

# From Probabilities to Categorical Beliefs: Going Beyond Toy Models

Igor Douven\*      Hans Rott†

## Abstract

According to the Lockean thesis, a proposition is believed just in case it is highly probable. While this thesis enjoys strong intuitive support, it is known to conflict with seemingly plausible logical constraints on our beliefs. One way out of this conflict is to make probability 1 a requirement for belief, but most have rejected this option for entailing what they see as an untenable skepticism. Recently, two new solutions to the conflict have been proposed that are alleged to be non-skeptical. We compare these proposals with each other and with the Lockean thesis, in particular with regard to the question of how much we gain by adopting any one of them instead of the probability 1 requirement, that is, of how likely it is that one believes more than the things one is fully certain of.

**1. Introduction.** There has been much recent interest in the question of how to reconcile purely qualitative and numerical representations of belief states. According to the so-called Lockean thesis (Foley [1992a; 1992b; 2009]), a proposition is believed (simpliciter, plainly) just in case its probability has at least a certain threshold value. This proposal works fine if the threshold is set to 1, but most regard this as imposing too stringent a requirement on belief. We are not, and should not be, *that* skeptical: there are many things we believe—and reasonably so—in the absence of full certainty.

On the other hand, the Lockean thesis becomes problematic for different reasons if the threshold is set to a value less than 1. In general, the resulting set of beliefs may then fail to be consistent, or fail to be closed under the relation of logical consequence. Many authors have thought that this would violate logical norms that are indispensable for belief, at least if our beliefs are to qualify as rational.<sup>1</sup>

Recently, Lin and Kelly [2012a; 2012b] and Leitgeb [2013; 2014; 2015; 2017] have proposed alternatives to the Lockean thesis. These authors aim to connect beliefs simpliciter with probabilities in a way that maintains logical closure while allowing beliefs simpliciter to be non-certain. In this paper, we assess these new proposals and compare them with the Lockean thesis with regard to the question of how much is gained by adopting them as opposed to settling for the probability 1 requirement. In other words, the question to be asked for each account is this: how likely is it, on the given account, that you will believe *more* than the things you are fully certain of? The answer to this question can be interpreted as a measure of how much the account improves upon the skepticism entailed by the probability 1 requirement.

Our general approach follows in the footsteps of Rott [2017]. Rott criticized Leitgeb's proposal as being overly skeptical, and argued that the otherwise similar theory of Lin and Kelly does much better in this respect, by calculating the exact likelihoods with which, on

---

\*Sciences, Normes, Décision (CNRS), Sorbonne University, igor.douven@paris-sorbonne.fr.

†Department of Philosophy, University of Regensburg, hans.rott@ur.de.

<sup>1</sup>From here on, we shall simply write “Lockean thesis” instead of “Lockean thesis with a threshold less than 1,” and use the label “probability 1 requirement” for what otherwise could have been called “the Lockean thesis with threshold 1.”

these proposals as well as on the Lockean thesis, a random probability distribution would give rise to an *informative* belief set, that is, a belief set containing non-certain beliefs.

However, Rott [2017] has an important limitation, in that the concrete cases considered in that paper concern only probability distributions over three and four worlds. In fact, it shares this limitation with Leitgeb's and Lin and Kelly's work, as these authors illustrate their claims with three-world models only. Toy models can be invaluable for developing intuitions and achieving a basic understanding of the mechanisms or processes that one is interested in. But by focusing exclusively on such models, we may be missing problematic features that only emerge in more realistic models. In the present case, we may not be able to draw any firm conclusions about how beliefs and probabilities are interconnected on the basis of how well (or poorly) any proposed bridge principles do in the context of models involving only a handful of worlds.

If three-world and four-world models are toy models, what would be realistic numbers of worlds to consider? Chi and Ohlsson [2005, 372] estimate that "a person's declarative knowledge base eventually approximates 1 million pieces of knowledge." It is not clear from the context whether they really mean "knowledge base" or only "belief base," and whether one "piece of knowledge" would correspond to one belief. Suppose they mean "belief base." If the beliefs are presumed to be logically independent, then we would need approximately  $2^{10^6} \approx 10^{301,030}$  worlds to represent our doxastic states.<sup>2</sup>

Our main interest is in how much "less" skeptical than the probability 1 requirement the Lockean thesis as well as Leitgeb's and Lin and Kelly's accounts are. Thus, we are interested in the likelihood of our having a personal probability distribution that gives rise to beliefs other than those we hold because we can rule out some worlds with certainty. And it is reasonable to assume that, of all the worlds that were just said to be needed to represent our doxastic states, a great number *can* be ruled out on the basis of the evidence in our possession. It is equally reasonable to assume, however, that the number of those that remain, in that they receive some positive probability, will be nowhere near the order of three or four. We continue the work begun in Rott [2017] by asking the question he asked, but now also considering probability models encompassing up to hundreds and even thousands of possible worlds.

