Why Bayesian Psychology Is Incomplete

Frank Döring†

Dept. of Philosophy
University of Cincinnati

Why Bayesian Psychology Is Incomplete

0. Abstract. Bayesian psychology, in what is perhaps its most familiar version, is incomplete: Jeffrey conditionalization is not a complete account of rational belief change. Jeffrey conditionalization is sensitive to the order in which the evidence arrives. This order effect can be so pronounced as to call for a belief adjustment that cannot be understood as an assimilation of incoming evidence by Jeffrey's rule. Hartry Field's reparameterization of Jeffrey's rule avoids the order effect but fails as an account of how new evidence should be assimilated.

1. Introduction. The following is an exercise in Bayesian rational psychology. Its contention is that Jeffrey conditionalization, of which classical conditionalization is a special case, is not a complete account of rational belief change. Bayesian psychology, in what is perhaps its most familiar version, is incomplete.

Bayesianism comes in many flavors. The Bayesian theorist I have in mind represents the beliefs of a rational agent by a probability function that is continually updated by conditionalizing on the incoming evidence. In the classical version of the story, evidence takes the form of a proposition receiving posterior probability 1. In Richard Jeffrey's version, the probability change is allowed to originate from new probability assignments short of unity to the members of a set of propositions that partitions the probability space. Classical conditionalization is Jeffrey conditionalization on an evidence partition consisting of just one proposition and its negation, with associated probabilities 1 and 0.

It is well known that classical conditionalization is irreversible, in the sense that no learning episode can be undone by any subsequent applications of the classical rule. There are, however, occasions on which the agent will want to undo a present commitment and revert to an earlier belief state. Such change lies beyond the purview of classical Bayesianism. For one thing, it requires that the agent keep a detailed record of earlier beliefs, which is not a standard Bayesian requirement. Secondly, it is a change that cannot be accomplished by conditionalization. Jeffrey conditionalization appears to obviate the need for this sort of change. For as long as no contingent proposition receives probability extremes 0 or 1, the effect of any update by Jeffrey's rule can be undone by subsequent applications of the same rule. As I hope to show, this fact does not obviate the need for a different form of belief change after all.

The outcome of a sequence of updates by Jeffrey conditionalization depends on the order in which the updating steps are carried out. Jeffrey pointed this out early on but didn't make much of it. Others began to worry, until Hartry Field (1978) showed how to reconceptualize the input into Jeffrey's rule in such a way as to make the order effect disappear. But Field's remedy, I will argue, is ineffective. The reason is that he assigns experience a role for which it is simply unsuited. This leaves us with Jeffrey's rule in its original form. Its order effect is serious enough to call for an occasional adjustment of beliefs that is not conditionalization on the evidence. Consequently, the rule provides at best a partial account of rational belief change.

Jeffrey never defended his rule as a complete account, nor am I aware of such a defense for the classical rule. But the fact that alternative approaches to belief change have had very little impact on Bayesian thinking suggests to me that there is too much complacency in the Bayesians community as regards the scope of its favorite accounts.

2. Classical Conditionalization. That classical conditionalization cannot be a complete account of rational belief change is widely appreciated. I shall therefore only briefly rehearse the case here. A classical Bayesian agent learns from experience by trading in a prior probability function $\underline{p}$ for a posterior function $\underline{p}_{\underline{e}}$ such that, for every proposition $\underline{x}$ in the common domain of $\underline{p}$ and $\underline{p}_{\underline{e}}$, $\underline{p}_{\underline{e}}(\underline{x}) = \underline{p}(\underline{x}|\underline{e})$, where $\underline{e}$ is a proposition that captures the deliverance of experience and $\underline{p}(\underline{e}) \neq 0$. The restriction to positive $\underline{p}(\underline{e})$ follows from the standard definition of conditional probability as a ratio of unconditional probabilities: the relevant conditional probabilities $\underline{p}(\underline{x}|\underline{e}) =_{def} \underline{p}(\underline{x}\ \&\ \underline{e})/\underline{p}(\underline{e})$ are defined only for positive $\underline{p}(\underline{e})$. The change sets $\underline{p}_{\underline{e}}(\underline{e})$ to 1 since $\underline{p}(\underline{e}|\underline{e}) = 1$, that is, in the new probability function, the evidence is accepted as certain.

