Franz Dietrich and Christian List
Reasons for (prior) belief in Bayesian epistemology

Article (Accepted version)
(Refereed)

http://eprints.lse.ac.uk

Reasons for (prior) belief in Bayesian epistemology[*]

Franz Dietrich

CNRS & University of East Anglia

Christian List

London School of Economics

8 January 2012

Abstract

Bayesian epistemology tells us with great precision how we should move from prior to posterior beliefs in light of new evidence or information, but says little about where our prior beliefs come from. It offers few resources to describe some prior beliefs as rational or well-justified, and others as irrational or unreasonable. A different strand of epistemology takes the central epistemological question to be not how to change one's beliefs in light of new evidence, but what reasons justify a given set of beliefs in the first place. We offer an account of rational belief formation that closes some of the gap between Bayesianism and its reason-based alternative, formalizing the idea that an agent can have reasons for his or her (prior) beliefs, in addition to evidence or information in the ordinary Bayesian sense. Our analysis of reasons for belief is part of a larger programme of research on the role of reasons in rational agency (Dietrich and List 2012a,b).

Keywords: Bayesian epistemology, doxastic reasons, prior and posterior beliefs, principle of insufficient reason, belief formation, belief change

1 Introduction

Bayesian epistemology tells us how we should change our beliefs in light of new evidence or information. It specifies what the impact of any given item of evidence or information should be on our beliefs, and what posterior beliefs we should arrive at after receiving such evidence or information (for an introduction, see, e.g., Bovens and Hartmann 2003). This picture of rational belief updating assumes, however, that we have started out with some *prior* beliefs, and one of Bayesianism's most widely recognized shortcomings is that it does not say anything about where those priors come from. Bayesianism by itself does not offer the resources to endorse some

1

prior beliefs as rational or appropriate, and to criticize others as irrational or at least unreasonable. Yet, we intuitively think of some prior beliefs as more reasonable than others. We tend to consider conspiracy theories, for example, less credible than more ordinary explanations of the same phenomena, even if we do not have enough evidence to rule them out in advance. In consequence, we tend to regard someone who assigns high prior probabilities to far-fetched theories as less reasonable than someone who assigns lower (albeit non-zero) probabilities to them. Nonetheless, a conspiracy theorist may be impeccably rational from a purely Bayesian perspective. Of course, there are proposals on how to distinguish 'good' Bayesian priors from 'bad' ones – the former are sometimes required to maximize entropy, for example – but these stipulations often seem rather *ad hoc* and do not generally 'ground' an agent's beliefs in a philosophically compelling manner. Thus Bayesian epistemology still lacks an established account of the formation of prior beliefs.

A very different but increasingly popular strand of epistemology takes the central epistemological question to be not how an agent should rationally move from his or her prior beliefs to his or her posterior ones, but what reasons justify a given set of beliefs in the first place. The focus of this approach is not on the constraints governing the relationship between prior and posterior beliefs, but on the way beliefs can be grounded or 'anchored' in certain underlying reasons. A key feature of a rational belief, on this picture, is that it is held for the right reasons, or at least for what the agent takes to be the right reasons. Reasons for belief are not the same as evidence or information, but they are, in some sense, more fundamental (on the distinction between evidence and reasons for belief, see, e.g., Foley 1991). This raises the question of what such reasons are, and below we propose a definition. Crucially, the notion of a reason for belief is absent from Bayesian epistemology.

The aim of this paper is to offer a new account of rational belief formation that closes some of the gap between Bayesian epistemology and its reason-based alternative. We formalize the idea that an agent can have reasons for his or her (prior) beliefs, in addition to the agent's evidence or information in the ordinary Bayesian sense. Our analysis of reasons for belief is part of a larger programme of research on the role of reasons in rational agency, and we have offered an analysis of reasons for preference and action elsewhere (Dietrich and List 2012a,b).[1] The formal concepts and results presented in this paper draw on that earlier work.

The paper is structured as follows. In Section 1, we introduce the way we think about beliefs and belief formation, distinguishing between ordinal and cardinal kinds of belief. In Section 2, we define the notion of a *doxastic reason* (a reason for belief) and discuss what it means for a proposition to become a doxastic reason for an agent. In Section 3, we state our first main representation theorem, which applies to ordinal

---

[1] Other related works on reasons for preferences include von Wright (1963), Liu (2010), and Osherson and Weinstein (2012).

beliefs, formally represented by credence orders over different epistemic possibilities. This result, which is an interpretationally new variant of an earlier theorem (Dietrich and List 2012a,b), shows that if two simple axioms are satisfied, an agent's beliefs across variations in his or her doxastic reasons can be represented in terms of a single underlying binary relation, which we call a *credibility relation*. This relation ranks different combinations of doxastic reasons in an order of credibility, which can be interpreted positively or normatively. Under a positive interpretation, it captures how strongly the agent is disposed to believe the various epistemic possibilities picked out by the different combinations of doxastic reasons. Under a normative interpretation, it captures how strongly the agent ought rationally to believe these possibilities. In Sections 4 and 5, we give some examples and draw attention to the surprising possibility of an intransitive credibility relation in the presence of fully consistent beliefs. In Section 6, we turn to the case of cardinal beliefs, formally represented by credence functions, and state our second main theorem, which provides a reason-based representation of an agent's cardinal beliefs. In Section 7, we briefly address a possible objection and reinforce our claim that reason-based belief formation cannot easily be redescribed in standard Bayesian terms. In Section 8, we make some concluding remarks.