Before we get started, some words about threshold values. Most who have considered the question of how beliefs and probabilities hang together agree that the threshold must be *at least* as high as .5, while for the reason mentioned, a threshold of 1 is generally regarded as being too demanding. Throughout the following, we will consider a number of different threshold values in the range from .5 to 1. Although we want to remain largely noncommittal on where exactly in that range the threshold should be located, we do note that, according to many authors, the threshold for rational belief is a *good deal* higher than .5. Some authors advocate very specific numbers, for a priori reasons. For example, Shear and Fitelson [2018] argue that the Lockean threshold should be at least the inverse of the golden ratio, that is,  $\sqrt{5}-1/2 = .6180$ . Easwaran [2016] gives reasons why an agent should believe all propositions  $A$  such that  $\Pr(A) > \omega/\rho+\omega$  and should believe no proposition  $B$  such that  $\Pr(B) < \omega/\rho+\omega$ , where  $\rho$  is a (constant) bonus for every true belief and  $\omega$  is a (constant) penalty for every false belief she holds. If we combine Easwaran's idea with Pruss' [2012] argument that  $\omega$  should equal  $\rho/\ln(4)-1$ , then we obtain that the Lockean threshold should be at least  $1/\ln(4) = .7213$ . Kaplan [1981, 308], Moser and Tlumak [1985, 128], and Kyburg [1990, 64] assume a threshold of .9, while Foley [1992a, 113] goes further still by postulating .99 as a reasonable value. Interestingly, .9 has been recommended by legal scholars (e.g., Kagehiro and Stanton [1985])

---

<sup>2</sup>We have more to say on the number of possibilities in Section 6.1.

as an explication of the notion of being beyond reasonable doubt, which is supposed to be a criterion for a conviction in the practice of criminal law.<sup>3</sup>

In Section 2, we state the various theses of interest in a formally precise way, and we summarize what these theses are already known to imply concerning skepticism. In the same section, we also explain the main mathematical technique we shall use for investigating possible skeptical implications in probability models much larger than the three-world and four-world models considered by previous authors. Sections 3–5 then present our formal results concerning the Lockean thesis, Lin and Kelly’s proposal, and Leitgeb’s proposal, respectively. Section 6 addresses two objections that some may want to level against our comparison of these three accounts.

**2. Setting the stage: Belief rules and Monte Carlo integration.** Let  $W = \{w_1, \dots, w_n\}$  be a finite set of mutually exclusive and jointly exhaustive possibilities (or possible worlds), and let  $\mathbf{p} = (p_1, \dots, p_n)$  be a probability distribution on  $W$  such that  $p_i$  is the probability of  $w_i$ . A subset  $A$  of  $W$  is called a *proposition (over  $W$ )*, and for any such proposition we define  $\Pr(A) = \sum\{p_i : w_i \in A\}$  (using the convention  $\sum \emptyset = 0$ ). Finally, let  $\theta$  be the belief threshold, where it is presupposed that  $.5 \leq \theta < 1$ . We are now going to introduce three *belief rules* and ask under what conditions each of these rules yields an *informative* belief set, by which we mean a belief set that contains at least one proposition with imperfect probability (i.e., with probability less than 1).

The most simple and natural belief rule for the threshold  $\theta$  is the one already mentioned above:

LOCKEAN THESIS (LT): The rational set of beliefs with respect to a probability function  $\Pr$  and a threshold  $\theta$  is  $\text{Bel} = \{A : \Pr(A) > \theta\}$ .

Thus, on LT,  $\mathbf{p}$  gives rise to an informative belief set just in case there is a proposition with a probability strictly between  $\theta$  and 1, which is true just in case

$$\exists i: 0 < p_i < 1 - \theta. \quad (\text{L})$$

In LT, absolute probabilities matter. In contrast, the belief rule proposed by Lin and Kelly—the so-called odds-threshold rule—eliminates a certain set of *relatively* improbable possibilities:

ODDS-THRESHOLD RULE (OR): The rational set of beliefs with respect to a probability function  $\Pr$  and a threshold  $\theta$  is  $\text{Bel} = \{A : X_\theta \subseteq A\}$ , where

$$X_\theta = \left\{ w_i \in W : \frac{p_i}{\max(\{p_j : w_j \in W\})} \geq \frac{1 - \theta}{\theta} \right\}.$$

This formulation of OR is not quite the same as that in Lin and Kelly [2012a, 537], where a threshold term of the form “ $1 - \theta$ ” is used, or that in Lin and Kelly [2012b, 970–971], where the form “ $1/\theta$ ” is used. Our presentation here involves a natural transformation of the threshold parameter that keeps  $\theta$  between .5 and 1 and permits a perfect comparability with the rules of Locke and Leitgeb.<sup>4</sup>

On OR,  $\mathbf{p}$  generates an informative belief set just in case

$$\exists i: 0 < p_i < \frac{1 - \theta}{\theta} \cdot \max(\{p_j : w_j \in W\}). \quad (\text{O})$$

<sup>3</sup>However, Magnussen et al. [2014] report that in actuality judges tend to convict a defendant as soon as they are around 80 percent certain of the defendant’s guilt.

<sup>4</sup>For details, see Rott [2017, Sect. 4]. It will turn out that the precise specification of the threshold value in OR loses much of its importance as the number of possibilities grows. See Table 3 below.