Changes by conditionalization are irreversible in two respects. First, no subsequent change by conditionalization can undo the effect of an earlier learning episode. The reason is that evidence accepted in the past will always retain probability 1 since, whenever $p(e_1) = 1$ and $p(e_2) \neq 0$, we have $p(e_1 \,\&\, e_2)/p(e_2) = p(e_2)/p(e_2) = 1$ so that, by conditionalization, $p_{e_2}(e_1) = p(e_1|e_2) = p(e_1 \,\&\, e_2)/p(e_2) = 1$. This means that earlier probability functions cannot be recovered by conditionalization. Second, new evidence that has been rejected at some point by having its probability set to zero cannot be accommodated later because the relevant conditional probabilities are no longer defined.

Ever since sense datum languages and Protokollsätze fell into disrepute, rational agents have had to contend with their fallibility as regards even the most basic evidence propositions. Even a perfectly rational Bayesian agent will, from time to time, have to revise his or her credence in an evidence proposition accepted at an earlier point. There are, presumably, more or less rational ways of performing this sort of change. Classical conditionalization in no way distinguishes between them. It is therefore not a complete account of rational belief change.

3. Jeffrey Conditionalization. Three decades ago, Richard Jeffrey proposed a rule that is more pliable and forgiving than classical conditionalization. The rule does not require probability 1 for evidence propositions, and its effects are reversible by subsequent application of the same rule. Suppose that experience causes the agent to assign new probabilities p_{new} to the cells $e_1,\ldots,e_n$ of a (finite) partition E of mutually exclusive and jointly exhaustive propositions. (As in the classical case,

$p_{old}(e)$ must be positive for all $e \in E$ so that $p_{old}(x|e)$ is defined.) Then the new probability assignment to an arbitrary proposition is given by

$$p_{new}(x) = \sum_{e \in E} p_{new}(e) \cdot p_{old}(x|e).$$

Jeffrey derives this rule from the condition that $p_{new}(x|e) = p_{old}(x|e)$ for each e in E for which $p_{new}(e) > 0$. This condition sometimes goes by the name *rigidity*. If one cell in the partition receives posterior probability 1, the prescriptions of Jeffrey's and the classical rule coincide. Intuitively, what goes on in Jeffrey conditionalization is that the agent's probability space gets carved up into disjoint regions corresponding to the cells of the partition, and then each of these regions is enlarged or shrunk independently of the others, subject to two provisos: that the regions' joint probability after the change equal 1 (normalization), and that all probability ratios within each region — the $p(x|e)/p(y|e)$ — be left intact (rigidity). We may think of each region as the surface of a balloon that is being inflated or deflated. The ratio between any two areas on a balloon's surface is unaffected by inflation or deflation. It is clear that, as long as no balloons are popped (no cell receives probability 0), such change is reversible. In classical conditionalization, by contrast, one of the balloons takes over while its complement vanishes. In the sequel, I will focus on the simplest case of Jeffrey conditionalization, in which the evidence partition for each learning episode consists of just two cells, representing one evidence proposition and its negation.

Jeffrey conditionalization is sensitive to the order in which the evidence comes in. This means that the probability function that results from first assigning some probability value r_1 to e_1 (and thus $1 - r_1$ to $\neg e_1$) and then probability r_2 to e_2 may differ from the function that results from first assigning r_2 to e_2 and subsequently assigning r_1 to e_1. Jeffrey notes this order dependence in passing as a curious but perfectly sensible feature of his scheme. The effect can, however, be quite dramatic and unreasonable, as the following schematic example shows. Let the agent's initial probability function be partially described by the following assignment:

$$p(AB) = p(\neg AB) = .05$$

$$p(A\neg B) = p(\neg A\neg B) = .45$$

Probabilities for all truth functions of A and B are thereby determined. In particular, p(A _ B) = p(AB) + p(¬AB) + p(A¬B) = .55. We shall now compare two updating sequences. In the first sequence, experience initially raises the probability of A _ B to 0.99 (thus lowering the probability of ¬A¬B to 0.01) and subsequently raises the probability of ¬A _ B to 0.99; in the second sequence these events occur in reverse order. According to Jeffrey's rule, raising the probability of A _ B to .99 has the following effect on the probability of AB:

p’(AB) = p’(A _ B) . p(AB | A _ B) + p’(¬(A _ B)) . p(AB | ¬(A _ B))