2   Beliefs and belief formation

We consider an agent's beliefs in some set X of *epistemic possibilities*. These possibilities can take a number of forms, depending on the context in which we wish to apply our framework. They can be possible worlds or states of the world, or alternative theories or hypotheses. All we assume is that the elements of X (i) are mutually exclusive and (ii) jointly exhaust the relevant possibilities.

A subset of X is called a *proposition*. It is said to be *true* of those possibilities contained in it, and *false* of all other possibilities. This terminology should be familiar in the case in which X is the set of all relevant possible worlds. While the standard definition of a proposition is *extensional* (identifying any proposition with the subset of X it picks out), it is often useful to model propositions as *intensional* objects, representing them for instance by sentences from a suitable language, and in a number of places below we implicitly do so.

In Bayesian epistemology, the agent's beliefs are usually represented by a *credence function* (subjective probability function) on X, which assigns to each possibility in X a real number between 0 and 1, with a sum-total of 1. To keep things simple, however, we begin by representing an agent's beliefs in an ordinal rather than cardinal way. So we represent the agent's beliefs by a *credence order* $\succsim$, a complete and transitive binary relation on X. For any two possibilities x and y, $x \succsim y$ means that the agent believes x at least as strongly as y; or, in the language of degrees of belief, the agent's

degree of belief in x is greater than or equal to his or her degree of belief in y. We further write $x \succ y$ if $x \succsim y$ but not $y \succsim x$ (the agent believes x more strongly than y), and $x \sim y$ if $x \succsim y$ and $y \succsim x$ (the agent believes x and y equally strongly). Later we return to the more standard representation of beliefs by credence functions.

Bayesian epistemology gives an account of how an agent's beliefs should rationally change in response to certain evidence or information. Suppose, for example, the agent receives some evidence that rules out some of the possibilities in X. He or she must then change his or her credence order in such a way as to rank any possibilities ruled out by the evidence below (or at least weakly below) any possibilities not ruled out, while not changing any other rankings. This is a form of *Bayesian updating*.

Here, however, we want to focus on how the agent arrives at his or her beliefs in the possibilities in X in the first place, before receiving any evidence that rules out certain possibilities. We call this the problem of *belief formation*. We can look at this problem from both positive and normative perspectives, i.e., we can ask either how an agent *actually* forms his or her beliefs, or how he or she *ought* rationally to do so. It is at this point that reasons come into play.

3   Reasons for belief

Reasons can be conceptualized in a number of ways. Some philosophers think of reasons as facts of a certain kind; others as certain kinds of properties of the possibilities under consideration; still others as mental states of the agent. Thomas Scanlon (1998, p. 67) defines a reason as 'a consideration that counts in favor of some judgment-sensitive attitude [e.g., belief or preference]'. We prefer to adopt a slightly more general definition, preserving the 'counting' aspect, but dropping the 'in favor of' aspect of Scanlon's definition. Thus, we think of a reason as a consideration – more precisely, a proposition – that has a particular kind of *relevance* (that 'counts') in relation to an agent's intentional attitudes, but where the relevance in question (the relevant *kind* of 'counting') can be spelt out in a variety of ways. It need not be as simple as counting 'in favor of' or 'against' something. The force of a particular reason might depend, for example, on which other reasons are present. We here focus on reasons for belief, rather than reasons for preference or action. We call such reasons *doxastic reasons*.

Since we want to develop a framework that can be interpreted positively as well as normatively, we further distinguish between *positive* and *normative* doxastic reasons. A *positive doxastic reason* is a proposition that is relevant for the agent's *actual* beliefs: it may affect the agent's actual belief in any possibility of which it is true vis-à-vis other possibilities. A *normative doxastic reason*, by contrast, is a proposition that is relevant for the beliefs the agent *ought* rationally to have: it may constrain the belief the agent ought to have in any possibility of which it is true vis-à-vis other

possibilities. Our definition scheme leaves open *how* a proposition must affect or constrain the agent's actual or ideally rational beliefs to count as a doxastic reason of one of these kinds. This may depend on factors such as the context and which other doxastic reasons are present. Our formal results below allow us to say more about these dependencies.

To illustrate the notions of positive and normative doxastic reasons, suppose the elements of X are scientific hypotheses. If a scientist has a high degree of belief in a particular hypothesis because of its elegance, the proposition that the hypothesis is elegant is a positive doxastic reason for the scientist's belief. This proposition is relevant for the scientist's *actual* beliefs, insofar as it affects the belief he *actually* has with respect to the given hypothesis, of which the proposition is true. If the scientist has a high degree of belief in a particular hypothesis despite its inconsistency, the proposition that the hypothesis is inconsistent is a normative doxastic reason for him *not* to believe it so strongly, contrary to his actual belief. This proposition is relevant for the beliefs he ought rationally to have: he ought to have a low degree of belief in any hypothesis of which the proposition is true (i.e., any inconsistent hypothesis), even though he does not actually do so.

Let D denote the set of doxastic reasons for the agent's beliefs regarding X in a given context or doxastic state. Depending on the intended interpretation, the reasons in D can be either positive or normative. Thus the set D contains all the propositions that are relevant for the agent's beliefs in the appropriate sense: either positively or normatively. The set D need not be, and often will not be, consistent. Different doxastic reasons may be true of different possibilities in X and may therefore pull the agent's beliefs in different directions. Our aim is to analyze the interplay between different doxastic reasons.