Finally, here is the belief rule that Leitgeb, in his stability account of rational belief, dubs “the Humean thesis” for the threshold  $\theta$ :<sup>5</sup>

HUMEAN THESIS (HT): The rational set of beliefs with respect to a probability function  $\Pr$  and a threshold  $\theta$  is  $\text{Bel} = \{A: \Pr(A|B) > \theta \text{ for all propositions } B \text{ such that } \neg B \notin \text{Bel} \text{ and } \Pr(B) \neq 0\}$ .<sup>6</sup>

In this version of HT, just as in the above version of LT, only thresholds less than 1 make sense. Leitgeb [2013, 1363] has shown that on HT,  $\Pr$  (or in our context equivalently,  $\mathbf{p}$ ) gives rise to an informative belief set just in case  $\Pr$  satisfies the following criterion:

$$\exists \text{ non-empty } S \subset W: 0 < \Pr(W \setminus S) < \frac{1 - \theta}{\theta} \cdot \min(\{p_i: w_i \in S\}). \quad (\text{H})$$

A proposition  $S$  satisfying H is called  $\theta$ -stable (or simply *stable*, when  $\theta$  is given) with respect to  $\Pr$ ; if in addition  $\Pr(S) < 1$ , it is called *informatively* ( $\theta$ -)stable.

We will use the same values for the parameter  $\theta$  in L and H. One justification for this is that the considerations bearing on the specific value of the threshold, given in the introduction, seem to pertain to the Lockean and the Humean threshold equally.<sup>7</sup> Given these identifications, it is easy to verify that H implies O, which in turn implies L, and that the converse implications do not hold. To see that H implies O, suppose  $S \subset W$  satisfies H and let  $p_{(1)}$  be the probability assigned to a least probable world in  $W \setminus S$ . Then  $p_{(1)} < (1-\theta)/\theta \cdot \min(\{p_i: w_i \in S\})$  and so, because  $\min(\{p_i: w_i \in S\}) \leq \max(\{p_i: w_i \in W\})$ , also  $p_{(1)} < (1-\theta)/\theta \cdot \max(\{p_i: w_i \in W\})$ , thereby satisfying O. To see that O implies L, assume a non-decreasing ordering of the  $p_i$ 's and suppose for reductio that there is a  $j < n$  such that  $0 < p_j < (1-\theta)/\theta \cdot p_n$  but not  $p_j < 1 - \theta$ . From this it follows that  $1 - \theta < (1-\theta)/\theta \cdot p_n$ , and thus  $\theta < p_n$ . On the other hand, since  $1 - \theta \leq p_j$  and  $p_j \leq 1 - p_n$  (notice that  $j \neq n$  and the  $p_i$ 's sum to 1), it follows that  $p_n \leq \theta$ , and we have a contradiction. That the converse implications do not hold can be seen already for  $n = 3$  and, for example,  $\theta = .8$ : the probability distribution  $\mathbf{p} = (1/9, 3/9, 5/9)$  satisfies O but not H, while the probability distribution  $\mathbf{p} = (1/6, 2/6, 3/6)$  satisfies L but not O.

To illustrate the above belief rules, we first note that vectors in the standard unit simplex (or probability simplex) of dimensionality  $n - 1$  can be interpreted as probability distributions on a set  $W$  of  $n$  possibilities, with the  $i$ -th vector component representing the probability of the  $i$ -th possibility (de Finetti [1962]). Thus, in particular, a probability distribution on  $W = \{w_1, w_2, w_3\}$  can be represented by a point in the so-called unit 2-simplex, that is, the two-dimensional simplex whose edges all have length 1. For any given combination of a belief rule and a value of  $\theta$ , we can then ask which regions in that simplex represent probability distributions giving rise to informative belief sets. Figure 1 answers that question for all three rules, assuming  $\theta = .8$  in each case.

Given that we are interested in measuring the degree to which LT, OR, and HT improve upon the skeptical solution, by countenancing belief in propositions whose probability is less than 1, all three rules can actually be simplified. This is because worlds with probability 0 can be ignored. To see why, note that given a probability distribution  $\mathbf{p} = (p_1, \dots, p_n)$  on  $W$ , if  $W' = \{w_i \in W: p_i > 0\}$  and  $|W'| = m < n = |W|$ , then  $\mathbf{p}$  represents a point in the unit  $(m - 1)$ -simplex spanned by the worlds in  $W'$ , which is a proper subspace of the unit  $(n - 1)$ -simplex. The process of determining whether  $\mathbf{p}$  gives rise to any non-certain

<sup>5</sup>See Leitgeb [2015, 152, 163; 2017, 76, 86–87]; notation adapted for uniformity of reading.

<sup>6</sup>Leitgeb points out that his formulation of the Humean thesis specifies sufficient and necessary conditions for the rationality of a belief set, but does not define a unique rational belief set for a given probability function  $\Pr$ . Uniqueness is guaranteed, however, if the threshold value  $\theta$  is given as well, and our formulation of HT fixes the threshold.

<sup>7</sup>Here we follow Rott [2017] again. In forthcoming work, we discuss the identification of the Lockean and Humean thresholds more extensively. Concerning the threshold value in O, see again note 4.