= p’(A _ B) . p(AB) / p(A _ B) + p’(¬(A _ B)) . p(A¬A) / p’(¬(A _ B))

= 0.99 . 0.05 / 0.55 + 0.01 . 0

= 0.09.

Similar calculations yield the values for A¬B and ¬AB. The two updating sequences can be tabulated as follows:

	A	¬A			A	¬A			A	¬A
B	5	5	_	B	9	9	_	B	47	47
¬B	45	45		¬B	81	1		¬B	1	5
					p’(¬A¬B) = 1%				p”(A¬B) = 1%	

	A	¬A			A	¬A			A	¬A
B	5	5	_	B	9	9	_	B	47	47
¬B	45	45		¬B	1	81		¬B	5	1
					p*(A¬B) = 1%				p**(¬A¬B) = 1%	

<u>Figure 1</u>

Probabilities are given here as percentages. The thick black line in each table divides the two cells of the evidence partition that underlies the change to the assignment depicted by that table. Focus on the bottom rows in the two rightmost tables. At the end of the first updating sequence, the probability of A¬B is one fifth of the probability of ¬A¬B; at the end of the second updating sequence it is the other way round. Equivalently, the probability of A given ¬B is 1/6 after sequence 1 and 5/6 after sequence 2. By playing with the numbers, this discrepancy can be brought as close to 1 as you please. If this doesn't seem dramatic enough yet, imagine a third updating step in which the probability of B is lowered close to 0. This would force the unconditional probabilities for A and ¬A near 0 and 1, with the roles of A and ¬A reversed in the two sequences. This seems wholly unjustified if there is nothing essential about the order in which experiences were made. What we would like to see is a symmetrical final state, one in which A¬B and ¬A¬B receive about the same probability.

We can shed some light on what is going on here by viewing the situation from the vantage point of classical conditionalization. Brian Skyrms showed that the effect of Jeffrey conditionalization can be obtained by classical conditionalization in a sufficiently rich probability space (cp. Jeffrey 1988). In order to mimic, for example, step 1 in the first sequence, we would have to embed the problem in a space that provides a proposition C expressing what is learned with certainty such that p(A _ B | C) = 0.99 and p($\underline{x}$ | (A _ B)C) = p($\underline{x}$ | A _ B) as well as p($\underline{x}$ | ¬A¬BC) =

$p(\underline{x} \mid \neg A \neg B)$. (The latter two requirements ensure that the condition from which Jeffrey derived his rule is met: probabilities conditional on the cells of the evidence partition are not affected by the change. C has to be "screened off" from any $\underline{x}$ by all cells of the evidence partition.) Classical conditionalization is not sensitive to the order of the evidence because we always have

$$\underline{p}_{\underline{y}\underline{z}}(\underline{x}) = \underline{p}_{\underline{y}}(\underline{x} \mid \underline{z}) = \underline{p}(\underline{x} \mid \underline{y}\underline{z}) = \underline{p}_{\underline{z}}(\underline{x} \mid \underline{y}) = \underline{p}_{\underline{z}\underline{y}}(\underline{x}).$$

It follows that if the two updating sequences in Figure 1 are viewed as the effects of classical conditionalization in a richer probability space, then the pair of evidence propositions for the first sequence must be different from the pair of evidence propositions for the second sequence. The difference will consist, presumably, in some reference to the order in which the evidence arrived. But intuitively, I think, this is a difference that shouldn't make a difference.