To mark the fact that the agent's beliefs depend on his or her set of doxastic reasons, we append the subscript D to the symbol $\succsim$, interpreting $\succsim_D$ as the agent's (actual or ideally rational) credence order in the event that D is his or her set of doxastic reasons in relation to the possibilities in X. Our analysis focuses on how $\succsim_D$ depends on D. From a positive perspective, the dependency of interest is that between the agent's set of *positive* doxastic reasons (represented by D under the positive interpretation) and the beliefs he or she *actually* has (represented by $\succsim_D$ under the positive interpretation). From a normative perspective, it is the dependency between the agent's set of *normative* doxastic reasons (represented by D under the normative interpretation) and the beliefs he or she *ought rationally* to have (represented by $\succsim_D$ under the normative interpretation). In line with our earlier notation, we write $\succ_D$ to represent the strict order corresponding to $\succsim_D$, and $\sim_D$ to represent the equivalence relation.

4 When does a proposition become a doxastic reason?

Not every proposition pertaining to the possibilities in X will be a doxastic reason for the agent's beliefs in a given context or doxastic state. As a matter of positive fact, only some propositions will be sufficiently salient for the agent to play the role of doxastic reasons for his or her beliefs. Which propositions do so in any given situation is a contingent psychological matter, which cannot be settled *a priori*.[2] Normatively, moreover, only some propositions are epistemically relevant for the agent's beliefs. Different epistemological theories give different answers to the question of which propositions they are. For example, is the fact that a particular scientific hypothesis is parsimonious or elegant a normative reason to believe the hypothesis? Is the fact that a particular scenario is psychologically salient also a normative reason to consider that scenario credible? Is a theory that has the hallmarks of a conspiracy theory less likely to be true? Different epistemological theories give different answers to such questions.

Furthermore, positive and normative doxastic reasons can be quite distinct from one another. We know from epistemology that agents sometimes form their beliefs on the basis of considerations – positive doxastic reasons – that are by no means normative reasons for those beliefs. Someone may believe something because it appeals to him or her in ways that are completely irrelevant from the perspective of a normative theory of epistemology.

To provide a model of a doxastic agent in actual as well as counterfactual circumstances, we must take the agent to have an entire family of credence orders, consisting of one credence order $\succsim_D$ for each possible set of doxastic reasons D. Although at any given time the agent will be in a single doxastic state, the reference to a family of credence orders expresses the fact that the agent may be disposed – or, under the normative interpretation, rationally required – to change his or her beliefs in certain ways when D changes.

To define this family of credence orders, we need to say what the different possible doxastic states of the agent are. Let $\mathcal{D}$ denote the set of all sets of doxastic reasons deemed possible: psychologically possible under a positive interpretation, normatively

---

[2] In our earlier work on reasons for preferences, we discuss three accounts of when a proposition becomes positively relevant (see Dietrich and List 2012a). The first is the 'abstract conceptualization' account, according to which a proposition attains positive relevance as soon as the agent conceptualizes it abstractly. We find this account insufficient. The second account is the 'qualitative understanding' account, according to which a proposition attains positive relevance only if in addition to being conceptualized by the agent it is also understood in a more demanding qualitative sense. We take this account to be more plausible than the first one, at least in some applications. The third account is the 'attentional salience' account, according to which a proposition gains positive relevance if and only if it is sufficiently salient for the agent or the agent explicitly attends to it. Arguably, this account best captures the role deliberation may play in influencing an agent's reasons for his or her attitudes.

admissible under a normative one. Each set D in $\mathcal{D}$ corresponds to one particular context or doxastic state the agent could possibly (or admissibly) be in. In the limit, $\mathcal{D}$ could be the set of all logically possible sets of propositions. However, for our first theorem, we only assume that $\mathcal{D}$ is intersection-closed, i.e.,

$$\text{for any } D_1, D_2 \in \mathcal{D}, \ D_1 \cap D_2 \in \mathcal{D}.$$

For our second theorem, we only assume that D is *either* intersection-closed, as just defined, *or* union-closed, meaning that

$$\text{for any } D_1, D_2 \in \mathcal{D}, \ D_1 \cup D_2 \in \mathcal{D}.$$

In fact, in the last clause, union-closure can even be weakened to the requirement that, for any pair of sets $D_1, D_2 \in \mathcal{D}$, $\mathcal{D}$ contain *some* superset of $D_1 \cup D_2$.

5 A reason-based representation of ordinal beliefs

We can now state our two central axioms on the relationship between an agent's set of doxastic reasons and his or her beliefs. They are the doxastic counterparts of two axioms on the relationship between practical reasons and preferences we have introduced elsewhere (Dietrich and List 2012a,b). We first state these axioms in relation to ordinal beliefs, i.e., beliefs represented by credence orders, and later restate them in the more standard, cardinal case of beliefs represented by credence functions.

Both axioms capture the idea that an agent's beliefs in the possibilities in X are constrained by the propositions that (i) characterize those possibilities and (ii) serve as doxastic reasons for the agent. Depending on whether we interpret our framework positively or normatively, the constraints either describe the agent's actual belief formation or specify how he or she ought rationally to form his or her beliefs.

The first axiom says that the agent has an equal degree of belief in any two possibilities of which the same doxastic reasons are true.

Axiom 1 (*Principle of insufficient reason*) *For any* $x, y \in X$ *and any* $D \in \mathcal{D}$,

$$\{R \in D : R \text{ is true of } x\} = \{R \in D : R \text{ is true of } y\} \implies x \sim_D y.$$

This axiom expresses, and perhaps generalizes, the idea underlying the well-known principle of insufficient reason, which goes back at least to Bernoulli (1654-1705) and Laplace (1749-1827). In its traditional form, the principle states that unless we have sufficient reason to distinguish between two possibilities over and above distinguishing them by name, we should consider them equiprobable. Our present variant formalizes the lack of sufficient reason as the absence of any doxastic reasons that discriminate between the possibilities in question.

