The Stability Theory of Belief

Hannes Leitgeb
Ludwig Maximilians University

This essay develops a joint theory of rational (all-or-nothing) belief and degrees of belief. The theory is based on three assumptions: the logical closure of rational belief; the axioms of probability for rational degrees of belief; and the so-called Lockean thesis, in which the concepts of rational belief and rational degree of belief figure simultaneously. In spite of what is commonly believed, I will show that this combination of principles is satisfiable (and indeed nontrivially so) and that the principles are jointly satisfied if and only if rational belief is equivalent to the assignment of a stably high rational degree of belief. Although the logical closure of belief and the Lockean thesis are attractive postulates in themselves, initially this may seem like a formal "curiosity"; however, as I am going to argue in the rest of the essay, a very reasonable theory of rational belief can be built around these principles that is not ad hoc but that has various philosophical features that are plausible independently.

I would like to thank Timothy Williamson, Philip Welch, Richard Pettigrew, Gerhard Schurz, David Makinson, David Etlin, Horacio Arló-Costa, Philip Ebert, Martin Smith, Ole Hjortland, Florian Steinberger, Franz Huber, Christian Feldbacher, and two anonymous referees and the editors of this journal for invaluable comments on earlier drafts of this essay. I would also like to thank Kevin Kelly, Hanti Lin, Sonja Smets, and Alexandru Baltag for their comments on a conference version of this essay (the comments, together with an extended abstract of this essay, have appeared in the proceedings volume of the Tsinghua Logic Conference, Beijing 2013; see Leitgeb 2013b). I am grateful to the Alexander von Humboldt Foundation for financially supporting this work.

Philosophical Review, Vol. 123, No. 2, 2014
DOI 10.1215/00318108-2400575

Published by Duke University Press

1. The Lockean Thesis and Closure of Belief under Conjunction

Each of the following three postulates on belief (*Bel*) and degrees of belief (P) for perfectly rational agents seems tempting, at least if taken just by itself:

P1 The logic of belief, in particular, the logical closure of belief under conjunction, that is: for all propositions A, B,

$$\text{if } Bel(A) \text{ and } Bel(B) \text{ then } Bel(A \wedge B).$$

P2 The axioms of probability for the degree of belief function P.

P3 The Lockean thesis (see Foley 1993, 140–41) that governs both *Bel* and P: there is a threshold r that is greater than $\frac{1}{2}$ and at most 1, such that for every proposition B, it holds that B is believed if and only if the degree of belief in B is not less than r, or more briefly,

$$Bel(B) \text{ if and only if } P(B) \geq r.$$[1]

P1 is entailed by the doxastic version of any normal system of modal logic for the operator '*Bel*', P2 is at the heart of Bayesianism, and P3 expresses the natural thought that it is rational to believe a proposition if and only if it is rational to have a sufficiently high degree of belief in it.

Yet this combination of rationality postulates is commonly rejected. And the standard reason for doing so is that, given P1–P2, there does not seem to be any plausible value of 'r' available that would justify the existence claim in P3.

Here is why: The first possible option, r being equal to 1, seems much too extreme; '$Bel(B)$ if and only if $P(B) \geq r$' would turn into the trivializing '$Bel(B)$ if and only if $P(B) = 1$' condition, by which all and only propositions of which one is probabilistically certain are to be believed. But this cannot be right, at least if it is taken as a requirement on believed propositions that is meant to hold in each and every context. For example: it is morning; I rationally believe that I am going to receive an e-mail today. However, I would not regard it as rational to buy a bet in which I would win one dollar if I am right, and in which I would lose a million dollars if I am wrong. But according to the usual interpretation of

[1] In many formulations of the Lockean thesis, a greater-than symbol is used instead of a greater-than-or-equal-to symbol, but since I am going to assume the underlying set of possible worlds to be finite, nothing will really hang on this choice of formulation. However, the "greater-than-equals" formulation will prove to be more convenient for my own purposes.

subjective probabilities in terms of betting quotients, I should be rationally disposed to accept such a bet if I believed the relevant proposition to the maximal degree of 1. Hence, I rationally believe the proposition even though I do not believe it with probability 1.

The remaining way to argue for the existence claim in P3 would be to turn to some value of 'r' that is less than 1; and as long as one considers '$Bel(B)$ if and only if $P(B) \geq r$' just by itself, this looks far more appealing and realistic. But then again, if taken together with P1 and P2, this option seems to run into the famous Lottery Paradox (see Kyburg 1961), to which I will return later.[2]

Therefore, in spite of the prima facie attractiveness of each of P1–P3, it just does not seem to be feasible to have all of them at the same time, which is why a large part of the classical literature on belief (or acceptance) can be categorized according to which of the three postulates are being preserved and which are dropped—as Levi (1967, 41) formulates it, "either cogency [my P1] or the requirement of high probability as necessary and sufficient for acceptance [my P3] must be abandoned." For instance, putting P2 to one side for now, Isaac Levi keeps P1 but rejects P3, while Henry Kyburg keeps P3 and rejects P1. Hempel (1962) still had included both P1 and P3 as plausible desiderata, although he was already aware of the tension between them.

In the following, I want to show that this reaction of dropping any one of P1–P3 is premature; it is in fact not clear that one could not have all of P1–P3 at once and the existence claim in P3 being true in virtue of some threshold $r < 1$.

The first step to seeing this is to note that P3, as formulated above, is *ambiguous* with respect to the position of the 'there is a threshold r' quantifier in relation to the implicit universal quantification over degree of belief functions P. According to one possible disambiguation, there is indeed no value of 'r' less than 1 so that 'for all B, $Bel(B)$ if and only if $P(B) \geq r$' could be combined consistently with P1 and P2. But according to a second kind of disambiguation, taking all of these assumptions together will in fact be logically possible, and it will be this manner of understanding P3 on which my stability theory of belief will be based.

Here is the essential point: we need to distinguish a claim of the form 'there is an $r < 1$ ... for all P...' from one of the form 'for all P... there is an $r < 1$ ...'. As we are going to see, the difference is crucial:

[2] A similar point can be made in terms of the equally well-known Preface Paradox; see Makinson 1965.

while it is *not* the case that

> there is an $r < 1$, such that for all P (on a finite space of worlds),

the logical closure of *Bel*, the probability axioms for *P*, and for all *B*, *Bel*(*B*) if and only if $P(B) \geq r$, are jointly satisfied, it *is* the case that

> for all P (on a finite space of worlds), there is an $r < 1$

such that the same conditions are jointly the case.

Let me explain why. I will start with what will be interpreted later on in sections 3 and 4 as a typical lottery example:

> **Example 1** (1a) Assume that $r = \frac{999{,}999}{1{,}000{,}001}$. Consider W to be a set $\{w_1, \ldots, w_{1{,}000{,}000}\}$ of one million possible worlds, and let P be the uniquely determined probability measure that is given by $P(\{w_1\}) = P(\{w_2\}) = \ldots = P(\{w_{1{,}000{,}000}\}) = \frac{1}{1{,}000{,}000}$. A fortiori, the axioms of probability are satisfied by *P*, as demanded by P2 above. At the same time, by the Lockean thesis (P3), it would follow that for every $1 \leq i \leq 1{,}000{,}000$, it is rational to believe the proposition $W - \{w_i\}$ (that is, W without $\{w_i\}$), as $P(W - \{w_i\}) = \frac{999{,}999}{1{,}000{,}000} \geq \frac{999{,}999}{1{,}000{,}001}$. Therefore, by P1, the conjunction (that is, intersection) of all of these propositions would rationally have to be believed as well; but this conjunction is nothing but the contradictory proposition $\varnothing$, which has probability 0 by P2, and which for that reason is not rationally believed according to P3. We end up with a contradiction. For r as being chosen before, we can determine a probability measure P, such that the logical closure of *Bel*, the probability axioms for P, and for all B, *Bel*(B) if and only if $P(B) \geq r$, do not hold jointly. By the same token, for every $\frac{1}{2} < r < 1$, a uniform probability measure can be constructed, such that these conditions are not satisfied simultaneously.
>
> (1b) Let W be the set $\{w_1, \ldots, w_{1{,}000{,}000}\}$ again, and assume the probability measure P to be again given by $P(\{w_1\}) = \ldots = P(\{w_{1{,}000{,}000}\}) = \frac{1}{1{,}000{,}000}$. But now set $r = \frac{1{,}000{,}000}{1{,}000{,}001}$: in that case, the only proposition that is to be believed according to the Lockean thesis is W itself, which has probability 1. Trivially, then, the set of believed propositions is closed under logic (including closure under conjunction), which is why the logical closure of *Bel*, the probability axioms for P, and for all B, *Bel*(B) if and only if $P(B) \geq r$, hold jointly. For P as being chosen before, we can determine a threshold r, such that all of our desiderata are satisfied.

It is evident that in case 1b, we were able to circumvent the contradiction from 1a by another trivializing method (just as opting for $r = 1$ and '*Bel*(B) if and only if $P(B) = 1$' had been trivializing before): given a P with a finite domain, we can push the threshold in the Lockean thesis sufficiently close to (though short of) 1 so that only those propositions that have probability 1 end up believed.

While the same method enables us to determine for every probability measure (over a finite set of worlds) a suitable threshold $r < 1$ and *Bel* such that P1, P2, and for all B, *Bel*(B) if and only if $P(B) \geq r$, are jointly the case, this is hardly satisfying; for once again, rational belief would be restricted to propositions of which one is probabilistically certain. The much more exciting observation is that in many cases one can do much better: it is possible to achieve the same result without trivializing consequences, in the sense that at least some proposition of probability *less than 1* happens to be believed.

Here is an example (to which we will return also in subsequent sections and which will be given a concrete interpretation in section 5):

Example 2 Let $W = \{w_1, \ldots, w_8\}$ be a set of eight possible worlds; one might think of these eight possibilities as coinciding with the state descriptions that can be built from three propositions A, B, C: w_1 corresponds to $A \wedge B \wedge \neg C$, w_2 to $A \wedge \neg B \wedge \neg C$, w_3 to $\neg A \wedge B \wedge \neg C$, w_4 to $\neg A \wedge \neg B \wedge \neg C$, w_5 to $A \wedge \neg B \wedge C$, w_6 to $\neg A \wedge \neg B \wedge C$, w_7 to $\neg A \wedge B \wedge C$, and w_8 to $A \wedge B \wedge C$. Let P be the unique probability measure that is defined by: $P(\{w_1\}) = 0.54$, $P(\{w_2\}) = 0.342$, $P(\{w_3\}) = 0.058$, $P(\{w_4\}) = 0.03994$, $P(\{w_5\}) = 0.018$, $P(\{w_6\}) = 0.002$, $P(\{w_7\}) = 0.00006$, $P(\{w_8\}) = 0$. Figure 1 depicts what this probability space looks like.