Figure 1: Points in the big triangle (the unit 2-simplex) represent probability distributions on  $W = \{w_1, w_2, w_3\}$ . Assuming  $\theta = .8$ , probability distributions giving rise to informative belief sets according to each of LT, OR, and HT fall in the dark gray region; those that give rise to informative belief sets according to LT and OR, but not according to HT, fall in the medium gray region; and those giving rise to informative belief sets according to LT but not according to the other rules fall in the light gray region. Probability distributions that do not give rise to informative belief sets on any rule fall in the white region.

beliefs is then strictly to be carried out in that subspace. Suppose, for example, that an agent is able to rule out with certainty all but three possible worlds. Then she believes with certainty the proposition consisting of those three worlds. Whether she holds any beliefs with non-certainty, according to one or the other belief rule, is now entirely a matter of into which region of the unit 2-simplex spanned by the three remaining worlds the point representing her probabilities falls.

Given a probability distribution  $\mathbf{p} = (p_1, p_2, \dots, p_n)$ , let  $(p_{(1)}, p_{(2)}, \dots, p_{(n)})$  be an ordering of the probabilities in  $\mathbf{p}$  such that  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ . In view of the foregoing observation, we can cut off from this ordering the probability 0 worlds and keep just the vector  $(p_{(k)}, p_{(k+1)}, \dots, p_{(n)})$ , where  $1 \leq k \leq n$ ,  $p_{(i)} = 0$  for  $i = 1, \dots, k-1$ , and  $p_{(i)} > 0$  for  $i = k, \dots, n$ . This allows us to restate the criteria L, O, and H more simply as the following conditions on probability distributions:

$$p_{(k)} < 1 - \theta; \tag{L^*}$$

$$p_{(k)} < \frac{1 - \theta}{\theta} \cdot p_{(n)}; \tag{O^*}$$

$$\exists i_{k \leq i < n}: p_{(k)} + \dots + p_{(i)} < \frac{1 - \theta}{\theta} \cdot p_{(i+1)}. \tag{H^*}$$

If and only if a probability distribution satisfies one of these conditions, then it gives rise to an informative belief set according to the corresponding rule.

The question that Rott [2017] asked, and that we will ask again, is how likely it is, on the above rules, to find a probability distribution that gives rise to an informative belief set. Unsurprisingly, and as Rott also shows, the answer is going to depend both on the value of  $\theta$  and on the size of the support of the distribution (the number of possibilities with non-zero probability). We understand this question to be asking how likely it is, given a support of size  $n$ , that an arbitrarily chosen distribution  $\mathbf{p}$  on this support gives rise to an informative belief set, where  $\mathbf{p}$  being arbitrary is construed as  $\mathbf{p}$  being drawn randomly from a Dirichlet distribution  $\text{Dir}(\mathbf{1})$ , with  $\mathbf{1}$  a vector of  $n$  1s. To clarify: a Dirichlet distribution is a probability distribution over probability distributions, or, as we shall say, a *likelihood* distribution over

Table 1: Likelihood of a probability distribution giving rise to an informative belief set for the belief rules at issue and for various values of  $\theta$ , for the cases of three and four worlds (from Rott [2017], precisified).

no. of worlds	rule	$\theta$					
		.6	.7	.8	.9	.95	.99
3	LT	1	.9900	.8400	.5100	.2775	.0591
	OR	.9643	.8552	.6667	.3876	.2088	.0443
	HT	.8460	.6374	.4053	.1819	.0837	.0154
4	LT	1	1	.9920	.7840	.4880	.1153
	OR	.9939	.9500	.8220	.5429	.3144	.0711
	HT	.8417	.6139	.3712	.1594	.0730	.0136

probability distributions. The intuitive way to think of a uniform such distribution is that all points in the unit  $(n - 1)$ -simplex have the same likelihood of being drawn in a sampling from that simplex. The uniformity assumption is justified to the extent that, a priori, it is no more likely that a rational person’s probabilities are represented by a point in one particular region of a simplex of the right dimension rather than by a point in a different region of that simplex.

Rott [2017] answers the foregoing question for the cases of three and four worlds; the results are given in Table 1. The likelihoods in the table are calculated by determining those regions in the unit 2- and unit 3-simplex, respectively, which represent distributions that satisfy the condition imposed by the given rule, for the various values of  $\theta$ : the likelihoods in the table correspond to the normalized Lebesgue measures of those regions (area relative to total area in the three-world case, volume relative to total volume in the four-world case).<sup>8</sup> However, once we go beyond the 3-simplex, the regions over which we must integrate to find the likelihoods corresponding to those in Table 1 quickly become very complicated, and their measures can no longer be analytically calculated. They can still be reliably approximated, however.

While working on the development of nuclear weapons at Los Alamos Scientific Laboratory in the 1940s, the physicist Enrico Fermi and the mathematicians Nicholas Metropolis, Stanisław Ulam, and John von Neumann invented the technique of Monte Carlo sampling, primarily for investigating the properties of elementary particles, but realizing that the technique could also be used for evaluating complicated and multidimensional integrals (Metropolis and Ulam [1949]). Meanwhile, Monte Carlo integration has become popular in many areas of science as a method of approximating the values of integrals that are difficult or even impossible to work out analytically.