The second axiom says that if the agent's set of doxastic reasons grows (say from D to D') but none of the added reasons (in $D'\backslash D$) is true of either of a given pair of possibilities x and y, then the agent's credal ranking of x and y remains unchanged.

Axiom 2 (Invariance of relative likelihoods under the addition of irrelevant reasons) *For any $x, y \in X$ and any $D, D' \in \mathcal{D}$ with $D' \supseteq D$,*

$$\big[\text{no } R \in D'\backslash D \text{ is true of } x \text{ or } y\big] \Longrightarrow \big[x \succsim_D y \Leftrightarrow x \succsim_{D'} y\big].$$

Axiom 2 is plausible since we do not expect an agent to change his or her credal ranking of x and y unless he or she has come to focus on at least one new doxastic reason that is true of x or y and is thereby 'relevant'. Without any such new reason, the agent has no sufficient reason for changing his or her belief in x relative to y.

By reinterpreting a theorem originally proved in the context of reason-based preferences (Dietrich and List 2012a,b), we can show that the two axioms just introduced are necessary and sufficient for a parsimonious representation of the agent's beliefs. The agent's credence orders across all variations in doxastic reasons can then be represented in terms of a single underlying binary relation, denoted $\geq$, which ranks different possible combinations of doxastic reasons in terms of how 'credible' they are. We call this relation a *credibility relation*. A *possible combination of doxastic reasons* – the relatum on both sides of $\geq$ – is a consistent subset of some underlying set $\mathcal{P}$ of all possible doxastic reasons.[3] (As is standard, a set of propositions is *consistent* if there is a possibility x in X of which all the propositions in the set are true.)

Theorem 1 *The agent's family of credence orders $(\succsim_D)_{D \in \mathcal{D}}$ satisfies Axioms 1 and 2 if and only if there exists a credibility relation $\geq$ over all possible combinations of doxastic reasons such that, for each $D \in \mathcal{D}$,*

$$x \succsim_D y \iff \{R \in D : R \text{ is true of } x\} \geq \{R \in D : R \text{ is true of } y\} \text{ for all } x, y \in X.$$

According to this representation, the agent's credibility relation generates the agent's beliefs as follows: in any doxastic state, the agent believes a possibility x more strongly than another possibility y if and only if the credibility relation ranks the combination of doxastic reasons that are true of x above the combination of doxastic reasons that are true of y. The two combinations of doxastic reasons that are being compared (i.e., $\{R \in D : R \text{ is true of } x\}$ and $\{R \in D : R \text{ is true of } y\}$) can be seen as characterizing the possibilities x and y through the lens of the agent's doxastic state.

---

[3] We might define $\mathcal{P}$, for instance, as the set of all those propositions that occur in at least one specification of D in $\mathcal{D}$.

6   Two examples

Some examples help to illustrate this reason-based representation of an agent's beliefs. Suppose an agent seeks to form beliefs regarding four simple possibilities, namely possible states of a particular refrigerator:[4]

> The pump (of the refrigerator) is broken and the ice (inside) is melting (bm).
>
> The pump is not broken but the ice is melting nonetheless ($\neg bm$).
>
> The pump is broken but (surprisingly) the ice is not melting ($b\neg m$).
>
> The pump is not broken and the ice is not melting ($\neg b\neg m$).

Suppose, further, the agent has not yet gathered any evidence about this particular refrigerator but instead seeks to form his or her prior beliefs on the basis of the antecedent credibility of the different possibilities (this antecedent credibility may of course depend on the agent's background beliefs about the world). For simplicity, suppose the only considerations (propositions) that may serve as reasons for the agent's prior beliefs are

$$B \quad : \quad \text{the pump is broken (formally } \{bm, b\neg m\}\text{), and}$$
$$M \quad : \quad \text{the ice is melting (formally } \{bm, \neg bm\}\text{),}$$

and any set of these reasons can be simultaneously active, i.e., $\mathcal{D}$ consists of all subsets of $\mathcal{P} = \{B, M\}$. Now it is conceivable that the agent's credence orders across variations in his or her doxastic reasons are the following:

$$D = \{B, M\} : \quad \neg b\neg m \; \succ_D \; bm \; \succ_D \; \neg bm \; \succ_D \; \neg b\neg m;$$
$$D = \{B\} : \quad \neg b\neg m \; \sim_D \; \neg bm \; \succ_D \; bm \; \sim_D \; b\neg m;$$
$$D = \{M\} : \quad \neg b\neg m \; \sim_D \; b\neg m \; \succ_D \; bm \; \sim_D \; \neg bm;$$
$$D = \varnothing : \quad \neg b\neg m \; \sim_D \; bm \; \sim_D \; \neg bm \; \sim_D \; \neg b\neg m.$$

This family of credence orders is to be understood as follows: if the agent has or focuses on both B and M as his or her doxastic reasons, then the agent's credence order over the four possibilities is the first one; if the agent has or focuses on B alone as his or her doxastic reason, then the agent's credence order is the second one; and so on. It is important to note that this is just one example of what the agent's credence orders in the four different possible doxastic states might look like (note, in particular, that the agent's set of doxastic reasons only *constrains* but does not by itself *determine* the corresponding credence order; it does so only in conjunction with the underlying credibility relation).

One can check that the present family of credence orders satisfies Axioms 1 and 2. According to our theorem, it must then be representable in terms of a single

---

[4] This example is inspired by a 'freezer' example given by Skorupski (1997).

underlying credibility relation. What is this relation? It is easy to see that it must satisfy

$$\varnothing > \{B, M\} > \{M\} > \{B\},$$

where $>$ denotes the strict relation induced by $\geq$. Thus the empty reason combination, representing the 'default' in which the pump is not broken and the ice is not melting, is deemed most credible; the combination $\{B, M\}$, in which the pump is broken and the ice is melting, is deemed second most credible; the combination $\{M\}$, in which the ice is melting without a broken pump as an accompanying reason, is deemed second least credible; and the combination $\{B\}$, in which the pump is broken without being 'accompanied' by melting ice, is deemed least credible. Consistently with our ordinary practices of belief formation, this credibility relation will then be reflected in the agent's (prior) credence order over the four possible states of the refrigerator, capturing the agent's assignment of (prior) probabilities to these states.