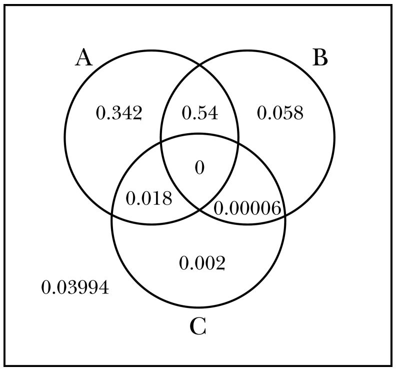

Figure 1. Example 2

Now consider the following six propositions,

$$\{w_1\}, \{w_1, w_2\}, \{w_1, \ldots, w_4\}, \{w_1, \ldots, w_5\}, \{w_1, \ldots, w_6\}, \{w_1, \ldots, w_7\}$$

only the last one of which has probability 1. Pick any of them, call it 'B_W', and let *Bel* be determined uniquely by stipulating that B_W is the least or strongest proposition that is believed, so that a proposition is believed if

> and only if it is entailed by (is a superset of) B_W; in other words: for all propositions $X \subseteq W$,
>
> $$Bel(X) \text{ if and only if } B_W \subseteq X.$$
>
> Finally, take $r = P(B_W)$ to be the relevant threshold. One can show that the so-determined *Bel*, *P*, and *r* satisfy the logical closure of *Bel*, the probability axioms for *P*, and for all *B*, *Bel*(*B*) if and only if $P(B) \geq r$. Once again, for our given *P*, there is a threshold *r*, such that all of our desiderata hold simultaneously. But this time, as far as the first five choices of B_W are concerned, there is in fact a proposition of probability less than 1 that is being believed. For example, if B_W is $\{w_1, w_2\}$, then $\{w_1, w_2\}$ is believed even though it has a probability of $0.882 < 1$.

What should we conclude from these examples? *Maybe it is possible to have one's cake and eat it, too*: to preserve the logic of belief and the axioms of probability while at the same time assuming consistently that the beliefs and degrees of belief of perfectly rational agents relate to each other as expressed by the Lockean thesis even for a threshold of less than 1.

The price to be paid for this proposal will be that not any old threshold in the Lockean thesis will do; instead the threshold must be chosen suitably depending on what the agent's beliefs and his or her degree of belief function are like. Whether this price is affordable or not, I will discuss later, but first I will turn to a different question: given a degree of belief function *P*, what are the belief sets *Bel* and thresholds *r* like that, together with *P*, satisfy all of our intended conditions? The answer will be given in section 2, in which P1–P3 will be made formally precise, and in which the intended belief sets and thresholds will be characterized in terms of a probabilistic notion of stability or resiliency. Based on this, I will formulate what will be called the "stability theory of belief" in section 3, which will postulate that belief corresponds to resiliently high probability, which is going to entail P1–P3; afterward, in the same section, I will outline the costs of accepting this theory: a strong form of context sensitivity of belief, where the context in question involves both the agent's degree of belief function *P* and the partitioning or individuation of the underlying possibilities. Section 4 explains what the theory predicts concerning the Lottery Paradox; the observed context sensitivity of belief will actually work to the theory's advantage there. In section 5, I will present an example of how the theory can be applied in other areas, in this case, to a problem in formal epistemology. Section 6 summarizes what has been achieved and, on these grounds, makes the case for the theory.

2. *P*-Stability

I begin by stating P1–P3 in full detail.

Let us consider a perfectly rational cognitive agent and his or her beliefs and degrees of belief at a fixed point of time. By ‘perfectly rational’, I mean only ‘*inferentially* perfectly rational’—so that the usual logical principles of doxastic closure and the principles of probability can be taken for granted for any such agent—but of course I do not assume, for example, that any such agent would be perfectly rational in the sense of believing all and only truths, or the like.[3]

Let W be a (nonempty) set of possible worlds. Throughout the essay, I will assume that W is *finite*; the theory that I am going to develop will work also in the infinite case, but I want to keep things as simple as possible here.

Given W, by a proposition, I mean any subset of W; so propositions will be regarded as sets of possible worlds. I will apply the standard terminology that is normally used for sentences also to propositions: when I say that a proposition is consistent, I mean that it is nonempty, and accordingly $\varnothing$ is the unique contradictory proposition; when I say that a proposition A is consistent with another proposition B, then this is: $A \cap B \neq \varnothing$; when I say that A entails B, this amounts to A being a subset of B; when I refer to the negation of A, I actually refer to its complement $W - A$ relative to W (which I will also denote by ‘$\neg A$’); the conjunction $A \wedge B$ of A and B is their intersection; and their disjunction $A \vee B$ is their union.

I represent the agent’s beliefs at the relevant time by means of a set *Bel* of propositions: the set of propositions believed by the agent in question at the time in question. Instead of ‘$A \in Bel$’ I will usually write $Bel(A)$.

This being in place, P1 was really a shorthand for the standard laws of doxastic logic adapted to the current propositional context (and disregarding introspection, which will not play any role in this essay):

[3] Ultimately, we should be concerned with *real-world* agents, but methodologically it seems like a good strategy to sort out the tension between belief and degrees of belief first for ideal agents—whom we strive to approximate—and only then for agents such as ourselves.

P1 For all propositions $A, B \subseteq W$:

- $Bel(W)$;
- not $Bel(\varnothing)$;
- if $Bel(A)$ and $A \subseteq B$, then $Bel(B)$;
- if $Bel(A)$ and $Bel(B)$, then $Bel(A \wedge B)$.

The first two clauses express that the agent believes that one of the worlds within his or her total set W of worlds is the actual world, and he or she does not believe the empty set to include the actual world. The other two clauses express the closure of belief under logical consequence.

Since W is finite by assumption, there can be only finitely many members of *Bel*; by P1, the conjunction of all of them, say, B_W, must also be a member of *Bel*, B_W must be consistent, and by the definition of B_W and by P1 again, the following must hold for every proposition B: $Bel(B)$ if and only if $B_W \subseteq B$.

Vice versa, assume there to be a consistent proposition B_W in *Bel*, such that for every proposition B: $Bel(B)$ if and only if $B_W \subseteq B$. Then it follows that P1 above is satisfied.

In other words, we can reformulate P1 equivalently as follows:

P1 [Reformulated] There is a consistent proposition $B_W \subseteq W$, such that for all propositions B:

- $Bel(B)$ if and only if $B_W \subseteq B$.

So P1 really amounts to a possible worlds model of belief: the agent implicitly or explicitly divides the set W of possible worlds into those that are serious possibilities for the agent at the time (see Levi 1984)—that is, serious candidates for what the actual world might be like—and those which are not. B_W is that set of serious possibilities, and it is determined uniquely given the belief set *Bel*.

Now I turn to P2: At the relevant point of time, let P be the agent's degree of belief or credence function, which I take to be defined for all subsets of W; in probabilistic terms, W is the sample space for P. Indeed, P2 assumes that P is a probability measure, and accordingly it states that:

P2 For all propositions $A, B \subseteq W$:

- $P(W) = 1$;
- if A is inconsistent with B, then $P(A \vee B) = P(A) + P(B)$;

- additionally, conditional degrees of belief can be introduced by

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

whenever $P(A) > 0$.

Since W was assumed to be finite, we may think of probabilities this way: they are assigned first to the singleton subsets of W—or, if one prefers, to the worlds in W themselves—and then the probabilities of larger sets are determined by adding up the probabilities of its singleton subsets. Because W is finite, we do not need to deal at all with the probabilities of infinite unions or intersections of propositions.

Finally, the Lockean thesis:

P3 There is an r with $\frac{1}{2} < r \leq 1$, such that for all propositions $B \subseteq W$:

$$Bel(B) \text{ if and only if } P(B) \geq r.$$

Now drop the existential quantifier 'there is an r' for a moment so that 'r' becomes a free variable and call the resulting open formula 'P3[r]' (read this as 'the Lockean thesis with threshold r'): if the interpretations of 'P' and '*Bel*' are fixed, then, depending on the value of 'r', P3[r] might turn out to be either true or false; and I will be interested in characterizing those values of 'r' for which it is true. I do allow for $r = 1$, but I will be particularly interested in choosing r so that propositions of probability less than 1 will also be believed by the agent. For the moment, I will focus especially on the right-to-left direction of the Lockean thesis (LT) with threshold r:

$$\mathrm{LT}^{\geq r > \frac{1}{2}}_{\leftarrow}: \text{ For all } B, Bel(B) \text{ if } P(B) \geq r.$$

This is because, with the right background assumptions, $\mathrm{LT}^{\geq r > \frac{1}{2}}_{\leftarrow}$ will actually turn out to be equivalent to P3[r], which is interesting in itself to observe. Other than that, in what follows, I could have worked just with P3[r] directly.

Now we are almost ready to spell out under what conditions P1, P2, and $\mathrm{LT}^{\geq r > \frac{1}{2}}_{\leftarrow}$ (or P3[r]) are jointly satisfied. In order to formulate the corresponding theorem, we will need one final probabilistic concept that is closely related, though not identical, to the notions of resiliency introduced by Skyrms (1977, 1980) within his theory of objective chance:

Definition 1 With P being a probability measure on the sample space W, we define for all $A \subseteq W$:

A is P-stable if and only if for all $B \subseteq W$, such that B is consistent with A and $P(B) > 0$:

$$P(A \mid B) > \frac{1}{2}.$$

Thus, a proposition is stable just in case it is sufficiently probable given any proposition with which it is compatible.

In order to get a feel for this definition, consider a consistent (nonempty) proposition A that is P-stable: one of the suitable values of 'B' above is the total set W of worlds—as W is consistent with A, and $P(W) = 1$—which is why P-stability entails that $P(A \mid W) = P(A) > \frac{1}{2}$; therefore, any consistent P-stable proposition A must have a probability greater than that of its negation. What *P-stability* adds to this is that *this is going to remain so* under the supposition of any proposition B that is consistent with A and for which conditional probabilities are defined—the high probability of A is resilient or robust.

It follows immediately from the axioms of probability that every proposition of probability 1 must be P-stable. For trivial reasons also, the empty proposition is P-stable. And it might seem that this will actually exhaust the class of P-stable sets since P-stability might seem pretty restrictive; but things will turn out to be quite different.

The relevance of P-stability is made transparent by the following representation theorem (I omit its proof, but it is not difficult at all):[4]

Theorem 1 Let W be a finite nonempty set, let *Bel* be a set of subsets of W, and let P assign to each subset of W a number in the interval $[0, 1]$. Then the following two statements are equivalent:

I. *Bel* satisfies P1, P satisfies P2, and P and *Bel* satisfy $\mathrm{LT}_{\leftarrow}^{\geq P(B_W) > \frac{1}{2}}$.

II. P satisfies P2, and there is a (uniquely determined) $A \subseteq W$, such that
- A is a nonempty P-stable proposition,
- if $P(A) = 1$, then A is the least subset of W with probability 1; and
- for all $B \subseteq W$:

$$\mathit{Bel}(B) \text{ if and only if } A \subseteq B$$

(and hence, $B_W = A$).

[4] See Leitgeb 2013a, n. 26 for the proof ('P-stable' in the present essay corresponds to 'P-stable$^{\frac{1}{2}}$' in Leitgeb 2013a).

This is a (universally quantified) equivalence statement: its left-hand side (I) summarizes all of our desiderata if for the moment we restrict ourselves just to one direction of the Lockean thesis and if we use $P(B_W)$ as the corresponding threshold; and the right hand-side (II) expresses that B_W is P-stable and if B_W has probability 1, then it is the least proposition of probability 1 (which must always exist for finite W).