In this paper, the technique will be used to approximate the likelihood that an arbitrary probability distribution gives rise to an informative belief set for models with many more worlds than four. This is not to say that Monte Carlo integration will allow us to (approximately) obtain the said likelihood for any model, no matter how large: the technique comes with a computational cost, which can become very steep. There is also the problem known as “the curse of dimensionality” (Bellman [1957]), which in our case is the problem that as the dimensionality of the simplexes grows, one has to increase the sample of probability distributions (the points sampled from the simplex) in order to maintain numerical stability of the integration, adding to the computational costs. Nevertheless, building on the technique

<sup>8</sup>If  $\theta = .5$ , then, as one easily verifies, the likelihood of finding a probability distribution that gives rise to an informative belief set equals 1 for each of LT, OR, and HT, however many worlds are considered. This does not mean that every distribution gives rise to a such a belief set: for instance, the flat distribution does not.

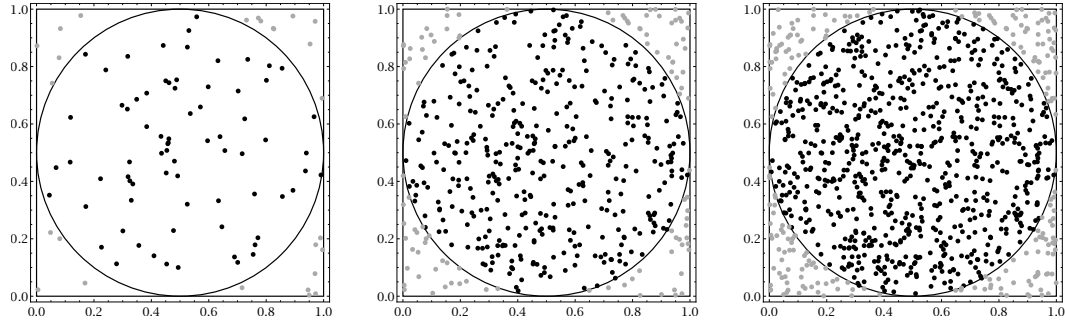


Figure 2: Three attempts to approximate the value of  $\pi$ , one by throwing 100 pebbles into a square (left panel), one by throwing 500 pebbles into the square (middle panel), and one by throwing 1,000 pebbles into the square (right panel). Hits are shown in black, misses in gray.

of Monte Carlo integration, we will be able to obtain estimates for fairly large numbers of worlds.

Because the technique will be so important in the following, and because not all readers will be familiar with it, we give a brief explanation, by means of an example that is frequently used to introduce Monte Carlo integration (see, e.g., Krauth [2006, Ch. 1]). Suppose we knew that the area of a circle is given by  $\pi r^2$  but did not know the value of  $\pi$ . Here is how we could estimate the latter: draw a  $1 \times 1$  square in the sand, draw the square's incircle, and start throwing pebbles into the square randomly (meaning that, for a pebble, any point in the square is as likely to be its landing place as any other). Now count all pebbles that landed in the circle (the hits) and divide their number by the total number of pebbles that were thrown (hits plus misses). Given that  $r = 0.5$  and so  $r^2 = 0.25$ , we have to multiply the result by 4 to obtain the desired estimate.

Figure 2 illustrates the idea, the panels showing, from left to right, the results of, respectively, 100, 500, and 1,000 pebbles having been thrown into the square (in reality, this was of course simulated by means of our computer's built-in random number generator). Using the aforementioned procedure, we find an approximate value of  $\pi$  of 3.28 in the first case, of 3.21 in the second, and of 3.16 in the third. To obtain a still better estimate of  $\pi$  ( $\approx 3.14$ ), we must either find more pebbles to throw or repeat the foregoing procedure a number of times and average the results.

In the following, we use essentially the same method to calculate the likelihoods of obtaining informative belief sets, for each of LT, OR, and HT, and given various values for the threshold. In each case this amounts to approximating the normalized Lebesgue measure of certain regions in probability simplexes.

**3. The Lockean belief rule.** To show in some detail how Monte Carlo integration can help to approximate the likelihood that an arbitrary probability distribution gives rise to an informative belief set, assuming LT, and later also OR and HT, it is easiest to first consider the cases of three and four possibilities, respectively, which can still be visualized. Just as we did with the square and the circle above, we now throw pebbles into the unit 2-simplex, let the computer check which of those lie in the region corresponding to satisfying condition  $L^*$  for a given value of  $\theta$  (see again Figure 1), and color black those pebbles that do (the hits) and gray those that do not (the misses). The left and middle panel of Figure 3 depict the results for  $\theta = .8$  and  $\theta = .9$ , respectively. We do the same for the unit 3-simplex. The right panel gives an impression of the results for  $\theta = .9$  in that space. (Obviously, in this case the pebbles-throwing metaphor no longer makes sense, as all pebbles would end up