A second example is also in line with how we ordinarily form our beliefs in the absence of specific evidence. The example resonates with the economic literature on 'salience' and 'focal points' in rational decision making (for a classic contribution, see Schelling 1960). Suppose one of us, say Christian, has agreed to meet his friend Rohit somewhere in Washington DC at 12 noon tomorrow, but the two of them have not agreed on a specific place, and they have no way to communicate. Moreover, Christian has no further evidence as to where Rohit is likely to expect him; he has never met up with Rohit in Washington before and does not even know how well Rohit knows Washington. Here are some possible meeting points:

> Union Station (u);
>
> Lincoln Memorial (l);
>
> White House (w);
>
> Hilary Clinton's apartment (h).

How will, or alternatively should, Christian form his beliefs on where Rohit might expect him? The following are possible considerations that might serve as doxastic reasons in his belief formation process:

A : The place in question is where one arrives in Washington (formally $\{u\}$).

F : The place in question is world-famous (formally $\{l, w\}$).

R : The place in question has restricted access (formally $\{w, h\}$).

So Union Square is the only relevant place where one arrives in Washington; both the Lincoln Memorial and the White House are world-famous; and both the White House and Hilary Clinton's apartment have restricted access due to security arrangements. Suppose, further, that any set of these reasons constitutes a possible doxastic state. In

particular, Christian's credence orders across variations in his set of doxastic reasons could look like this:

$$
\begin{aligned}
D = \{A, F, R\}: & \quad u \succ_D l \succ_D w \succ_D h; \\
D = \{A, F\}: & \quad u \succ_D l \sim_D w \succ_D h; \\
D = \{A, R\}: & \quad u \succ_D l \succ_D w \sim_D h; \\
D = \{F, R\}: & \quad l \succ_D w \succ_D u \succ_D h; \\
D = \{A\}: & \quad u \succ_D l \sim_D w \sim_D h; \\
D = \{F\}: & \quad l \sim_D w \succ_D u \sim_D h; \\
D = \{R\}: & \quad l \sim_D u \succ_D w \sim_D h; \\
D = \varnothing: & \quad l \sim_D u \sim_D w \sim_D h.
\end{aligned}
$$

Again, this is just one example of what the family of credence orders across variations in the agent's doxastic state might look like. As in the earlier example, it is easy to check that the present family of credence orders satisfies Axioms 1 and 2. So what is the underlying credibility relation? It must satisfy:

$$
\{A\} > \{F\} > \{F, R\} > \varnothing > \{R\}.
$$

Thus the reason combination $\{A\}$ is considered most credible, the combination $\{F\}$ second most credible, the combination $\{F, R\}$ third most credible, the empty combination second least credible, and the combination $\{R\}$ least credible. Accordingly, Christian will believe it to be antecedently much more credible that Rohit will meet him in a place where one arrives in Washington DC than, for instance, in Hilary Clinton's apartment, of which the only salient thing to be said is that it has restricted access. Again, the kind of reasoning captured by this example can be seen as a stylized version of the way doxastic reasons feature in our ordinary belief formation.

It should be clear at this point how our reason-based representation of an agent's ordinal beliefs works. But we should take a closer look at the nature of the agent's credibility relation that is guaranteed by this representation.

7   Is the credibility relation transitive?

While we have assumed that the agent's credence order in any doxastic state is transitive, this does not settle the question of whether this property carries over to the underlying credibility relation. Surprisingly, the assumptions introduced so far are insufficient to ensure a transitive credibility relation. We show this by giving an example structurally equivalent to an example we have given in the context of reasons for preferences (Dietrich and List 2012a).

Suppose a scientist seeks to rank three alternative hypotheses, h_1, h_2 and h_3, in an order of prior probability. For simplicity, let us assume that the hypotheses can

be assessed with respect to three different properties: whether they are

$$P \quad : \quad \text{parsimonious,}$$
$$C \quad : \quad \text{coherent with other theoretical assumptions, and}$$
$$W \quad : \quad \text{explanatorily wide-ranging.}$$

Crucially, each hypothesis exhibits only two of the three virtues: h_1 fails to be explanatorily wide-ranging, h_2 fails to be coherent with other theoretical assumptions, and h_3 fails to be parsimonious, as summarized in the following table:

	P	C	W
h_1	true	true	false
h_2	true	false	true
h_3	false	true	true

Now the scientist's (prior) credence order over the three hypotheses depends on which of the three virtues serve as doxastic reasons in his or her belief formation. It is conceivable, for example, that his or her beliefs across variations in doxastic reasons are as follows:

$$
\begin{aligned}
D = \{P,C,R\} : & \quad h_1 \sim_D h_2 \sim_D h_3, & \\
D = \{P,C\} : & \quad h_1 \succ_D h_2 \succ_D h_3, & (1) \\
D = \{C,W\} : & \quad h_3 \succ_D h_1 \succ_D h_2, & (2) \\
D = \{P,W\} : & \quad h_2 \succ_D h_3 \succ_D h_1, & (3) \\
D = \{P\} : & \quad h_2 \sim_D h_1 \succ_D h_3, & \\
D = \{C\} : & \quad h_3 \sim_D h_1 \succ_D h_2, & \\
D = \{W\} : & \quad h_3 \sim_D h_2 \succ_D h_1, & \\
D = \varnothing : & \quad h_1 \sim_D h_2 \sim_D h_3. &
\end{aligned}
$$