Summing up: If P and *Bel* are such that P1, P2, and the right-to-left direction of the Lockean thesis with threshold $P(B_W)$ are satisfied, where B_W is the least believed proposition that exists by P1, then B_W must be P-stable. And if given P and a P-stable proposition (which, if it has probability 1, is the least of that kind), then one can determine *Bel* from that P-stable proposition, so that P and *Bel* satisfy all of the desiderata, and the given P-stable proposition is the strongest believed proposition B_W. Or once again, in other terms: assuming that P and *Bel* make condition (I) from above true carries exactly the same information as assuming that P is a probability measure and the least believed proposition is P-stable (and, if it has probability 1, is the least proposition of probability 1).

One can show even more: Either side of the equivalence statement that is embedded in the theorem above actually implies the *full* Lockean thesis with threshold $P(B_W)$, that is, for all propositions B: $Bel(B)$ if *and only if* $P(B) \geq P(B_W) > \frac{1}{2}$. Consequently, one can replace '$\mathrm{LT}^{\geq P(B_W) > \frac{1}{2}}_{\leftarrow}$' in condition (I) by P3[$P(B_W)$] (the Lockean thesis with threshold $P(B_W)$) and still the equivalence holds. This means that although one might have thought that one could do just with the right-to-left half of the Lockean thesis, once one throws in enough of the logic of belief, there is no such halfway house—one always ends up with the full Lockean thesis.

The threshold term '$P(B_W)$' as employed in the Lockean thesis above is really the only choice given the logic of belief: For by P1 there must be a least believed proposition B_W; therefore, if one also wants the Lockean thesis with threshold r to be satisfied, the threshold r cannot exceed $P(B_W)$; and while r may well be a bit smaller than $P(B_W)$, it cannot be so small that some proposition ends up believed on grounds of the Lockean thesis that is not at the same time a superset of B_W, or otherwise the definition of B_W would be invalidated. Hence, in the present context, if one wants an instance of P3[r] to be satisfied at all, one may just as well use P3[$P(B_W)$] from the start—for given P1, any such P3[r] must determine the same beliefs as P3[$P(B_W)$] anyway.

By the theorem from above, in a context in which P1 and P2 have already been presupposed, we can therefore reformulate postulate P3 from before as follows:

P3 [Reformulated] B_W is P-stable, and if $P(B_W) = 1$ then B_W is the least proposition $A \subseteq W$ with $P(A) = 1$.

From the theorem above it also follows that if one has complete information about what the P-stable sets for a given probability measure P are like, then one knows exactly how to satisfy P1–P3 from above for this very P: either one picks a P-stable set of probability less than 1—if there is such—and uses it as B_W; or one uses the least proposition of probability 1 for that purpose.

Fortunately there is an algorithm that makes it very easy to compute precisely those P-stable sets over which condition (II) in our theorem quantifies. I will not go into the details,[5] but the algorithm is based on the simple fact that (for finite W)

> A is P-stable if and only if either $P(A) = 1$ or for all $w \in A$, $P(\{w\}) > P(W - A)$.[6]

If we apply the algorithm to example 1 from section 1, the only set B_W so constructed is W itself, which is at the same time the least proposition of probability 1. The corresponding threshold $P(B_W)$ is 1, but one might just as well choose some number that is less than, but sufficiently close to, 1 instead.

[5] Here is a sketch of the algorithm: Assume that $W = \{w_1, \ldots, w_n\}$, and $P(\{w_1\}) \geq P(\{w_2\}) \geq \ldots \geq P(\{w_n\})$. If $P(\{w_1\}) > P(\{w_2\}) + \ldots + P(\{w_n\})$, then $\{w_1\}$ is the first, and least, nonempty P-stable set, and one moves on to the list $P(\{w_2\}), \ldots, P(\{w_n\})$. For example, if $P(\{w_2\}) > P(\{w_3\}) + \ldots + P(\{w_n\})$, then $\{w_1, w_2\}$ would be the next P-stable set. On the other hand, if $P(\{w_1\}) \leq P(\{w_2\}) + \ldots + P(\{w_n\})$, then consider $P(\{w_2\})$. If it is greater than $P(\{w_3\}) + \ldots + P(\{w_n\})$, then $\{w_1, w_2\}$ is the first P-stable set, and one moves on to the list $P(\{w_3\}), \ldots, P(\{w_n\})$; but if $P(\{w_2\})$ is less than or equal to $P(\{w_3\}) + \ldots + P(\{w_n\})$, then consider $P(\{w_1\}), P(\{w_2\}), P(\{w_3\})$; and so forth. The procedure is terminated when the least subset of W of probability 1 is reached.

[6] See Leitgeb 2013a, sec. 2.5, for the proof ('P-stable' in the present essay corresponds to 'P-stable$^{\frac{1}{2}}$' in Leitgeb 2013a). In the computer science literature, a compatibility condition on probability measures and strict total orders of worlds has been formulated that is similar to this equivalent reformulation of P-stability: compare the "big-stepped probabilities" of Benferhat, Dubois, and Prade (1997) and Snow's (1998) "atomic bound systems."

In the case of example 2, as promised, the algorithm determines (starting at the bottom):

- $\{w_1, w_2, w_3, w_4, w_5, w_6, w_7\} \quad (r = 1.0)$
- $\{w_1, w_2, w_3, w_4, w_5, w_6\} \quad (r = 0.99994)$
- $\{w_1, w_2, w_3, w_4, w_5\} \quad (r = 0.99794)$
- $\{w_1, w_2, w_3, w_4\} \quad (r = 0.97994)$
- $\{w_1, w_2\} \quad (r = 0.882)$
- $\{w_1\} \quad (r = 0.54)$

For instance, if $\{w_1, w_2\}$ is taken to be the least believed proposition B_W, then all of P1–P3 are satisfied, and the same holds for $\{w_1, w_2, w_3, w_4\}$; in contrast, neither $\{w_1, w_2, w_3\}$ nor $\{w_1, w_2, w_4\}$ will do. To the right of the list of P-stable sets above, I have stated the corresponding thresholds $r = P(B_W)$ that are to be used in P3. The bravest option would be to use $r = 0.54$ as a threshold, in the sense that it yields the greatest number of believed propositions: all the supersets of $\{w_1\}$. The other extreme is $r = 1$ (or something just a bit below that), which is the most cautious choice: the only propositions believed by the agent will then be $\{w_1, w_2, w_3, w_4, w_5, w_6, w_7\}$ and W itself. All the other thresholds lie somewhere in between these two extremes; for example, the Lockean threshold $P(B_W)$ for $B_W = \{w_1, w_2\}$ is 0.882.

The six P-stable sets taken together look very much like one of David Lewis's "spheres systems" in his semantics for counterfactuals (see Lewis 1973): for every two of them, one is a subset of the other or vice versa. And indeed one can prove in general, including the infinite case, that if there is a P-stable proposition A with $P(A) < 1$ at all, then the set of all such propositions A is well-ordered with respect to the subset relation, and the least P-stable proposition of probability 1 is a proper superset of all of them.[7]

One final example: figure 2 shows the equilateral triangle that represents geometrically all probability measures on the set $\{w_1, w_2, w_3\}$ of worlds. For example: the w_1-corner represents the measure that assigns 1 to $\{w_1\}$ and 0 to the other two singletons; the center point represents the uniform measure that assigns $\frac{1}{3}$ to each singleton set; the closer one moves from the center toward the w_1-corner, the greater the probability of $\{w_1\}$; and so forth. The ordered numbers in the interior small triangles encode the P-stable sets for the probability measures that are represented

[7] See Leitgeb 2013a, theorem 4 'P-stable' in the present essay corresponds to 'P-stable$^{\frac{1}{2}}$' in Leitgeb 2013a, and P is also assumed to be countably additive.

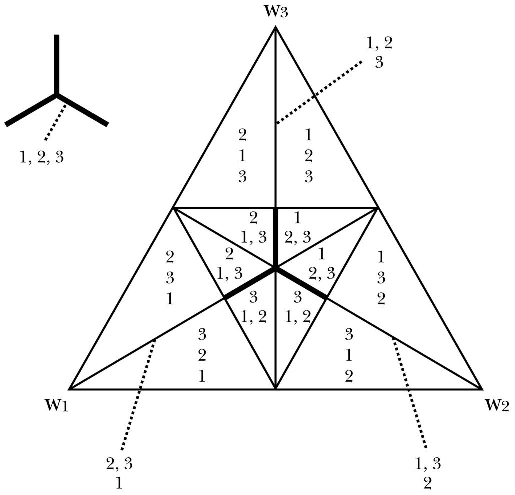

Figure 2. P-stable sets for $W = \{w_1, w_2, w_3\}$

by points within the respective triangles: for example, all P that are represented by points in the lower of the two small triangles adjacent to the w_1-corner have $\{w_1\}$, $\{w_1, w_2\}$, $\{w_1, w_2, w_3\}$ as P-stable sets; the ordered numbers $\substack{3\\2\\1}$ are the indices of worlds that, in this order, generate the P-stable sets if read from below—so worlds whose indices appear further down in a numerical array carry more probabilistic weight than the worlds whose indices appear higher up. Accordingly, every measure that is represented by a point in the upper of the two small triangles adjacent to the w_1-corner has $\{w_1\}$, $\{w_1, w_3\}$, and $\{w_1, w_3, w_2\}$ as its P-stable sets. Intuitively, all of this makes sense: in both of the small triangles, w_1 counts as the most plausible world because geometrically all of the corresponding measures are close to the w_1-corner; w_2 is more plausible than w_3 in the lower triangle because, from the viewpoint of w_1, this triangle belongs to the w_2-half of the whole equilateral triangle; things are just the other way round in the upper of the two small triangles. If one moves closer to the center again, the resulting systems of P-stable sets become more coarse grained, that is, the number of P-stable sets decreases; for example, no singleton set is P-stable anymore. Furthermore, probability measures that are represented by points that are close to each other in the triangle have similar sets of P-stable propositions.

The only points in the full equilateral triangle that represent probability measures for which *there are no P-stable propositions of probability less than 1 at all* are: the vertices and all points on the bold line segments that meet at the center of the triangle. In particular, the uniform probability measure at the center allows only for W to be stable. This gives us: *almost all* probability measures P have a least P-stable set of probability less than 1.[8] Hence, for *almost all* probability measures P, there exist an $r < 1$ and a *Bel*, such that *Bel* is closed logically, for all B it holds that $Bel(B)$ iff $P(B) \geq r$, and where there is a B, such that $Bel(B)$ and $P(B) < 1$. The same can be shown to be true if there are more than three, but still finitely many, possible worlds.

Returning to our discussion in section 1 (but using the notation for postulates that was used in the present section), we find that: for all P (on a finite space of worlds), there is an $r < 1$ such that P1, P2, P3[r] are jointly satisfied. And, additionally, almost always there is a nontrivializing way of satisfying P1, P2, P3[r] with $r < 1$ so that at least some proposition of probability less than 1 is believed.

3. The Theory and Its Costs

The results from the last section suggest a theory of belief and degrees of belief for perfectly rational agents that consists of the following three principles:

P1 There is a (uniquely determined) consistent proposition $B_W \subseteq W$, such that for all propositions B: $Bel(B)$ if and only if $B_W \subseteq B$.

P2 The axioms of probability hold for the degree of belief function P.

P3 B_W is P-stable, and if $P(B_W) = 1$ then B_W is the least proposition $A \subseteq W$ with $P(A) = 1$.