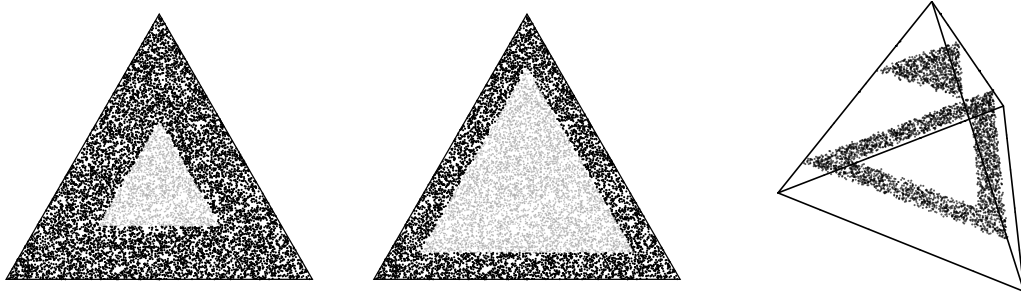


Figure 3: Left and middle panel showing hits (black) and misses (gray) among 10,000 probability distributions randomly sampled from a uniform Dirichlet distribution on the unit 2-simplex, for threshold values .8 and .9, respectively, with hits and misses defined according to LT. Right panel showing two arbitrary slices of plot showing hits (in black) among 100,000 probability distributions randomly sampled from a uniform Dirichlet distribution on the unit 3-simplex, assuming  $\theta = .9$ . (It is only by showing slices that one can give an impression of the hit and miss regions, and in particular of the fact that much of the interior does not belong to the hit region; misses are not colored, to further enhance visibility of the structure.)

lying on the bottom facet of the simplex.) The likelihoods for these cases are now simply found by dividing the number of hits (an approximation of the Lebesgue measure of the region in which the probability distributions satisfying  $L^*$  are to be found) by the totality of pebbles, that is, the sum of hits and misses (an approximation of the Lebesgue measure of the simplex as a whole).

More generally, assume that we have  $n$  possibilities (which, recall, can all be taken to have positive probability, for the reason mentioned above), where  $\Delta^n$  is the unit  $n$ -simplex,

Table 2: Likelihood of a distribution giving rise to an informative belief set for various values of  $\theta$ , determined via Monte Carlo integration, applying the Lockean thesis LT.

no. of worlds	$\theta$					
	.6	.7	.8	.9	.95	.99
3	1	.9902	.8401	.5102	.2773	.0593
4	1	1	.9920	.7845	.4874	.1152
5	1	1	1	.9376	.6843	.1860
6	1	1	1	.9899	.8313	.2659
7	1	1	1	.9993	.9241	.3536
8	1	1	1	1	.9721	.4418
9	1	1	1	1	.9916	.5303
10	1	1	1	1	.9980	.6124
12	1	1	1	1	1	.7542
14	1	1	1	1	1	.8592
16	1	1	1	1	1	.9269
18	1	1	1	1	1	.9658
20	1	1	1	1	1	.9855
25	1	1	1	1	1	.9990
30	1	1	1	1	1	1
50	1	1	1	1	1	1
100	1	1	1	1	1	1
250	1	1	1	1	1	1
500	1	1	1	1	1	1
1000	1	1	1	1	1	1



and where  $L_\theta^n$  is the region in  $\Delta^n$  satisfying  $L^*$  on the assumption that the threshold value is  $\theta$ . Then the above procedure helps to approximate the likelihood that a probability distribution gives rise to an informative belief set, applying LT with threshold  $\theta$ , via the following approximate equation:

$$\frac{\int \cdots \int_{L_\theta^{n-1}} dV}{\int \cdots \int_{\Delta^{n-1}} dV} \approx \frac{\# \text{ of hits}}{\# \text{ of hits} + \# \text{ of misses}}. \quad (\text{AEQ})$$

As explained before, hits and misses refer to probability distributions  $\mathbf{p}$  randomly drawn from the Dir( $\mathbf{1}$ ) distribution, with  $\mathbf{1}$  being the vector of 1's of length  $n$ , and with hits being defined as the  $\mathbf{p}$  with  $\mathbf{p} \in L_\theta^{n-1}$  and misses as the  $\mathbf{p}$  with  $\mathbf{p} \notin L_\theta^{n-1}$ . As also explained, the approximation is expected to be the better the more probability distributions we sample, though obviously there is the issue of computational costs to be considered.

The algorithm for carrying out the Monte Carlo integration for LT was written in the Julia Language (Bezanson et al. [2017]). Table 2 reports values for the various thresholds that were obtained using this algorithm. Here, as well as in the Monte Carlo procedures carried out for OR (see Section 4), computations were always based on a sampling of 1,000,000 probability distributions. Comparison of the values for the  $n = 3$  and  $n = 4$  cases with the corresponding analytically determined values in Table 1 gives reason to believe that the approximations are within reasonable bounds. We are reporting likelihoods for up to 1,000 worlds, but we could easily go on for a while, for the computational cost for LT turns out to be quite low. However, given that the trend is so clear already from the shown outcomes, there would be no point in continuing.

A still more compelling reason for not carrying the Monte Carlo method any further for LT is that it can be seen quite simply that the likelihood of finding an informative belief set, given LT with any admissible value of  $\theta$ , is bound to be 1 for large enough models. For observe that, whatever the exact value of  $\theta$ , there will be a number  $n$  of doxastically possible worlds such that  $1/n < 1 - \theta$ , and given this or any larger number of worlds, there must be at least one world with probability less than  $1 - \theta$ .

**4. The odds-threshold rule.** To obtain the likelihoods of finding an informative belief set when applying Lin and Kelly's rule OR, we proceed in the same way as in the case of LT. Specifically, we again use the approximate equation (AEQ), although now of course with the integral in the numerator ranging over the region defined by the condition imposed by  $O^*$ . Figure 4 visualizes the results for  $\Delta^2$  and  $\Delta^3$ .