This family of credence orders can be shown to satisfy Axioms 1 and 2. So there must exist an underlying credibility relation that can generate all of these credence orders. Surprisingly, however, this credibility relation fails to be transitive. To see this, consider the rows labelled (1), (2) and (3), which display the scientist's credence orders when his or her set of doxastic reasons is one of $\{P,C\}$, $\{C,W\}$ or $\{P,W\}$. To represent the scientist's beliefs in the doxastic states corresponding to these rows, his or her credibility relation must have the following properties:

$$
\begin{aligned}
\{P,C\} \quad > \quad \{P\} > \{C\}, \\
\{C,W\} \quad > \quad \{C\} > \{W\}, \\
\{P,W\} \quad > \quad \{W\} > \{P\},
\end{aligned}
$$

and so the relation must be intransitive: $\{P\}$ is deemed more credible than $\{C\}$, $\{C\}$ more credible than $\{W\}$, and yet $\{W\}$ more credible than $\{P\}$, a cyclical sequence of binary comparisons.

Interestingly, this intransitivity in the scientist's credibility relation goes along with the scientist's having a perfectly transitive credence order in any given doxastic state: for any set D of doxastic reasons, the corresponding credence order $\succsim_D$ is a perfectly transitive relation, as we have seen. So, if the scientist is always in a 'stable' doxastic state, the intransitivity in the underlying credibility relation will never show up in his or her actual beliefs.

But the intransitivity can come to the surface if different contexts somehow 'activate' different sets of doxastic reasons. Imagine, for example, that the comparison of any two hypotheses makes precisely those doxastic reasons salient that distinguish these hypotheses. So, in a comparison between h_1 and h_2, the scientist's doxastic reasons will be C and W; in a comparison between h_2 and h_3, they will be P and C; and in a comparison between h_1 and h_3, they will be P and W. With these shifts in doxastic reasons across different pairwise comparisons, the scientist will consider h_1 more credible than h_2, h_2 more credible than h_3, and yet h_3 more credible than h_1, despite having perfectly consistent beliefs *relative to any fixed set of doxastic reasons*.

On the other hand, we can also find conditions under which an intransitive credibility relation is ruled out, so that the patterns of belief just discussed can never occur. In particular, the identified intransitivity could never occur if there existed additional hypotheses h_4 to h_7 of which precisely one or none of the three possible doxastic reasons, P, C and W, were true. Then the scientist's credence order over the extended set of hypotheses h_1 to h_7 in the doxastic state $D = \{P, C, W\}$ would constrain his or her credibility relation to rank all possible combinations of reasons, including the three singletons $\{P\}$, $\{C\}$ and $\{W\}$, transitively.

As we have observed elsewhere (Dietrich and List 2012a), this counterfactual stipulation would give the scientist a kind of 'Olympian perspective' from which he or she could (i) identify one possibility in X corresponding to each possible combination of doxastic reasons (which instantiates all and only the reasons in it) and (ii) by ranking these possibilities, rank all the corresponding reason combinations transitively.

A more technical result generalizes this observation. If the set of possible sets of doxastic reasons $\mathcal{D}$ is union-closed and the underlying set $\mathcal{P}$ of possible doxastic reasons is *weakly independent* – meaning that, for every consistent subset $S \subseteq \mathcal{P}$ there exists some possibility $x \in X$ of which, among the possible reasons in $\mathcal{P}$, all and only those reasons in S are true – then the credibility relation will always be transitive; i.e., it will be a credibility *order* – not just some *relation* – over all possible combinations of doxastic reasons (for a proof, see Dietrich and List 2012a).

8 A reason-based representation of cardinal beliefs

We can now go beyond the representation of beliefs in terms of credence orders and consider the more standard representation in terms of credence functions (subjective probability functions). The agent's beliefs are thus represented by a credence function C from the set of possibilities X into the interval from 0 and 1, with a sum-total of 1.[5] The credence of any proposition (a subset of X) is defined in the usual way by adding up the credences of all the possibilities of which the proposition is true.

To indicate that the agent's beliefs depend once again on his or her doxastic reasons, we append the subscript D to C, interpreting C_D as the agent's (actual or ideally rational) credence function when D is his or her set of doxastic reasons in relation to the possibilities in X. We are interested in how C_D depends on D, quantifying, as before, over all possible sets of doxastic reasons in $\mathcal{D}$. So the object under investigation is now a family of credence functions $(C_D)_{D \in \mathcal{D}}$ rather than a family of credence orders $(\succsim_D)_{D \in \mathcal{D}}$.

The cardinal variants of our two axioms on the relationship between the agent's doxastic reasons and his or her beliefs are directly analogous to the earlier ordinal variants. Again, the constraints given by the axioms can be interpreted either as characterizing the agent's actual belief formation or as specifying how he or she ought rationally to form his or her beliefs. Apart from substituting a credence function for a credence order, the first axiom is identical to the earlier one, requiring the agent to have an equal degree of belief in any two possibilities of which the same doxastic reasons are true.