In a nutshell: *Belief is determined by a proposition of resiliently or stably high subjective probability (in the sense of P-stability).* As it were, the "grounds" (that is, the set B_W) of a perfectly rational agent's belief system must be characterized by a probabilistic stability property. Call P1–P3 the *stability theory of belief.*[9]

[8] The term 'almost all' can be made precise by means of the so-called Lebesgue measure that one finds defined in typical textbooks in measure theory.

[9] This is not the place to delve into the history of stability conceptions of belief and knowledge. Let me just point out that if Loeb (2002) is right, Hume held the view in his *Treatise of Human Nature* that beliefs are stable dispositions to have ideas of a high degree of vivacity. If we understand 'degree of vivacity' in terms of 'degree of belief', the following might thus be called the *Humean thesis on belief: It is rational to believe a proposition just in case it is rational to have a stably high degree of belief in it.* The stability theory of belief in this essay constitutes one possible way of making this Humean thesis formally precise.

By what we have shown in the previous section, it follows from this that rational belief is closed under logic, the rational degree of belief function obeys the axioms of probability, and the Lockean thesis relates belief and degrees of belief, which is what we started from in the first section. In fact, if taken together, P1–P3 as stated in this section are equivalent to the postulates stated in section 1. And we also found in the last section that for almost all P it is possible to satisfy P1, P2, P3[r] by means of a P-stable proposition B_W for which $r = P(B_W) < 1$. If measured by these consequences, P1–P3 above seem to make for a very nice normative theory of theoretical rationality as far as belief and degrees of belief are concerned—normative, because the theory is concerned with the beliefs and degrees of beliefs of perfectly rational agents.

By calling P1–P3 a 'theory', I do not mean anything like a *complete* theory of theoretical rationality for belief, let alone of the rationality of belief in general:

For a start, one would have to supplement P1–P3, which are synchronic in nature, with *diachronic* principles. It is quite clear how this would go: P1–P3 are meant to hold for all *Bel* and P at arbitrary times t. In order to add an account of how to proceed from one time t to another time t' between which all that the agent learns is some piece of evidence E for which $P(E) > 0$, one would extend P2 by maintaining that P ought to be updated by conditionalizing it on E: for all B, $P_{new}(B) = P(B \,|\, E)$. Accordingly, as recommended by belief revision theory (see AGM 1985 and Gärdenfors 1988), one would add to P1 the principle that, given some piece of evidence E that is consistent with B_W and that is therefore also consistent with every proposition believed by the agent, *Bel* ought to be updated so that Bel_{new} is the set of supersets of the new strongest believed proposition $B_W^{new} = B_W \cap E$. All of that would be consistent with P1–P3, in the sense that if *Bel* and P satisfy P1–P3, then Bel_{new} and P_{new} also satisfy the corresponding conditions that are imposed by P1–P3 on them: for B_{new} can be shown to be P_{new}-stable again.[10] Over and above that, one would also have to add principles of update on pieces of evidence E that have probability 0 or that contradict some of the agent's present beliefs (I will return to the latter case briefly in section 5).[11]

[10] Note that it is not the case for all P that if B_W is the *least* P-stable set, then B_W^{new} is the *least* P_{new}-stable set again. This is very easy to see directly, but it can also be derived from a much more general result proven by Lin and Kelly (2012b); see their corollary 1.

[11] The formal details of such joint diachronic principles for belief and subjective probability are worked out in Leitgeb 2013a.

If, finally, the resulting theory were also extended by adequate principles of *practical* rationality—a belief-desire model of action on the one hand, Bayesian decision theory on the other—the resulting package might well be suitable as a theory of rationality for belief and degrees of belief more generally. But I will restrict our discussion to P1–P3 again for the rest of the essay.

Before I turn to the potential downsides of the theory, let me make clear that it leaves a lot of room for interpretation, even substantially diverging interpretation. In particular, because of the centrality of the probabilistic notion of P-stability, one might think that the stability theory necessarily amounts to a *reductive* account of belief in terms of probability; however, such a view would be misguided.

First of all, while P3 above demands that the strongest believed proposition B_W is P-stable, it does not determine with *which* P-stable set the proposition B_W ought to be identified, and as we know already, there might be more than one choice.[12] Only if P3 were strengthened appropriately—for example, by postulating B_W to be the *least* P-stable set (which must always exist for finite W)—would one be able to explicitly define 'B_W' and hence '*Bel*' in terms of 'P', and thus be able to reduce belief to degrees of belief.[13] But we did not presuppose any such strengthening of P3 above.

Second, although the Lockean thesis is very often understood such that *Bel* can be determined from P by applying the thesis from right to left, *and therefore P is prior to Bel*, the latter 'therefore' part is not actually contained in the thesis itself. After all, the Lockean thesis with threshold r is merely a universally quantified material equivalence statement that says for all B, either $Bel(B)$ and $P(B) \geq r$, or not $Bel(B)$ and $P(B) < r$. This does *allow for* probability being prior to belief, but it does not necessitate it.

For instance, one might want to defend the Lockean thesis in conjunction with the view that *belief is prior to probability*; then the thesis is a constraint on P of the form that, given *Bel* and r, the measure P ought to be such that all and only the believed propositions are to be assigned a probability greater than or equal to r.

[12] If the sample space W is infinite, then one can prove that there are even probability measures P for which there exist infinitely many P-stable propositions of probability less than 1.

[13] That is precisely the route that I follow in Leitgeb 2013a.

More plausibly, *neither side of the Lockean thesis might be taken as prior to the other.* In that case, the thesis is a simultaneous constraint on *Bel*, P, and r, which might be regarded as a normative principle of coherence or harmony between two ontologically and conceptually distinct systems of belief, that is, the system of all-or-nothing belief and the system of quantitative belief. In order for an agent to be rational, the two systems must cohere with each other as expressed by the Lockean thesis.

I will leave open which of these interpretations is the most plausible one.[14] But all of these interpretations are consistent with P1–P3. And in all of these interpretations, belief ends up as some kind of coarse graining of probability, for, by the Lockean thesis, believing a proposition is always equivalent to abstracting away from all the different degrees of belief that a proposition might have as long as it is not less than r. For the same reason, all of the uncountably many probability measures represented by points within one and the same little triangle in figure 2 yield one and the same system of finitely many P-stable sets. In other words, in the transition from P to *Bel*, information is being lost, which was to be expected, as 'P' expresses a quantitative concept, while '*Bel*' expresses a qualitative one. But none of this entails that belief is reducible to subjective probability.

The stability theory of belief and degrees of belief looks almost too good to be true. Where have the paradoxes gone? Why is it that, all of a sudden, closure of belief under conjunction does not work against the Lockean thesis anymore? There must be a catch.

And there is. For the rest of the present section, I will discuss the two kinds of costs that follow from the principles of stability theory. These are, on the one hand, (C1) the sensitivity of the threshold in the Lockean thesis to P, and on the other, (C2) the sensitivity of *Bel* to partitionings of the set W of worlds, where, additionally, thresholds that are not particularly close to 1 demand there to be a small number of worlds or partition cells in B_W. Or to sum up these costs: *there is a serious sensitivity of belief to the context.* In the next section, I will then deal specifically with the Lottery

[14] This said, I prefer the last option according to which neither belief nor subjective probability is prior to the other. I should add that in those interpretations in which one type of belief is said to be prior to the other, one would also need to specify the *kind* of priority that one has in mind; and of course it is perfectly possible, for example, that probability is claimed to be *ontologically* prior to belief, while at the same time belief is regarded as *epistemologically* prior to probability (since beliefs seem more easily accessible than subjective probabilities). Hence, much more would have to be said about the kind of priority in question.

Paradox. Ultimately the goal will be to evaluate whether the benefits of the theory outweigh its costs.

First, according to the stability theory, only particular thresholds r [$=P(B_W)$] are permissible to be used in the Lockean thesis, as follows from the results in the last section. Which thresholds one is permitted to choose depend on what 'P' and 'B_W' in '$P(B_W)$' refer to, that is, the probability measure P and the belief set *Bel*. Furthermore, B_W is itself constrained to be P-stable. So overall, if one grants the stability theory, one must learn to live at least with the fact that

> **C1** The range of permissible choices of threshold in the Lockean thesis codepends on the agent's degree of belief function P.[15]

Let us take a step back, for a moment. What determines the choice of threshold in the Lockean thesis more generally? The usual answer is: *the context.* Compare:

> The level of confidence an agent must have in order for a statement to qualify as *believed* may depend on various features of the context, such as the subject matter and the associated doxastic standards relevant to a given topic, situation, or conversation. (James Hawthorne 2009, 73)

What this means exactly depends on whether the Lockean thesis is meant to govern the ascription of belief or the belief states themselves. In the first case, it is possible that the agent, say, x, who ascribes beliefs to an agent, y, is distinct from y. In the second case, only one agent, y, is relevant, that is, the agent whose belief states are in question. Either way, the respective threshold r in the Lockean thesis functions as a "level of cautiousness" since demanding a greater lower boundary of the probabilities of believed propositions is more restrictive than demanding a smaller lower boundary. But in the first interpretation in terms of belief *ascription*, with P being fixed, the greater r becomes, the more demanding the resulting contextually determined concept of belief and hence the more cautiously x must ascribe beliefs to y. Whereas in the second interpretation, the greater r becomes, the more restrictive the constraint on y's *belief set* becomes in the sense that y aims to be more cautious about his or her beliefs: the context in question is then what might be called y's

[15] Once again, this does not mean that degrees of belief must be determined prior to the choice of any such threshold r. For instance, for given r, a measure P might be determined so that r is the probability of some P-stable set.

own *context of reasoning*, and it comprises everything that determines *y*'s own doxastic standards at a time.

While the stability theory is open to both interpretations, in what follows I will go for the second one. In the terms of the corresponding debate on knowledge, I aim at something like a sensitive moderate invariantism for belief rather than a proper contextualist understanding. Indeed, if in the following quotation, 'knowledge' is replaced by 'belief', then I subscribe to the resulting statement:

> The kinds of factors that the contextualist adverts to as making for ascriber-dependence—attention, interests, stakes, and so on—[have] bearing on the truth value of knowledge claims only insofar as they [are] the attention, interests, stakes, and so on of the subject. (John Hawthorne 2004, 157)

Even if the agent's degree of belief function is kept fixed, if what is salient to an agent changes, then his or her beliefs might change; the more that is at stake for the agent, the more it might take him or her to believe, and so on. The question is really: how much risk is the agent willing to take whose beliefs are in question? And according to the stability theory, the subject's degree of belief function P must be counted among the factors that codetermine the answer at the relevant time; it is the subject's attention, interest, stakes, . . . , *and his or her degree of belief function* that are relevant here.

