+b_i} \frac{c_i}{\sum_{j=1}^{n} c_j}(1 - a_0 - b_0)$ *for* $i = 1, \ldots, n$.

Proof A stationary point is a point in the feasible region of an optimisation problem where all partial derivatives are zero. First note that the values of $\mathbf{x}^*$ and y_1^* are nonnegative by construction, as are the values

y_0^* and $y_2^*, \ldots, y_n^*$ implied by $(\mathbf{x}^*, y_1^*)$ and (6) and (7), which means that (9), (10) and (11) are satisfied. The partial derivatives of $g(\mathbf{x}, y_1)$ are:

$$
\begin{aligned}
\frac{\partial g(\mathbf{x}, y_1)}{\partial y_1} &= -\frac{b_1}{y_1} + \frac{b_0}{1 - x_0 - \frac{x_1+y_1}{c_1}\sum_{i=1}^n c_i}\frac{\sum_{i=1}^n c_i}{c_1} \\
&\quad - \sum_{i=2}^n \frac{b_i}{\frac{c_i}{c_1}(x_1 + y_1) - x_i}\frac{c_i}{c_1}, \\
\frac{\partial g(\mathbf{x}, y_1)}{\partial x_0} &= -\frac{a_0}{x_0} + \frac{b_0}{1 - x_0 - \frac{x_1+y_1}{c_1}\sum_{i=1}^n c_i}, \\
\frac{\partial g(\mathbf{x}, y_1)}{\partial x_1} &= -\frac{a_1}{x_1} + \frac{b_0}{1 - x_0 - \frac{x_1+y_1}{c_1}\sum_{i=1}^n c_i}\frac{\sum_{i=1}^n c_i}{c_1} \\
&\quad - \sum_{i=2}^n \frac{b_i}{\frac{c_i}{c_1}(x_1 + y_1) - x_i}\frac{c_i}{c_1}, \\
\frac{\partial g(\mathbf{x}, y_1)}{\partial x_i} &= -\frac{a_i}{x_i} + \frac{b_i}{\frac{c_i}{c_1}(x_1 + y_1) - x_i}, \qquad i = 2, \ldots, n.
\end{aligned}
$$

And indeed, if we evaluate these at $(\mathbf{x}^*, y_1^*)$ we get:

$$
\begin{aligned}
\frac{\partial g(\mathbf{x}^*, y_1^*)}{\partial y_1} =& \frac{\sum_{i=1}^n c_i}{c_1}\left(\frac{b_0}{1-a_0-(1-a_0-b_0)} - \frac{a_1+b_1}{1-a_0-b_0}\right) \\
& - \sum_{i=2}^n \frac{a_i+b_i}{\frac{c_1}{\sum_{j=1}^n c_j}(1-a_0-b_0)} \\
=& \frac{\sum_{i=1}^n c_i}{c_1}\left(1 - \frac{1-a_0-b_0}{1-a_0-b_0}\right) = 0, \\
\frac{\partial g(\mathbf{x}^*, y_1^*)}{\partial x_0} =& -\frac{a_0}{a_0} + \frac{b_0}{1-a_0-(1-a_0-b_0)} = 0, \\
\frac{\partial g(\mathbf{x}^*, y_1^*)}{\partial x_1} =& \frac{\sum_{i=1}^n c_i}{c_1}\left(\frac{b_0}{1-a_0-(1-a_0-b_0)} - \frac{a_1+b_1}{1-a_0-b_0}\right) \\
& - \sum_{i=2}^n \frac{a_i+b_i}{\frac{c_1}{\sum_{j=1}^n c_j}(1-a_0-b_0)} \\
=& \frac{\sum_{i=1}^n c_i}{c_1}\left(1 - \frac{1-a_0-b_0}{1-a_0-b_0}\right) = 0, \\
\frac{\partial g(\mathbf{x}^*, y_1^*)}{\partial x_i} =& -\frac{\sum_{j=1}^n c_j}{c_i}\frac{a_i+b_i}{1-a_0-b_0} + \frac{a_i+b_i}{\frac{c_i}{\sum_{j=1}^n c_j}(1-a_0-b_0)} \\
=& \; 0, \qquad i = 2, \ldots, n. \qquad \square
\end{aligned}
$$

The relative entropy function f is known to be convex. Since g takes the same values as f but on a domain restricted by linear constraints (which is hence a convex domain) g is also a convex function. For a convex function g with a stationary point $(\mathbf{x}^*, y_1^*)$ in its domain it is known that $g(\mathbf{x}^*, y_1^*)$ is the minimum of g on that domain. Hence by lemma 10 $(\mathbf{x}^*, y_1^*)$ solves the optimisation problem and yields the values for $\Pr_1$ (using (6) and (7) to obtain the missing y_i^*'s). In particular, it holds for all i that

- $\Pr_1(A) = \sum_{i=1}^{n} x_i^* + y_i^* = \sum_{i=1}^{n} \frac{c_i}{\sum_{j=1}^{n} c_j}(1 - a_0 - b_0)$
  $= 1 - a_0 - b_0 = \Pr_0(A),$
- $\Pr_1(C|A \wedge B_i) = \frac{x_i^*}{x_i^* + y_i^*} = \frac{a_i}{a_i + b_i} = \Pr_0(C|A \wedge B_i),$
- $\Pr_1(C|\neg A) = \frac{x_0^*}{x_0^* + y_0^*} = \frac{a_0}{a_0 + b_0} = \Pr_0(C|\neg A).$ □