The order effect is an embarrassment against which a rational agent will want to safeguard. This requires, in the first place, a memory of previous belief states enabling the agent to notice the effect. I see no problem endowing the agent with such a faculty. The memory will be of obvious use in counteracting the effect. The most straightforward approach would be to revert to the original probabilities and try to assimilate the later evidence all at once. There is no unique recipe for this. The easiest solution in the case at hand is to merge the two updating steps into one by compounding their constraints (the respective posteriors). This requires that the constraints be compatible, which, in this case, they are. The new assignment to accomodate would be: $p(A\neg B) = p(\neg A\neg B) = 0.01$. In order to apply Jeffrey's rule, we need an evidence partition. We can easily construct one by complementing the two propositions whose probabilities were directly affected, so that the change now originates in $\{A\neg B, \neg A\neg B, B\}$, with $p(B) = 1 - (p(A\neg B) + p(\neg A\neg B)) = 0.98$. Jeffrey conditionalizing in one step on the new assignments yields a posterior distribution in which both pieces of evidence receive equal weight:

	A	¬A			A	¬A
B	5	5	–	B	49	49
¬B	45	45		¬B	1	1

Figure 2

There are many alternatives to this solution. One is to fuse the two pieces of evidence together using a rule first proposed by Arthur Dempster, and subsequently Jeffrey-conditionalize on the new evidence probabilities. In the present example, this will generate practically the same result as the first procedure. In general, however, the two procedures are not equivalent. Nor are they universally applicable.

Since a detailed discussion of Dempster's rule would take us too far afield, I offer only a brief hint for the cognoscienti, based on Shafer (1976). The "frame of discernment” Θ in the example is the set of all maximal conjunctions that can be formed from A and B and their negations. (You may think of Θ as a set of possible worlds.) Θ’s powerset, 2^{Θ}, is accordingly the set of all truth functions of A and B. The content of the first piece of evidence can be represented by the (convex) set of all probability functions over 2^{Θ} that assign 0.01 to ¬A¬B and 0.99 to A _ B. This set happens to correspond to a Dempster-Shafer belief function Bel_1 over 2^{Θ} with corresponding assignment m_1 of “basic probability numbers” to ¬A¬B and A _ B of 0.01 and 0.99, respectively. ¬A¬B and A _ B are the “focal elements” of m_1, the only propositions in 2^{Θ} to which the function assigns a non-zero value. The second piece of evidence determines in like fashion an assignment m_2 to A¬B and ¬A _ B of 0.01 and 0.99, respectively. (Warning! This construction won’t always work since there is no general correspondence between convex sets of classical probability functions and Dempster-Shafer belief functions.) The two assignments m_1 and m_2 can be merged, via Dempster’s rule, into their “orthogonal sum” m, which works out to be:

$m(\neg A\neg B) = m(A\neg B) = 0.01 \,.\, 0.99 / (1 - 0.01^2) \approx 0.01$; $m(B) = 0.99^2 / (1 - 0.01^2) \approx 0.98$.

Since the focal elements of m form a partition on 2^{Θ}, and since the values that m assigns to the cells of the partition sum to 1, we can interpret m directly as a (partial) classical probability function assigning values m(¬A¬B), m(A¬B), and m(B) to ¬A¬B, A¬B, and B, respectively. Jeffrey conditionalization on the partition {¬A¬B, A¬B, B} with these values yields the result in the second table in Figure 2 to within

1/1000 of a percentage point. The fact that this result is intuitively attractive in the example should not be read as an endorsement of Dempster's rule. I, for one, am skeptical about the rationality of the rule.

It is not important for the present discussion whether or not there is a unique rational procedure for assessing the joint impact of probabilistic evidence. What is important is that there are more or less rational approaches to the problem and that Jeffrey conditionalization offers no solution. Consequently, Jeffrey conditionalization cannot be all there is to rational belief change.