This should not be too surprising. Why should the choice of threshold in the Lockean thesis be allowed to depend on the agent's attention and interests but not on the agent's degree of belief function? After all, all of them are salient components of the agent's state of mind. Or from the viewpoint of decision theory: Assume that the Lockean thesis is taken for granted but only the choice of the corresponding threshold is left unresolved. How would a good Bayesian determine the right threshold in the corresponding context? He or she would view the whole situation as a decision problem: Should I choose the threshold in the Lockean thesis to be r_1, or should I choose it to be r_2, or . . . ? The outcome of each such choice of threshold would be a particular set of beliefs, which would be determined by plugging in that threshold in the Lockean thesis. These possible outcomes would be evaluated in terms of their utilities, and ultimately, by the tenets of standard decision theory; a threshold ought to be chosen that maximizes the expected utility of these outcomes. Hence, given the relevant utility measure *and his or her subjective probability measure*, he or she would choose a threshold so that the expected utility of the choice is maximal. In this way, obviously, P would codetermine the threshold r in the Lockean thesis simply because the expected utility of choosing one threshold rather than another codepends on P: with the utility measure being fixed, different probability measures P might well determine different ranges of permissible thresholds that all maximize expected utility relative to P. This is just like in the stability theory, in which different probability measures P may determine different ranges of permissible thresholds that all correspond to the probabilities of sets that are stable relative to P. So the dependency of r on P should not be particularly problematic in itself.

Still one might wonder: in the case of example 2 as discussed in the first two sections, why is one allowed to choose $r = 0.882$ or $r = 0.97994$ as a threshold—corresponding to B_W being either of the P-stable sets $\{w_1, w_2\}$ and $\{w_1, w_2, w_3, w_4\}$, respectively—but not, say, $r = 0.94$, which is the probability of the P-*un*stable set $\{w_1, w_2, w_3\}$?

An analogy might help here. It is well known that for some purposes, we conceive of *properties* so that every set of individuals whatsoever is guaranteed to be the extension of some property; but then again, for other purposes, we may want to restrict properties just to "natural" ones, so that not every set of individuals may count as an extension of a property in this restricted sense—a standard move in semantics, metaphysics, philosophy of science, and other areas (see, e.g., Lewis 1983). What 'natural' means exactly may differ from one area to the next, but in each case, natural properties ought to "cut nature at its joints," in some sense.

Now let us apply the same thought in the present context. For some purposes, for which the logic of belief is not relevant, we may conceive of the threshold in the Lockean thesis in the way that every threshold whatsoever can be combined with every probability measure whatsoever. But then again, for other purposes for which the logic of belief is an issue, we may want to restrict thresholds just to "natural" ones, so that not every threshold can be combined with every probability measure. Natural thresholds ought to "cut probabilities at their joints," and

r is natural with respect to P if and only if
there exists an A, such that $r = P(A)$ and for all $w \in A$,
$P(\{w\}) > P(W - A)$

may be just the kind of "probability cutting" that is appropriate here. As pointed out in the previous section, if $P(A) < 1$, then the condition that for all $w \in A$, $P(\{w\}) > P(W - A)$, is equivalent to A being P-stable. And if A is the least set of probability 1 (and W is finite), then it also holds that for all $w \in A$, $P(\{w\}) > P(W - A)$.

Or analogously: if one is interested only in the logic of belief, then every consistent proposition whatsoever may be a candidate for the strongest believed proposition B_W. However, in a context in which both belief and degrees of belief are of interest, only "probabilistically natural" propositions may count as candidates for B_W, and P-stability may be just the right notion of naturalness since it belongs to the same ballpark as other "natural" notions of stability or resiliency or robustness in statistics (see Skyrms 1977, 1980), economics (see Woodward 2006), metaphysics (see Lange 2005), and beyond. Hence, the fact that P1–P3 impose more constraints on r than P3 would do just by itself and that P1–P3 impose more constraints on B_W than P1 would do just by itself should not be thought to speak against the theory.

Now for the second, and more substantial, worry: according to the stability theory of belief, it turns out that

C2(i) belief is partition dependent, and

C2(ii) generally, the smaller the partition cells are in terms of probability, the greater the probabilities of believed propositions must be in order for P1–P3 to be satisfied.

Let me explain this in detail (still presupposing W to be finite). It is quite common in applications of probability theory that even when initially P had been defined for all subsets of W, there might be a context in which not all subsets of W are actually being required for the purposes in question. For example, if one is interested only in the possible outcomes of a lottery, then only the propositions of the form *ticket* 1 *wins*, *ticket* 2 *wins*, . . . together with their logical combinations will be relevant; accordingly only the probabilities of such propositions will count. Formally, this can be achieved by introducing a *partition* Π on W: a set of pairwise disjoint nonempty subsets u_i of W, such that the union of these sets u_i is just W again. For example, in the lottery case, initially W might have been the set of all metaphysically possible worlds, but then a set of partition cells u_i might have been introduced, such that any such set u_i would be the set of all worlds in which ticket i wins.[16] Such partition cells u_i might then be viewed themselves as "coarse-grained" possible worlds in

[16] Let me disregard the question of whether such a class u_i would actually be a set or rather a proper class of worlds.

which all differences between two distinct metaphysically possible worlds within one and the same cell would be ignored; the probabilities of these "pseudo-worlds" would be given by $P(u_i)$, and only unions of such sets u_i would be considered propositions in the relevant context.

If one wants to make all of that completely precise, one needs to build up a new probability space that has the set Π of all partition cells as its sample space, where propositions are now subsets of Π, and in which a whole new probability measureP_Π is being defined in terms of P. The probability space in example 1 from section 1 could be seen as arising from precisely that procedure, with each "coarse-grained world" in W corresponding to a particular ticket winning in a fair lottery of one million tickets.

If the context changes again, and one needs to draw finer distinctions than before—for example, it is not just relevant which ticket wins but also who bought the ticket—one may refine the partition accordingly, so that what had been one partition cell u_i before is being broken up into several new and smaller partition cells. Or one can afford to draw coarser distinctions—for example, it is not relevant anymore which ticket wins but only whether ticket 1 wins or not—and hence the partition is made coarser, so that what had been several partition cells before are now being fused into just one large partition cell.

In each case, the probabilities of the partition cells and of their unions are determined from the original probability measure P that is defined for all subsets of W, or equivalently: where the original probability measure is given with respect to the maximally fine-grained partition whose partition cells are just the singleton sets $\{w\}$ for $w \in W$. For it does not really matter whether W equals $\{w_1, \ldots, w_n\}$ or whether the set of "worlds" is considered equal to the maximally fine-grained partition $\Pi = \{\{w_1\}, \ldots, \{w_n\}\}$ of W; whether the probability measure is P or whether it is the measure P_Π that assigns to the singleton set $\{\{w_i\}\}$ the same number that P assigns to the singleton set $\{w_i\}$; more generally, it does not matter whether P_Π assigns to $X \subseteq \Pi$ the number that P assigns to $\cup X$, that is, to the set of members of members of X; and in terms of the intended interpretation of propositions, it does not matter whether the proposition that ticket 1 is drawn is $\{w_1\}$ or $\{\{w_1\}\}$, and so forth. Accordingly, in the following, I will move back and forth between such numerically distinct but formally equivalent constructions of worlds, propositions, and probability measures, without much additional comment.

Since operating with partitions is a natural and useful doxastic procedure, it is important to determine what happens to an agent's beliefs when partitions are introduced and changed. If P1–P3 are taken for granted, the answer is: C2(i) refining a partition may lead to a change of beliefs, in particular, to a loss of beliefs; and C2(ii) whatever the partition is like, in order for P1–P3 and $P(B_W) < 1$ to be satisfied, the probability of every singleton subset of B_W must be greater than the probability of $W - B_W$, whether the members of B_W are some "maximally fine-grained worlds" in W or some more or less coarse-grained partition cells on W. I will illustrate finding C2(i) in terms of an example, and I will demonstrate C2(ii) and its consequences by means of a little calculation:

Example 1–Reconsidered

Let $W = \{w_1, \ldots, w_{1,000,000}\}$ be a set of 1,000,000 possible worlds again, where each world w_i corresponds to ticket i being drawn in a fair lottery. Accordingly, let P be the uniform probability measure that is given by $P(\{w_1\}) = \ldots = P(\{w_{1,000,000}\}) = \frac{1}{1,000,000}$ again.

Now introduce the partition

$$\Pi = \{\{w_1\}, \{w_2, \ldots, w_{1,000,000}\}\},$$

of W, or in other words: the agent is interested only in whether ticket 1 wins or not. Consider the partitions cells $\{w_1\}$ and $\{w_2, \ldots, w_{1,000,000}\}$ as new coarse-grained worlds and Π as the resulting new set of such worlds. Based on our original P, we can then define a new probability measure P_Π, for which Π serves as its sample space, and where P_Π assigns probabilities to subsets of Π as expected: $P_\Pi(\{\{w_1\}\}) = \frac{1}{1,000,000}$, $P_\Pi(\{\{w_2, \ldots, w_{1,000,000}\}\}) = \frac{999,999}{1,000,000}$, $P_\Pi(\{\{w_1\}, \{w_2, \ldots, w_{1,000,000}\}\}) = 1$, $P_\Pi(\emptyset) = 0$. The new probability for a set X results from applying the original probability measure P to $\cup X$ (the set of members of W that are members of the partition cells in X); in particular, $P_\Pi(\{\{w_1\}\}) = P(\{w_1\})$ and $P_\Pi(\{\{w_2, \ldots, w_{1,000,000}\}\}) = P(\{w_2, \ldots, w_{1,000,000}\})$.

The algorithm from section 2 (as sketched in note 6) tells us then that the corresponding P_Π-stable sets are

$$\{\{w_2, \ldots, w_{1,000,000}\}\} \text{ and } \{\{w_1\}, \{w_2, \ldots, w_{1,000,000}\}\},$$

the first one of which has a probability slightly less than 1, while the second one has a probability of exactly 1.

Finally, let $B_W^\Pi = \{\{w_2, \ldots, w_{1,000,000}\}\}$ and $r = P_\Pi(\{\{w_2, \ldots, w_{1,000,000}\}\})$: then all of P1–P3 are satisfied, and since $\{\{w_2, \ldots, w_{1,000,000}\}\}$ is nothing but the negation of the proposition $\{\{w_1\}\}$, this means that the agent believes that ticket 1 will not win (relative to Π).

In order to drive the point home, let us now maximally refine Π to Π' again so that one is interested again in which ticket will be drawn; or equivalently, simply use the original W and P again. Then, as observed

Published by Duke University Press

> already in section 2, W is the only P-stable set, and our theory demands that $B_W = W$. Consequently, the agent does *not* believe that ticket 1 will not win (relative to the most fine-grained available partition). That is, refining a partition can lead to a change of beliefs.

In section 4, I will return to this example, when I will evaluate its consequences for the Lottery Paradox. So much concerning C2(i), for the moment.

And about C2(ii) from above: this is just the alternative characterization of P-stable sets again that we had observed in section 2, but we will see that it makes better sense to address its consequences in a context in which one discusses the workings of partitions.

As stated in section 2, a set A is P-stable if and only if either $P(A) = 1$ or for all $w \in A$, $P(\{w\}) > P(W - A)$. So, by P1–P3, if we are dealing with the "nontrivial" case $P(B_W) < 1$, every singleton subset of B_W must have a probability greater than $\neg B_W$, whether the worlds in question are the "given" worlds in W or some more or less coarse-grained "pseudo-worlds" as determined from some partition Π of W. Either way, consequently, for all $w \in B_W$, $P(\{w\}) > 1 - P(B_W)$, and hence, for all $w \in B_W$, $P(B_W) > 1 - P(\{w\})$. In words, if the probability of some serious candidate for the actual world is really small, then $P(B_W)$, and hence the probability of *every* believed proposition, must be really high, or otherwise P1–P3 could not hold jointly. Or contrapositively: if $P(B_W)$, or for that matter the probability of *some* believed proposition, is not particularly high, then the probabilities of all worlds or partition cells in B_W cannot be particularly low either. For instance, if one wants P1–P3 to hold, and the agent ought to believe some proposition of probability 0.91, then all worlds or partition cells in B_W need to have a probability of more than 0.09. That is, B_W cannot contain more than ten worlds or partition cells. Or if P1–P3 are meant to be satisfied, and the agent ought to believe some proposition of probability 0.98, then all worlds or partition cells in B_W must have a probability of more than 0.02. Therefore, B_W cannot contain more than forty-eight worlds or partition cells. And if we let the number of members of B_W go to infinity, then the probability of B_W, and thus of every believed proposition, must tend to 1 in the limit.