The computational costs are higher than for calculating the likelihoods for LT—for OR, the maximum of each sampled distribution has to be selected, and also has to be scaled—but only slightly so, and here we went on to calculate likelihoods for cases all the way up to 25,000 possibilities. Table 3 only reports likelihoods for up to the case of 1,000 possibilities, which suffices to make manifest the general pattern we found: the likelihoods quickly reach the value of 1, for all threshold values, and then remain there. It may be noted that the likelihoods do not go to 1 quite as fast as in the case of LT, but that is immaterial, given that, even with 1,000 possibilities, we are still not anywhere near a realistic number of possibilities (and the rationale behind the current endeavor is the idea that how the rules differ in toy models is of little interest). Note, incidentally, that comparing the results for  $n = 3$  and  $n = 4$  with the corresponding likelihoods in Table 1 again shows that the Monte Carlo integration yields excellent approximations.<sup>9</sup>

<sup>9</sup>We also ran simulations for a combination of the odds-threshold rule OR with a high-probability requirement in the manner of Locke, or more precisely, for the rule that puts  $\text{Bel} = \{A: X_\theta \subseteq A\}$  if  $\Pr(X_\theta) > \theta$ , and  $\text{Bel} = \{W\}$  otherwise. (Thanks to Gerhard Schurz for bringing up this idea.) This rule behaves oddly and is in general much too skeptical. It is particularly skeptical for comparatively *low* values of  $\theta$ . For such values, OR excludes so many worlds from the set  $X_\theta$  that  $X_\theta$  mostly ends up having a rather low probability—one that does not surpass  $\theta$ .

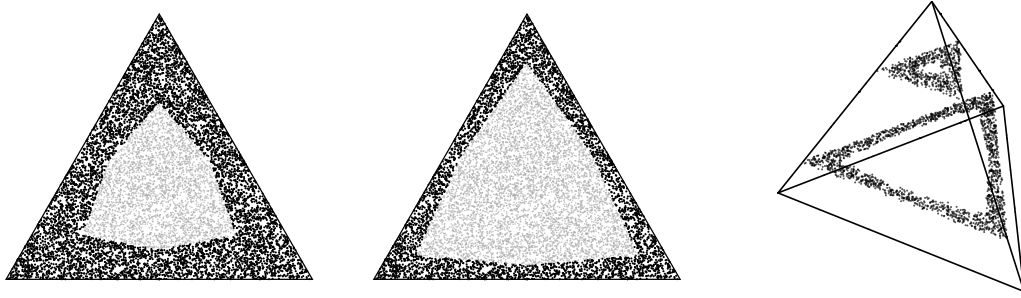


Figure 4: Same results as shown in Figure 3, but now for OR. (For explanation, see the caption of Figure 3.)

We had a quick way of seeing why, for large numbers of worlds, the likelihood of finding an informative belief set when applying LT had to be 1 for all reasonable values of  $\theta$ . To our knowledge, no similarly easy argument is available in the case of OR. We can nevertheless gain some understanding of why the likelihoods develop as suggested by the integration results reported here, by using (i) information about the uniform Dirichlet distribution, which we are assuming in all our Monte Carlo integrations; (ii) insights from order statistics; and (iii) techniques for transforming probability distributions.

It is known that if  $\mathbf{p} = (p_1, \dots, p_n) \sim \text{Dir}(\mathbf{1})$ , then  $p_i \sim \text{Beta}(1, n - 1)$ , for all  $i$ , where  $\text{Beta}(1, n - 1)$  is the beta distribution with parameters 1 and  $n - 1$ .<sup>10</sup>

Table 3: Likelihood of a distribution giving rise to an informative belief set for various values of  $\theta$ , determined via Monte Carlo integration, applying the odds-threshold rule OR.

no. of worlds	$\theta$					
	.6	.7	.8	.9	.95	.99
3	.9641	.8554	.6662	.3868	.2087	.0440
4	.9939	.9498	.8216	.5428	.3140	.0710
5	.9990	.9830	.9070	.6644	.4120	.0990
6	.9998	.9943	.9524	.7554	.4995	.1279
7	1	.9983	.9760	.8245	.5752	.1567
8	1	.9994	.9879	.8749	.6422	.1870
9	1	.9998	.9939	.9110	.7001	.2169
10	1	.9999	.9970	.9374	.7493	.2447
12	1	1	.9992	.9694	.8259	.3024
14	1	1	.9998	.9849	.8803	.3553
16	1	1	1	.9926	.9182	.4068
18	1	1	1	.9965	.9442	.4555
20	1	1	1	.9982	.9620	.4993
25	1	1	1	.9998	.9862	.5993
30	1	1	1	1	.9949	.6823
50	1	1	1	1	.9999	.8785
100	1	1	1	1	1	.9910
250	1	1	1	1	1	1
500	1	1	1	1	1	1
1000	1	1	1	1	1	1

<sup>10</sup>See Kotz, Balakrishnan, and Johnson [2000, Ch. 49] on Dirichlet distributions and Johnson, Kotz, and Balakrishnan [1995, Ch. 25] on beta distributions. The formulas for the probability density function (pdf) and the cumulative distribution function (cdf) of a  $\text{Beta}(1, n - 1)$  distribution are given in the Appendix; the details do not matter here.