Axiom 1* (Principle of insufficient reason) *For any $x, y \in X$ and any $D \in \mathcal{D}$,*

$$\{R \in D : R \text{ is true of } x\} = \{R \in D : R \text{ is true of } y\} \Longrightarrow C_D(x) = C_D(y).$$

The second axiom is also essentially the same as its earlier counterpart, except that it is in one respect more specific. Recall that the earlier second axiom required that if the agent's set of doxastic reasons grows from D to D' but none of the added reasons in $D' \backslash D$ is true of either of a given pair of possibilities x and y, then the agent's subjective likelihood of x relative to y remains unchanged. In the ordinal case, this simply meant that $x \succsim_D y \Leftrightarrow x \succsim_{D'} y$. In the cardinal case, we have to be more specific in saying what it means for the relative likelihood of x and y to remain unchanged for the agent. Here we require the agent's subjective *likelihood ratio* of x and y to remain unchanged, i.e., we replace the clause $x \succsim_D y \Leftrightarrow x \succsim_{D'} y$ with $\frac{C_D(x)}{C_D(y)} = \frac{C_{D'}(x)}{C_{D'}(y)}$. Thus the second axiom becomes the following:

---

[5] To avoid technicalities, we assume that X is countable and that each x in X is assigned non-zero probability.

Axiom 2* (Invariance of relative likelihoods under the addition of irrelevant reasons) *For any $x, y \in X$ and any $D, D' \in \mathcal{D}$ with $D' \supseteq D$,*

$$\left[\text{no } R \in D' \backslash D \text{ is true of } x \text{ or } y \right] \implies \frac{C_D(x)}{C_D(y)} = \frac{C_{D'}(x)}{C_{D'}(y)} \text{ for any } x \in X.$$

What is the consequence of Axioms 1* and 2*? It should be evident that an underlying credibility *order* is not sufficient to encode the information needed to represent a family of (cardinal) credence functions $(C_D)_{D \in \mathcal{D}}$. Instead, the agent's beliefs are now representable in terms of an underlying real-valued *credibility function* (as distinct from credence function), denoted γ, which encodes cardinal credibility assessments over possible combinations of doxastic reasons.

Theorem 2 *The agent's family of credence functions $(C_D)_{D \in \mathcal{D}}$ satisfies Axioms 1* and 2* if and only if there exists a real-valued credibility function γ over all possible combinations of doxastic reasons such that, for each $D \in \mathcal{D}$,*

$$C_D(x) = \frac{\gamma(\{R \in E : R \text{ is true of } x\})}{\sum\limits_{x' \in X} \gamma(\{R \in E : R \text{ is true of } x'\})}.$$

According to this cardinal representation, the agent's credibility function generates the agent's beliefs as follows: in any doxastic state, the degree of belief the agent assigns to any possibility is proportional to the credibility (according to γ) of the set of doxastic reasons that are true of that possibility. The factor of proportionality, $1/ \sum\limits_{x' \in X} \gamma(\{R \in E : R \text{ is true of } x'\})$, depends on the doxastic state D and ensures that the agent's degrees of belief across all possibilities x in X add up to 1.

Restricted to the relevant combinations of doxastic reasons – those needed to generate the agent's beliefs – the credibility function γ is unique up to a positive multiplicative constant. The function is fully unique (on the relevant domain) if we introduce the convention that γ adds up to 1, meaning that the credibility function has the formal properties of a probability function over the set of all possible combinations of doxastic reasons.

9 A possible objection

Before concluding with some more general reflections, we need to address one possible objection. Can the kind of reason-based belief formation we have discussed not simply be remodelled in ordinary Bayesian terms? In other words, could it not simply be the case that the agent forms his or her beliefs in doxastic state D by ordinary Bayesian updating instead of some reason-based process? He or she would start with uniform Bayesian priors across the possibilities in X, take into account some relevant information, which is functionally equivalent to our set of doxastic reasons D, and

perform a Bayesian update that results in a posterior credence (subjective probability) function identical to C_D.

The reply is simple. Recall the example of meeting a friend in Washington DC. Our theory allows the agent, for instance, to arrive at a credence function of the form:

$$
\begin{aligned}
C_D(u) &= 0.49; \\
C_D(l) &= 0.3; \\
C_D(w) &= 0.2; \\
C_D(n) &= 0.01,
\end{aligned}
$$

where u, l, w, and n are the four possible meeeting points, as before. These beliefs make perfect sense when the agent's set of doxastic reasons includes all three considerations we have mentioned: where one arrives in Washington, whether the place in question is famous, and whether there is restricted access. But could one possibly form these beliefs through Bayesian updating, starting from uniform priors? The answer is negative. After learning the truth of some proposition – the input of a Bayesian update – the agent's degree of belief in some possibilities, namely those of which the proposition is false, would have to be set to zero, and all other possibilities would still have to be assigned credences proportional to the original priors; this is a simple consequence of Bayes's theorem. So it is never possible to assign posterior degrees of belief 0.49, 0.3, 0.2 and 0.01 to four different basic possibilities (elements of X) if the priors were 0.25, 0.25, 0.25 and 0.25.

There are only two ways in which we might arrive at these non-uniform posteriors via ordinary conditionalization. The first is to apply Jeffrey conditionalization as opposed to standard Bayesian conditionalization (see, e.g., Jeffrey 2004). In Jeffrey conditionalization, the evidence or information learnt need not take the form of a proposition, but can be a probability distribution. Thus the agent may begin with uniform priors and then simply learn the required non-uniform probability distribution across the four different meeting points in Washington. In this case, the evidential or informational input would be the probability distribution itself. It is unclear, however, whether this qualifies as an illuminating explanation of the agent's belief formation or just as a redescription of the facts that were to be explained in the first place.