4. Field's Proposal. Hartry Field (1978) offered a recipe to circumvent the order effect of Jeffrey conditionalization. He suggests that Jeffrey's rule is wrongly parameterized in terms of posterior probabilities $\underline{p}_{new}(\underline{e})$ as "input factors". The rule, he argues, should instead be parameterized in terms of a factor $\underline{\alpha}$ defined as $\underline{p}_{new}(\underline{e})/\underline{p}_{old}(\underline{e}) \cdot \underline{p}_{old}(\neg\underline{e})/\underline{p}_{new}(\neg\underline{e})$. (Field actually uses '$\underline{\alpha}$' for the logarithm of this factor, but since the logarithm is introduced solely for convenience, I shall omit it.) Experience, says Field, delivers these factors, not posterior probabilities. The factors record the rate of change in odds rather than the absolute result of the change. (Given the usual definition of odds — $\underline{o}(\underline{x}) = \underline{p}(\underline{x})/(1-\underline{p}(\underline{x}))$ —, $\underline{\alpha} = \underline{o}_{new}(\underline{e})/\underline{o}_{old}(\underline{e})$. $\underline{\alpha}$ is in effect a <u>Bayes factor</u> between $\underline{e}$ and $\neg\underline{e}$.) Whereas Jeffrey's original rule gives the new probability assignment as a function of the old assignment and new probabilities for the cells in an evidence partition, Field conditionalization gives the new assignment as a function of the old assignment and a <u>change factor</u> for the probabilities of the propositions in the evidence partition. The form in which change is propagated through the rest of the agent's probabilities is the same on both accounts. But Field's rule, unlike Jeffrey's, is commutative. Changing first the odds for $\underline{e}_1$ by a factor $\underline{r}_1$ and then the odds for $\underline{e}_2$ by a factor $\underline{r}_2$ results in the same final probabilities as proceeding in reverse order, first changing the odds for $\underline{e}_2$ by $\underline{r}_2$ and then changing the odds for $\underline{e}_1$ by $\underline{r}_1$. In Field's view, the order effect we observed earlier is simply tied to a wrong conception of what constitutes the input into Jeffrey's rule.

Field conditionalization is certainly attractive. But can it work? Does experience issue change factors not probabilities? Field offers one small piece of theoretical evidence for his claim that this is what experience does or anyway should do. He points out that the logarithm of the change factor is 0 when there is no change in odds and suggests that 0 is the right value to represent an "uninformative" episode of sensory stimulation. The problem with this suggestion is that informativeness is clearly relative to agents' beliefs. For example, whether the sound of the ambulance siren is informative for me depends on whether I already noticed the ambulance car. The input parameter, by contrast, should be belief independent, as Field insists elsewhere in his paper. It "represents the effects that the sensory stimulation has <u>by itself</u> (independently of the value of $\underline{p}$ [the agent's prior probability function])" (Field 1978, p. 363). Yet reasonable change in odds is clearly not belief independent. Consider Jeffrey's sort of example. A piece of cloth looks bluish green under neon lights. How should the observation affect your probability (or odds) for the cloth's being blue? It clearly depends. If your prior for blue was very low, you should move to a higher posterior; if your prior was very high, you should move to a lower posterior. The magnitude and even the direction of the change depends on your priors. Yet if the observation issued Field's factor $\underline{\alpha}$, the direction of change would have to be the same in either case. Since this would be wrong, the observation had better not issue $\underline{\alpha}$.

To say that the input factor cannot be an odds-change factor is one thing; to say that it is a posterior probability is another. For Field, the factor cannot be a posterior probability because posteriors are dependent on priors, not just on the evidence: "it is clear that the probability $\underline{q}$ which I attach to an observation sentence $\underline{E}$ after I have been subjected to a sensory stimulation will depend not only on the sensory stimulation but also on the probability I attached to $\underline{E}$ before the stimulation" (ibid. p. 363). I find this less evident than Field. Yet I know of no proof that experience furnishes posteriors that override the agent's priors. I am therefore willing to concede that there may be cases where priors should be given some weight. But this is perfectly compatible with thinking of experience as furnishing probabilities. Posteriors over the evidence partition can then be seen as resulting from a compromise between the probabilities drawn from experience and the agent's priors, and updating can proceed by Jeffrey's original rule.