Observations C2(i) and C2(ii) should make the limitations of the stability theory of belief quite clear. How serious are they, and what, if anything, can I say in defense of the theory?

C2(i) suggests that P1–P3 taken together make belief dependent on, or relativized to, partitions. If, as is plausible, we count the agent's choice of partition as belonging to the context of reasoning in which the agent's beliefs "take place," or if the partition is at least determined from such a context, then we might say that belief ends up relativized to contexts. But this should not take us by surprise anymore: We have already seen that, according to P1–P3, the threshold in the Lockean thesis—and thus what the agent believes—depends on the context (comprising the agent's attention, interests, stakes, degrees of belief, and so forth). We have also made clear already that this does not entail any kind of priority of probability over belief.

What we have established now is that the agent's manner of partitioning *W* ought to be also included in the context on which the agent's beliefs depend. But in view of the general impact that the context has on belief according to the stability theory, this is hardly a big deal at this point of argumentation.

Furthermore, there are quite a few well-known and successful theories around that presuppose probabilities of some sort and for which the same relativization to partitions can be observed—take Levi's (1967) theory of acceptance, in which partition cells are regarded as the "relevant answers" to a question posed by the agent, or Skyrms's (1984) theory of objective chance, in which partition cells are "natural hypotheses" that derive from the causal-statistical analysis of a subject matter. In the former theory, what an agent *accepts* at a time depends on what he or she regards as relevant answers, and in the latter theory, the *chance* of an event depends on what hypotheses the agent regards as natural candidates and how the agent distributes his or her subjective probabilities over them.

Third, C2(i) does not just affect the stability theory but really a much wider class of theories of belief/acceptance and probability, as proven by Lin and Kelly (2012a, secs. 13–14). Stability theory's partition dependence of belief is just a special case of the general phenomenon that Lin and Kelly refer to as lack of "question-invariance" of acceptance.

Fourth, there are even some empirical findings on the so-called Alternative-Outcomes Effect that seem to support the view that belief is partition sensitive (see, e.g., Windschitl and Wells 1998): if possible scenario outcomes are presented to people in terms of different partitions (e.g., you hold three raffle tickets and seven other people each hold one vs. you hold three and another person holds seven), then the participants' *numerical* probability estimates of the focal outcomes remained unaffected, while their corresponding *nonnumerical* or qualitative certainty estimates turned out to be sensitive to the partitions. I do not claim that I could rationally reconstruct these experimental results on the basis of the stability account; and, as always, it is not so clear what kind of bearing empirical results like these should have on a normative theory of rational belief; but at least such findings indicate that actual beliefs of actual people are indeed partition dependent.

Finally, while the stability theory has it that an agent's beliefs may change from one partition to another, there are also some invariances: the same logical closure conditions apply to believed propositions relative to every partition whatsoever; relative to every partition, the probability of every believed proposition must exceed that of its negation (by the Lockean thesis); and one can also derive a couple of cross-partition laws. For example, take a partition to be given. A set B_W has been determined to be the strongest believed proposition. Now coarsen the partition inside of B_W (or do not change anything there) and repartition anyway you want outside of B_W. However, do not repartition so that any partition cells from inside of B_W and from outside of B_W end up in the same cell in the new partition. If you abide by these constraints on repartitioning, then the original B_W still determines a set that is also P-stable on the new partition. Only if a partition is altered on B_W without making it coarser there, previously P-stable sets may no longer have stable counterparts after repartitioning. So it is not as if the theory entailed that changing partitions would always affect an agent's beliefs in some erratic and unpredictable manner.[17]

My overall diagnosis is that while belief certainly becomes more strongly dependent on contexts than one might have hoped for, no decisive argument against P1–P3 emerges from C2(i).

Now for C2(ii) from above, that is, if $P(B_W) < 1$, then the probability of every world or partition cell in B_W must be greater than the probability of $W - B_W$. Given P1–P3, this leaves agents with the following options: (A) Either they believe some proposition with a probability that is not particularly close to 1, in which case they can make only very few distinctions in terms of serious possibilities in B_W; or (B) they believe only propositions that have probabilities very close to, or identical with, 1, in which case they are flexible about drawing fine-grained distinctions within B_W; or (C) they opt for a position in between the two extremes. Let us assume that W itself is very fine grained in the sense of containing a lot of worlds. Then, by means of partitioning, the obvious manner of realizing (A) is to introduce a partition that is very coarse grained on B_W;

[17] I am grateful to an anonymous referee whose comments on this part were invaluable.

for (B) a very fine-grained partition on B_W will be the right choice; and (C) will be the case if agents opt for a partition that lies somewhere in between.

Option (B) should be appealing to all those who defend a view according to which believed propositions ought to have a degree of belief of 1 *in general*, for option (B) approximates that kind of position. Examples in the relevant literature would be Levi (1980), van Fraassen (1995), Arló-Costa (2001), Arló-Costa and Parikh (2005); and a closely related position is held by Williamson (1998) if 'belief' is replaced by 'knowledge' and 'subjective probability' by 'epistemic probability'. All of these proposals also share with the present theory the assumption that ideal belief (or, in Williamson's case, ideal knowledge) is closed under logic. By invoking further resources—for instance, by starting from a primitive conditional probability measure (Popper function) P, as van Fraassen (1998), Arló-Costa (2001), and Arló-Costa and Parikh (2005) have done—one might even finesse P1–P3 so that option (B) would get even closer to some of these proposals, for example, by singling out only particular sets of probability 1 or only particular sets of very high probability as believed propositions. Consequently, P1–P3 cannot be much worse off than these proposals, as P1–P3 allow for them, or for something close to them, to be realized. But P1–P3 are also less restrictive than these proposals by not turning option (B) into a general requirement *in all contexts*.

Option (A) ought to be attractive to anyone who favors the Lockean thesis with a "realistic" threshold that is not particularly close to 1; examples include Kyburg (1961, 1970), or more recently, Kyburg and Teng (2001), Foley (1993), Hawthorne and Bovens (1999), and Hawthorne and Makinson (2007). Of course, in contrast with the current theory, these proposals do not include the closure of belief under conjunction, but that might be because they think *one could not have* it in the presence of the Lockean thesis anyway, which is not right given what we found in the first two sections. The downside of P1–P3 if compared to these other Lockean proposals is the additional constraint that in order to realize (A), one needs to reason relative to sufficiently likely serious possibilities only.

But how severe is this constraint? Is it really plausible to assume that when we have beliefs and when we reason on their basis, we always take into account *every maximally fine-grained possibility whatsoever*? Instead, in typical everyday contexts, we might reason relative to some contextually determined partition of *salient* and *sufficiently likely* alternatives. Say, for some reason in some context, we are interested only in whether the three propositions *A*, *B*, *C* are the case or not. Hence, the possible worlds or partition cells on which we concentrate are precisely all of the logical combinations of these three propositions,

$$A \wedge B \wedge C, A \wedge B \wedge \neg C, A \wedge \neg B \wedge C, \ldots, \neg A \wedge \neg B \wedge \neg C,$$

and we take into account only the propositions that can be built from them. For instance, in formal epistemology, when one studies confirmation or coherence or learning, one typically does so by means of "small" probability spaces that may well correspond to what is required by option (A). Indeed, when I turn to a concrete application of my theory in section 6, I will deal with precisely such a situation in which only the logical combinations of three propositions happen to be relevant. More generally, when we represent an argument from natural language in logical terms, we usually follow Quine's (1960, 160) *maxim of shallow analysis* and end up with a formalization in terms of, say, just a couple of propositional letters.[18] When people draw inferences in everyday situations, then, according to what is perhaps the empirically most successful theory of reasoning in cognitive psychology—Johnson-Laird's theory of mental models—they do not do so by representing infinitely many superfine-grained possibilities but rather by representing the, usually very few, distinctions that are required in order to build a model of the situation. And so on. In all of these cases, it seems that satisfying P1–P3 along the lines of option (A) should be perfectly viable. It is only when one's attention gets directed toward a great number of case distinctions that belief ever gets closer to having a probability of 1. Adapting the title of Lewis 1996, rational belief also turns out to be elusive then.

Finally, the stability theory of belief allows for continuous transitions between options (A) and (B) and hence for the compromise option (C). All of these options are still governed by the same set of general principles, that is, P1–P3.

Let us take stock. If P1–P3 are satisfied, and thus their consequences C1, C2(i), and C2(ii) are true as well, the following picture of our perfectly rational agent emerges: The agent must hold his or her beliefs, and reason upon them, always relative to a context that involves the agent's attention, interests, stakes, the degrees of belief function *P*, and more. The context must include or determine a partition of the

[18] "A *maxim of shallow analysis* prevails: *expose no more logical structure than seems useful* for the deduction or other inquiry at hand." Quine 1960, 160.

underlying set of presumably very fine-grained worlds into more or less coarse-grained partition cells that figure as "pseudo-worlds" in the subsequent reasoning processes. Additionally, the context restricts the permissible thresholds in the Lockean thesis to a range of natural candidates that are given by the probabilities of *P*-stable sets. From these thresholds, whether implicitly or explicitly, the agent needs to choose the one that is to be used for the Lockean thesis: the greater the threshold, the more cautious the agent will be about his or her beliefs; but the greater the threshold, the greater also the number of serious possibilities that the agent is potentially able to distinguish.

In this way, the agent is able to maintain the logic of belief, the axioms of probability, and the Lockean thesis simultaneously. The price to be paid is this very dependency of belief on contexts. Accordingly, while the logic of beliefs does hold locally within every context, logical inferences across contexts are not licensed unrestrictedly. But P1–P3 also guarantee some doxastic invariances across contexts. Moreover, in a lot of everyday and scientific contexts, agents may restrict themselves to coarse-grained possibilities without loss, and the fallback position of reasoning in terms of the most fine-grained partition is available to them too, in which case P1–P3 amount to a more conservative "probability 1" account of belief (or something close to it).

While it is always hard to weigh the benefits of a theory against its limitations, so far, the logic of belief, the axioms of probability, and the Lockean thesis seem to do quite well against the drawbacks of contextualization.

In the next section, I will put the theory to the test again by considering how well it does in the face of paradox.

4. Application to the Lottery Paradox

"Solving" a paradox by a theory usually involves the following ingredients: the theory should avoid the absurd conclusion of the paradox; it should preserve some, or many, of the original premises of the paradox; the theory should explain why some of the premises need to be given up; and it should explain why those premises that are given up appeared to be true initially, by explaining—and maybe explaining away—the intuitions that seemed to warrant these premises.