The second way to remodel the agent's belief formation using ordinary Bayesian conditionalization is to deny that the set X correctly captures the underlying set of possibilities, and to refine the space of possibilities, by 'splitting' the elements of X into more fine-grained possibilities. Given the resulting more fine-grained space of possibilities, it may then be possible for the agent to arrive at non-uniform credences across different possible meeting points (such as 0.49, 0.3, 0.2 and 0.01) via ordinary Bayesian updating from uniform priors (such as 0.25, 0.25, 0.25, 0.25). The individual meeting points (Union Station, Lincoln Memorial etc.) would no longer be the most

basic possibilities, but they would correspond to different non-singleton subsets of a refinement of the original set X. The input of a Bayesian update – namely a proposition picking out a subset of that refined set of possibilities – might then rule out more possibilities underlying some meeting points (e.g., more finely individuated possibilities that each correspond to meeting at Hilary Clinton's apartment) than possibilities underlying others (e.g., more finely individuated possibilities that each correspond to meeting at Union Station). While this move is theoretically possible and perhaps even plausible in some cases, it may require us to sacrifice theoretical parsimony by ascribing to the agent a very complex ontology of epistemic possibilities. (This inflation in the ascribed ontology of possibilities would be particularly significant in the case of repeated belief changes of the sort discussed here.) Addressing this issue in detail is beyond the scope of this paper, but we wish to acknowledge it as something that merits further attention.[6]

10   Concluding remarks

Just as standard rational choice theory does not say where an agent's preferences come from (see Dietrich and List 2012b), so Bayesian epistemology does not give a satisfactory account of where the agent's prior beliefs come from. We have tried to sketch such an account, introducing a reason-based model of belief formation.

Our account provides many new conceptual resources, which we can only hint at rather than fully develop here. Since our account can capture both the relationship between positive doxastic reasons and an agent's actual beliefs and the relationship between normative doxastic reasons and the beliefs an agent ought rationally to have, we are in a position to formalize notions such as:

- *beliefs that are explained by positive doxastic reasons, but not justified by normative doxastic reasons*: here, the agent's positive doxastic reasons employed in his or her belief formation are not genuine normative doxastic reasons, according to our normative background theory of epistemology;
- *beliefs that are justifiable, but not actually justified, by the right reasons*: here, there exist some normative doxastic reasons that would justify the agent's beliefs, given a suitable credibility relation or function, but these are not the positive doxastic reasons actually employed by the agent in his or her belief formation;
- *beliefs that are formed on the basis of the right reasons versus the wrong ones*: here, the agent's positive doxastic reasons are genuine normative doxastic reasons (which still leaves open whether the agent's credibility relation or function is a normatively appropriate one);

---

[6]In ongoing, unpublished work, we argue against the attempt to remodel reason-based attitude formation in purely informational terms.

- *beliefs that are formed on the basis of the right credibility relation or function versus the wrong one*: here, the agent's credibility relation or function is a normatively appropriate one (which this still leaves open whether the agent's positive doxastic reasons are genuine normative ones);
- *beliefs that are formed on the basis of the right reasons and the right credibility relation or function*: here, the agent's actual belief formation coincides with the normatively ideal one – the epistemically impeccable case.

Similarly, our account points towards a novel taxonomy of different sources of disagreement between two or more agents' beliefs. Differences in belief may stem from:

- different doxastic reasons *and* different credibility relations or functions;
- different doxastic reasons but the same credibility relation or function;
- the same doxastic reasons but different credibility relations or functions;
- and, finally, differences in evidence or information in the ordinary Bayesian sense.

By providing these conceptual resources, our approach allows us to capture, among other things, the role of reasons in one or several agents' deliberation about their beliefs, over and above the familiar pooling of evidence or information. This is relevant, for instance, to debates on Aumann's agreement theorem. Robert Aumann (1976) famously proved that if different agents have the same prior beliefs and common knowledge of one another's rationality, they cannot agree to disagree. Our account suggests that even if different agents have the same credibility function and common knowledge of one another's rationality, they may still agree to disagree, namely when they have different doxastic reasons. This observation, in turn, raises some important questions about how an agent's doxastic reasons would be affected by various kinds of intra- and interpersonal deliberation.

Of course, the larger research programme of which the present paper is part continues to be work in progress, but we hope to have illustrated the usefulness of a reason-based approach to Bayesian epistemology.

References

[1] Aumann, Robert (1976) 'Agreeing to disagree', *The Annals of Statistics* 4(6): 1236-1239.

[2] Bovens, Luc, and Stephan Hartmann (2003) *Bayesian Epistemology*, Oxford (Oxford University Press).

[3] Dietrich, Franz, and Christian List (2012a) 'A Reason-Based Theory of Rational Choice', *Nous*, in press.

[4] Dietrich, Franz, and Christian List (2012b) 'Where do preferences come from', *International Journal of Game Theory*, forthcoming.

[5] Foley, Richard (1991) 'Evidence and Reasons for Belief', *Analysis* 51(2): 98-102.

[6] Jeffrey, Richard (2004) *Subjective Probability: The Real Thing*, Cambridge (Cambridge University Press).

[7] Liu, Fenrong (2010) 'Von Wright's "The Logic of Preference" revisited', *Synthese* 175(1): 69-88.

[8] Osherson, Daniel, and Scott Weinstein (2012) 'Preferences based on reasons', *Review of Symbolic Logic*, in press.

[9] Scanlon, Thomas (1998) *What we owe to each other*, Cambridge, MA (Harvard University Press).

[10] Schelling, Thomas (1960) *The Strategy of Conflict*, Cambridge, MA (Harvard University Press).

[11] Skorupski, John (1997) 'Reasons and Reason', in Garrett Cullity and Berys Gaut (eds.), *Ethics and Practical Reason*, Oxford (Oxford University Press), pp. 345-367.

[12] von Wright, Georg Henrik (1963) *The Logic of Preference*, Edinburgh (Edinburgh University Press).