Suppose that in the example in Figure 1 the agent does not simply accept as posteriors the probabilities over the evidence partitions issued by experience. Let the agent instead determine posteriors as averages of the priors and the probabilities from experience. This slightly dampens the order effect, but it doesn't change its nature. In Figure 3, two change sequences are tabulated. In the first sequence, the priors enter into the average calculation for the posteriors to be fed into Jeffrey's rule with a weight of 0.1 while the incoming evidence is accorded a weight of 0.9. In the second sequence, both weights are 0.5. The order effect shows up in the difference between the cells in the bottom rows of the two rightmost tables. Both sequences possess "mirror images" in which a reversal of the updating steps leads to the same reversal of the probabilities in these two cells that we observed in Figure 1.

	A	¬A
B	5	5
¬B	45	45

–

	A	¬A
B	8.6	8.6
¬B	77.4	5.4

p'(¬A¬B) =
0.1.p(¬A¬B) + 0.9.0.01

–

	A	¬A
B	34.8	34.8
¬B	8.6	21.8

p"(A¬B) =
0.1.p'(A¬B) + 0.9.0.01

	A	¬A
B	5	5
¬B	45	45

–

	A	¬A
B	7	7
¬B	63	23

p'(¬A¬B) =
0.5.p(¬A¬B) + 0.5.0.01

–

	A	¬A
B	12.9	12.9
¬B	32	42.2

p"(A¬B) =
0.5.p'(A¬B) + 0.5.0.01

Figure 3

So even if experience doesn't determine posteriors all by itself, Jeffrey conditionalization may well remain applicable, displaying the same order effect as before. I conclude that Field has neither shown that Jeffrey conditionalization is inapplicable nor that his own scheme is a viable alternative.

5. Incremental Updating. Both standard Bayesian accounts of belief change are incomplete: classical conditionalization because it is irreversible, Jeffrey conditionalization because of its order effect. In either case, the limitation is due to a conception of learning as <u>incremental</u> updating. The impact of new evidence is exhausted by the change it effects in the agent's probability function at the time at which it arrives. There is no prevision in the model for an assessment of the joint impact of pieces of evidence that arrived at different times. Yet it is exactly this sort of considered "look back" that would support the preferred symmetric solution in the example, one where there is no (or no significant) discrepancy in the probabilities for A¬B and ¬A¬B.

Incremental updating is rational insofar as it minimizes change in belief at each step in the learning process (probabilities conditional upon the evidence remaining fixed). Yet, as we saw, locally minimal changes can add up to a change that violates more global constraints of minimality or symmetry. The Bayesian procedures for incremental belief change thus fail to serve globally the very same ideal of rationality they serve locally. For the purpose of globally (or less locally) minimal change, procedures of a different sort are required. In a more complete account, the incremental updating schemes will have to be supplemented with non-incremental schemes. Naturally, demanding such supplementation is one thing, providing it is quite another. I suspect that the same need for supplementation arises for non-Bayesian schemes of belief management, but I cannot argue this point here because the schemes in question are not yet sufficiently developed to determine sequences of belief revision. In the currently most developed approach to the revision of beliefs as represented by sets of accepted sentences or propositions, the Alchourrón-Gärdenfors-Makinson approach, no revision rules have been specified that would allow iterated revision. The problem is that the rules that have been proposed exploit some parameter like epistemic entrenchment that is external to the representation of belief, and it is unclear how this parameter should be adjusted when beliefs are revised.[1]

References

Boutilier, Craig (1995), "On the Revision of Probabilistic Belief States." Notre Dame Journal of Formal Logic 36.1: 158-183.

Field, Hartry (1978), "A Note on Jeffrey Conditionalization." Philosophy of Science 45: 361-367.

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

Jeffrey, Richard C. (1988), "Conditioning, Kinematics, and Exchangeability," in Brian Skyrms and William L Harper (eds.), Causation, Chance, and Credence. Dordrecht: Kluwer, Vol. 1, pp. 221-255.

Shafer, Glenn (1976), A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press.

Footnotes

† Thanks to Chris Gauker and Rob Rynasiewicz for helping me sort out my thoughts on Bayesian updating.

1.. For an overview over the state of the art, see the Winter 1995 special issue of the Notre Dame Journal of Formal Logic, in particular Boutilier's contribution.