I want to argue that the theory from the last section does solve the Lottery Paradox. (I will not deal with the Preface Paradox; to the extent to which the Preface Paradox resembles the Lottery Paradox, similar considerations apply, but the Preface story involves additional complications that I do not want to get into here.) My main task will be to interpret and evaluate two of the formal examples that we had already encountered before: example 1 from section 1 and example 1–reconsidered in the last section.

A fair lottery of one million tickets will be played. By the Lockean thesis, a rational agent ought to believe of each ticket that it will not win because each ticket is very likely to lose. But it is also plausible that belief is closed under conjunction and that the agent's degrees of belief should reflect the fairness of the lottery. Taking these together leads to contradiction, along the lines of what was pointed out in (1a) of example 1 in section 1.

What does the joint stability theory of belief and degrees of belief predict concerning this paradox? First, for $W = \{w_1, \ldots, w_{1{,}000{,}000}\}$ and P being uniform over W again, it suggests that a partition of the underlying set of worlds needs to be determined. The salient options are:

- In a context in which the agent is interested in *whether ticket i will be drawn*; for example, for $i = 1$: Let Π be the corresponding partition $\{\{w_1\}, \{w_2, \ldots, w_{1{,}000{,}000}\}\}$. The resulting probability measure P_Π is given by P so that:

  $$P_\Pi(\{\{w_1\}\}) = \frac{1}{1,000,000}, P_\Pi(\{\{w_2, \ldots, w_{1,000,000}\}\}) = \frac{999,999}{1,000,000}$$

  As determined in example 1–reconsidered, there are two P_Π-stable sets, and one of the two possible choices for the strongest believed proposition B_W^Π is $\{\{w_2, \ldots, w_{1{,}000{,}000}\}\}$. If B_W^Π is chosen as such, our perfectly rational agent believes of ticket $i = 1$ that it will not be drawn, and of course P1–P3 are satisfied.

  For example, this might be a context in which a single ticket holder—the person holding ticket 1—would be inclined to say of his or her ticket: "I believe it won't win."

- In a context in which the agent is interested in *which ticket will be drawn*:

  Let Π' be the corresponding partition that consists of all singleton subsets of W, or equivalently: keep W as it is. Consequently, the probability measure $P_{\Pi'}$ can be identified with P again, and it is distributed uniformly over the one million alternatives.

> As mentioned in example 1–reconsidered, the only P-stable set—and hence the only choice for the strongest believed proposition B_W—is W itself: our perfectly rational agent believes that some ticket will be drawn, but he or she does not believe of any ticket that it will not win.[19] Of course, P1–P3 are satisfied again.
>
> For example, this might be a context in which a salesperson of tickets in a lottery would be inclined to say of each ticket: "It might win" (that is, it is not the case that I believe that it won't win).

The same relativization to partitions had been exploited already by Levi (1967, 40), in order to analyze Lottery-Paradox-like situations.

In either of the two contexts from before, the theory avoids the absurd conclusion of the Lottery Paradox; in each context, it preserves the closure of belief under conjunction; and in each context, it preserves the Lockean thesis for some threshold ($r = \frac{999{,}999}{1{,}000{,}000}$ in the first case, $r' = 1$ in the second case)—all of this follows from P-stability and the theorem from section 2. In the first Π-context, the intuition is preserved that, in some sense, one believes of ticket i that it will lose since it is so likely to lose. In the second Π'-context, the intuition is preserved that, in a different sense, one should not believe of any ticket that it will lose since the situation is symmetric with respect to tickets, as expressed by the uniform probability measure, and of course some ticket must win. Finally, by disregarding or mixing the contexts, it becomes apparent why one might have regarded all of the premises of the Lottery Paradox as true. But according to the present theory, contexts should not be disregarded or mixed: partitions Π and Π' differ from each other, and different partitions may lead to different beliefs, as observed in the last section and as exemplified in the Lottery Paradox. Accordingly, the thresholds in the Lockean thesis may have to be chosen differently in different contexts, and once again, this is what happens in the Lottery Paradox—which makes good sense: in the second Π'-context, by uniformity, the agent's degrees of belief do not give him or her much of a hint of what to believe. That is why the agent ought to be supercautious about her beliefs in that

[19] Douven and Williamson (2006) prove on very general grounds that if a probability space is "quasi-equiprobable" (their term)—a generalization of uniform or equiprobable probability measures—the corresponding belief set must either consist only of propositions of probability 1 or it must include a proposition of probability 0. B_W coinciding with W falls under the first disjunct, of course.

context; hence the maximally high threshold. In contrast, in the first Π-context, the agent's degrees of belief are strongly biased against ticket *i* being drawn. That is why the agent may afford to be brave in terms of his or her beliefs about *i* not winning in that context. No contradictory conclusion follows from this since, according to the stability theory, it is not permissible to apply the conjunction rule for beliefs across different contexts.

This seems to be a plausible rational reconstruction and solution (in the sense specified before) of the Lottery Paradox, based on the theory from the last section.

I conclude that the stability theory handles the Lottery Paradox quite successfully. The context sensitivity of belief that was observed in the previous section actually works to the theory's advantage here since one can analyze the different reasons for assuming the various premises in the paradox in terms of different contexts, without running into contradictions. And the contexts in question arise naturally—from the interest in a particular ticket winning or not, or the interest in which ticket will be winning.

In the next section, I will turn to an application of the theory apart from such paradoxical circumstances.

5. An Application in Formal Epistemology

Sometimes, when we analyze a concept, problem, or question on the basis of subjective probabilities, we still want to be able to express our findings also in terms of beliefs. Or the other way round. Or we want to refer both to belief and probability right from the start. In all of these cases, a joint theory of belief and degrees of belief is required.

In this section, I will present an example of the first kind by applying the stability theory of belief in the context of Bayesian formal epistemology.

By the *secular acceleration of the moon*, one refers to the phenomenon that the movement of the moon around the earth appears to accelerate slowly. Astronomers had been aware of this for a long time, and in the nineteenth century, they wanted to explain the phenomenon by means of the physics at the time, that is, Newtonian mechanics, which turned out to be a nontrivial problem.

In logical terms, when T is the relevant part of Newtonian mechanics, H is a conjunction of auxiliary hypotheses including the assumption that tidal friction does not matter, and E is the observational evidence for the moon's secular acceleration, then T and H together logically imply $\neg E$. In other words: T, H, and E are not jointly satisfiable. So given E, either T or H needs to be given up, and it is not clear which—a classical Duhem-Quine case of underdetermination of theories by evidence, or so it seems.

That is where the Bayesian story begins: Dorling (1979) argues that this apparent instance of underdetermination vanishes as soon as one takes into account subjective probabilities. For that purpose, he reconstructs what might be called the "ideal" astrophysicist's degrees of belief at the time. Obviously, this is all fictional, but that is how it goes with rational reconstructions, and Dorling does a sophisticated job of deriving the probability measure on systematic grounds. He ends up with precisely the probability measure from example 2 as discussed in the first two sections, with 'T' replacing 'A', 'H' replacing 'B', and 'E' replacing 'C'; compare figure 1 from section 1. Hence, $T = \{w_1, w_2, w_5, w_8\}$, $H = \{w_1, w_3, w_7, w_8\}$, and $E = \{w_5, w_6, w_7, w_8\}$. Since '$T$', '$H$', and '$E$' are treated like propositional letters here, the probability of $T \wedge H \wedge E$ needs to be set to 0 "by hand," for the logically omniscient ideal astrophysicist at the time already knew that this conjunction could be ruled out. Accordingly, in example 2, the probability of $\{w_8\}$ had been set to 0. The probability space as a whole is a typical case of a Bayesian philosopher of science abstracting away from all further complications, such as the precise propositional contents of the single axioms of T, the various conjuncts of H, and the various data that are summarized by E. In terms of coarse graining, when I introduce beliefs into this Bayesian model further below, I will thus be heading for option (A) from section 3.

Now what is the Bayesian response to the Duhem-Quine case? The prior probability measure P assigns a high degree of belief to Newtonian mechanics, it assigns a degree of belief to the conjunction of the auxiliary hypotheses that is greater than what it assigns to its negation, and it assigns initially a tiny probability to E:

$$P(T) = 0.54 + 0.342 + 0.018 = 0.9,\ \ P(H) = 0.54 + 0.058 + 0.00006 = 0.59806,\ \ \text{and}\ P(E) = 0.002 + 0.00006 = 0.02006.$$

A perfectly rational Bayesian agent would then update his or her degrees of belief by the relevant evidence E: the resulting new degrees of belief are

$$P_{new}(T) = P(T \mid E) = 0.8976, P_{new}(H) = P(H \mid E) = 0.003,$$

$$P_{new}(E) = P(E \mid E) = 1.$$

This means that, after taking into account the observational data, the ideal astrophysicist at the time still ought to have assigned a high degree of belief to Newtonian mechanics. Moreover, she had become certain about the evidence, but she should have assigned only a tiny degree of belief to the conjunction of the auxiliary hypotheses. And that is pretty much what happened in actual history: physicists gave up some of the auxiliary assumptions, including the one of tidal friction being negligible, but of course they continued to support Newtonian mechanics. No Duhem-Quine problem emerges: a success story of Bayesianism.

This said, Dorling (1979, 179) mentions that "while I will insert definite numbers so as to simplify the mathematical working, nothing in my final *qualitative* interpretation will depend on the precise numbers"; and that better be right—because of the fictional character of P, it would be ridiculous if any of Dorling's findings depended on his precise choice of numbers. Dorling (1979, 180) also states that "scientists always conducted their serious scientific debates in terms of finite *qualitative* subjective probability assignments to scientific hypotheses," the idea being that scientists never put forward numerical degrees of belief in their academic debates. Instead, they argue that some hypothesis is highly plausible, that given some hypothesis some other hypothesis is not very plausible at all, or the like.[20]

However, Dorling does not seem to have the resources available to derive the intended *qualitative* interpretation of his probabilistic results in any systematic matter, or to prove the *robustness* of his interpretation under slight modifications of numbers, or to offer any precise account of *qualitative* subjective probability assignments.[21]

But there is an obvious way of filling this gap: by expressing Dorling's findings by means of the qualitative concept of belief, based on a joint theory of belief and subjective probability. And the stability theory

[20] In Dorling's (1979, 180) own terms: scientists use expressions such as "more probable than not," "very probable," "almost certainly correct," "so probable as to be almost necessary," and so on.

[21] Sometimes by 'qualitative probability' one means *comparative* probability: probability theory based on the primitive predicate 'is at least as likely as'. And that is certainly available to Dorling. But at the same time, that is not how Dorling (1979, 179) understands 'qualitative probability'. As he points out, in order for his example to work, "H should have been regarded at the time as more probable than not and T should have been regarded as substantially more probable than H." In order to make these locutions precise, he concludes, "something semi-quantitative is necessary" for which comparative probability is not sufficient.

of belief seems to be the obvious choice for this purpose, for the following reasons:

First, Dorling's argument seems to rely, if only tacitly, on the following inference step: he determines that, after taking account of evidence, the probability of T is high and the probability of H is tiny, *from which he concludes that T ought to be maintained, but H ought to be abandoned.* After all, he wants to justify why scientists gave up on H but not on T, and giving up is still a binary act. It is hard to see anything else to be in operation here than a version of the Lockean thesis, which is what P3 offers.

Second, according to the stability theory, and as I argued in section 3, belief turns out to be a coarse-grained version of subjective probability, again due to the presence of the Lockean thesis. So when we translate facts about P into facts about *Bel* by means of the Lockean thesis, we know that a lot of information is being abstracted away—infinitely many probability measures will correspond to one and the same belief set. What is more, we have seen in figure 2 that probability measures whose geometric representations are close to each other also yield similar P-stable sets and hence similar candidates for B_W. Therefore, if we can confirm Dorling's diagnosis about underdetermination in terms of the ideal astrophysicist's *beliefs* as determined by the stability theory, we can be quite certain that he was right when he claimed that his interpretation did not "depend on the precise numbers."

Third, scientists do seem to express their own beliefs and criticize the beliefs of others when they conduct "their serious scientific debates," and they also apply the standard logical rules, including closure under conjunction, when they do so: picture a scientist writing A on a blackboard and then later B, arguing that both are satisfied, and then imagine another scientist stopping his colleague from writing $A \wedge B$ further down below—this would certainly seem at odds with scientific practice. Which gives us P1.

So the all-or-nothing concept of belief, with P1 and P3 from section 3 in the background, seems to be precisely what is required to supply Dorling with the lacking theoretical resources. Since P2 is a given anyway by Bayesian lights, the stability theory of belief is what emerges.

In sections 1 and 2, we already determined the six P-stable sets that result from Dorling's choice of numerical values. According to the stability theory, a perfectly rational agent's beliefs at the time need to be given by one of these P-stable sets. We settle for the bravest possible choice in light of the fact that the probability of H is not particularly high; this gives us:

$$B_W = \{w_1\} (r = 0.54)$$

At this point, the agent believes Newtonian mechanics, the conjunction of the auxiliary hypotheses, and the negation of E—that is: $Bel(T)$, $Bel(H)$, $Bel(\neg E)$—as well as all of their logical consequences, for example, $Bel(T \wedge H \wedge \neg E)$. *Bel* and P taken together satisfy the Lockean thesis with $r = 0.54$ as a threshold. We also know from the previous sections that if that Lockean threshold had not been given by the probability of a P-stable set, then belief would not have been closed under conjunction. For example, it might have been the case then that $Bel(T)$, $Bel(H)$, and $Bel(\neg E)$ without $Bel(T \wedge H \wedge \neg E)$ being the case at the same time.

Just as in the probabilistic story from before, the next step for the agent is to update his or her beliefs by means of $E = \{w_5, w_6, w_7, w_8\}$. Since E contradicts B_W, that is, since the agent had expected $\neg E$ to be true beforehand, this is a case of proper belief *revision* in the sense of AGM (1985) and Gärdenfors (1988). The standard method of revision in such a case (see Grove 1988 for the details), given a sphere system of doxastic fallback positions, is for the agent to move to the least sphere that is consistent with the evidence, to intersect it with the evidence, and to use the resulting set B_W^{new} of worlds as the new strongest believed proposition. Formally this is just like a Lewis-Stalnaker semantics for conditionals, where one considers the least sphere that is consistent with the antecedent proposition, one intersects the two, and then one determines which consequent propositions are supersets of that intersection.[22]

If we use the total set of P-stable propositions as the obvious choice of sphere system (recall section 2), then the least P-stable set that is consistent with E is

$$\{w_1, \ldots, w_5\}.$$

Intersecting it with E yields

$$B_W^{new} = \{w_5\}.$$

Therefore, the propositions that the agent believes after the update are precisely the supersets of $\{w_5\}$.[23]

[22] I take what David Lewis called the "Limit Assumption" for granted here (1973).

[23] More on the corresponding stability account of belief *revision* (or *conditional* belief) can be found in Leitgeb 2013a.

This means that after taking into account the observational data, the ideal astrophysicist at the time still ought to have believed Newtonian mechanics. Moreover, she should have taken on board the evidence but also believed the *negation* of the conjunction of the auxiliary hypotheses. In short:

$$Bel_{new}(T), Bel_{new}(\neg H), Bel_{new}(E);$$

and, accordingly,

$$Bel_{new}(T \wedge \neg H \wedge E).$$

Once again, this is exactly what happened in actual history. And all of this is consistent with stability theory and with the previous purely probabilistic considerations since B_W^{new} turns out to be P_{new}-stable again (where $P_{new}(.) = P(. \mid E)$).[24] We can thus confirm Dorling's intended qualitative conclusions by applying the stability theory of belief to what would otherwise be a purely Bayesian, and hence quantitative, theory.

6. Summary

I have presented a theory of belief and degrees of belief that combines three parts, P1–P3, that are usually thought to lead jointly to trivialization or inconsistency; in particular, the theory includes the closure of rational belief under conjunction and the Lockean thesis on rational belief. In the first two sections, I made it clear that, actually, neither trivialization nor inconsistency follows from these assumption. In section 3, I gave the official formulation of the theory, which I called the "stability theory" because of the central notion of *P*-stability that figures in it and that indeed entails the closure of belief under conjunction and the Lockean thesis. But I also discussed the main cost of the theory: a strong form of sensitivity of belief to context. In particular, the theory entails that what an agent believes rationally will depend crucially on how the underlying space of possibilities is partitioned. However, I argued that the benefits of the theory seemed to outweigh its limitations. In section 4, I showed that the theory is able to handle the Lottery Paradox. Finally, section 5 dealt with a concrete application of the theory to a problem in formal epistemology, which demonstrated that this joint theory of belief and

[24] This is not just a random coincidence. From the principles of stability theory, one can derive such correspondence results for conditionalization and belief revision *in general*; see Leitgeb 2013a.

degrees of belief is more than just the sum of doxastic logic and subjective probability theory taken together. All of this speaks in favor of the theory, which I thus offer as an alternative to existing theories of belief or acceptance.[25]

References

Alchourrón, Carlos E., Peter Gärdenfors, and David Makinson (AGM). 1985. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *Journal of Symbolic Logic* 50: 510–30.

Arló-Costa, Horacio. 2001. "Bayesian Epistemology and Subjective Conditionals: On the Status of the Export-Import Laws." *Journal of Philosophy* 98, no. 11: 555–98.

Arló-Costa, Horacio, and Rohit Parikh. 2005. "Conditional Probability and Defeasible Inference." *Journal of Philosophical Logic* 34: 97–119.

Benferhat, Salem, Didier Dubois, and Henri Prade. 1997. "Possibilistic and Standard Probabilistic Semantics of Conditional Knowledge." *Journal of Logic and Computation* 9: 873–95.

Dorling, Jon. 1979. "Bayesian Personalism, the Methodology of Scientific Research Programmes, and Duhem's Problem." *Studies in the History and Philosophy of Science Part A* 10, no. 3: 177–87.

Douven, Igor, and Timothy Williamson. 2006. "Generalizing the Lottery Paradox." *British Journal for the Philosophy of Science* 57, no. 4: 755–79.

Foley, Richard. 1993. *Working Without a Net.* Oxford: Oxford University Press.

Gärdenfors, Peter. 1988. *Knowledge in Flux.* Cambridge, MA: MIT Press.

Grove, Adam. 1988. "Two Modellings for Theory Change." *Journal of Philosophical Logic* 17: 157–70.

Hawthorne, James. 2009. "The Lockean Thesis and the Logic of Belief." In Huber and Schmidt-Petri 2009, 49–74.

Hawthorne, James, and Luc Bovens. 1999. "The Preface, the Lottery, and the Logic of Belief." *Mind* 108: 241–64.

Hawthorne, James, and David Makinson. 2007. "The Quantitative/Qualitative Watershed for Rules of Uncertain Inference." *Studia Logica* 86: 247–97.

Hawthorne, John. 2004. *Knowledge and Lotteries.* Oxford: Oxford University Press.

Hempel, Carl G. 1962. "Deductive-Nomological vs Statistical Explanation." In *Minnesota Studies in the Philosophy of Science* 3, ed. H. Feigl and G. Maxwell, 98–169. Minneapolis: University of Minnesota Press.

[25] One can show that this stability theory of belief derives not only from postulates such as P1–P3 of this essay but also from alternative sets of postulates that are plausible independently; see Leitgeb 2013a for more details.

Huber, Franz, and Christoph Schmidt-Petri, eds. 2009. *Degrees of Belief.* Synthese Library 342. Dordrecht: Springer.

Kyburg, Henry E., Jr. 1961. *Probability and the Logic of Rational Belief.* Middletown, CT: Wesleyan University Press.

———. 1970. *Probability and Inductive Logic.* Toronto: MacMillan.

Kyburg, Henry E., Jr., and Cho Man Teng. 2001. *Uncertain Inference.* Cambridge: Cambridge University Press.

Lange, Marc. 2005. "Laws and Their Stability." *Synthese* 144, no. 3: 415–32.

Leitgeb, Hannes. 2013a. "Reducing Belief Simpliciter to Degrees of Belief." *Annals of Pure and Applied Logic* 164: 1338–89.

———. 2013b. "The Stability Theory of Belief. A Summary." (Extended abstract.) In *Logic across the University: Foundations and Application–Proceedings of the Tsinghua Logic Conference, Beijing*, ed. J. van Benthem and F. Liu, 47–54. Volume 47: Studies in Logic. London: College Publications.

Levi, Isaac. 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science.* Cambridge, MA: MIT Press.

———. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability and Chance.* Cambridge, MA: MIT Press.

———. 1984. *Decisions and Revisions.* Cambridge: Cambridge University Press.

Lewis, David K. 1973. *Counterfactuals.* Oxford: Blackwell.

———. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–77.

———. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74, no. 4: 549–67.

Lin, Hanti, and Kevin T. Kelly. 2012a. "A Geo-Logical Solution to the Lottery Paradox." *Synthese* 186, no. 2: 531–75.

———. 2012b. "Propositional Reasoning That Tracks Probabilistic Reasoning." *Journal of Philosophical Logic* 41, no. 6: 957–81.

Loeb, Louis E. 2002. *Stability and Justification in Hume's Treatise.* Oxford: Oxford University Press.

Makinson, David. 1965. "The Paradox of the Preface." *Analysis* 25, no. 6: 205–7.

Quine, Willard van Orman. 1960. *Word and Object.* Cambridge, MA: MIT Press.

Skyrms, Brian. 1977. "Resiliency, Propensities, and Causal Necessity." *Journal of Philosophy* 74, no. 11: 704–13.

———. 1980. *Causal Necessity.* New Haven: Yale University Press.

———. 1984. *Pragmatics and Empiricism.* New Haven: Yale University Press.

Snow, Paul. 1998. "Is Intelligent Belief Really Beyond Logic?" In *Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference*, 430–4. American Association for Artificial Intelligence.

Van Fraassen, Bas C. 1995. "Fine-Grained Opinion, Probability, and the Logic of Full Belief." *Journal of Philosophical Logic* 24: 349–77.

Williamson, Timothy. 1998. "Conditionalizing on Knowledge." *British Journal for the Philosophy of Science* 49: 89–121.

Windschitl, Paul D., and Gary L. Wells. 1998. “The Alternative-Outcomes Effect.” *Journal of Personality and Social Psychology* 75, no. 6: 1411–23.

Woodward, Jim. 2006. “Some Varieties of Robustness.” *Journal of Economic Methodology* 13, no. 2: 219–40.