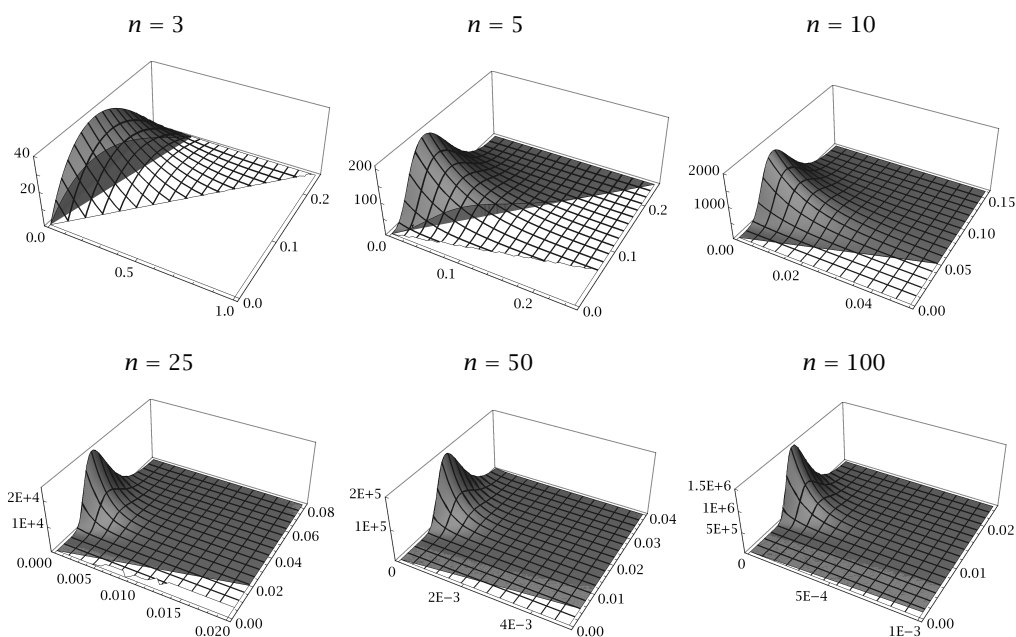


Figure 5: For various values of  $n$ , plots of the joint pdfs of the minimum and the scaled maximum of  $n$  independent and identically distributed random variables all following a beta distribution with parameters 1 and  $n - 1$ . In gray, the part of the distributions lying in the region of the domain where the minimum is smaller than the scaled maximum. (The scaling of the maximum is for  $\theta = .8$ .)

Order statistics is the branch of statistics which studies the distributions of least elements, second least elements, and so on, of statistical samples. Results in this area almost exclusively concern samples of random variables that are independent and identically distributed (David and Nagaraja [2003]). While in our case the random variables (the  $p_i$ 's) are identically distributed—they all come from the same beta distribution, as said—they are obviously not independent, given that they must sum to 1: if the values of  $n - 1$  elements of  $\mathbf{p} = (p_1, \dots, p_n) \sim \text{Dir}(\mathbf{1})$  are known, then so is the value of the remaining one. But we can still approximate the distribution of the various order statistics—the minimum, the maximum, their joint distribution, and so on—by assuming independence, given that the  $p_i$ 's are “quasi-independent,” in that each  $p_i$  is independent of the *normalized* vector of the  $p_j$ 's with  $j \neq i$ .<sup>11</sup> Intuitively, to be informed about the value of one element of  $\mathbf{p} = (p_1, \dots, p_n)$  tells us *near to nothing* about any of the remaining values if  $n$  is sufficiently large. Where  $\mathbf{X} = (X_1, \dots, X_n)$  with the  $X_i$  independent and identically distributed with cumulative distribution function (cdf)  $F(x)$  and probability density function (pdf)  $f(x)$ , order statistics gives us the pdf of the  $j$ -th order statistic  $X_{(j)}$ , for any  $j$ , and also the joint pdf of  $X_{(j)}$  and  $X_{(k)}$ , for any  $j$  and  $k$  (see Arnold, Balakrishnan, and Nagaraja [2008, 10] and also the Appendix). In particular, it gives the joint distribution of the minimum,  $X_{(1)}$ , and maximum,  $X_{(n)}$ , of  $\mathbf{X}$ .

Finally, given the joint pdf of the minimum and the maximum, we can find the *transformed* joint distribution that scales the maximum by a given constant, using standard techniques as for instance described in Rice [2007, Ch. 3.6].

With this machinery in place, the joint pdf of the two random variables that matter to the evaluation of OR, to wit, the minimum of  $\mathbf{p} = (p_1, \dots, p_n) \sim \text{Dir}(\mathbf{1})$  and the maximum of

<sup>11</sup>In the statistical literature, this property of quasi-independence is commonly referred to as *neutrality*; see Connor and Mosimann [1969].

that distribution scaled by  $r_\theta := (1-\theta)/\theta$ , can be calculated to be

$$\frac{n}{r_\theta} \cdot (n-1)^3 \left( (1-x)^{n-1} - \left(1 - \frac{y}{r_\theta}\right)^{n-1} \right)^{n-2} (1-x)^{n-2} \left(1 - \frac{y}{r_\theta}\right)^{n-2}.$$

Figure 5 plots the above as a function of  $x$  and  $y$ , for various values of  $n$  ranging from 3 to 100, and assuming  $\theta = .8$ . From this figure, we immediately understand the pattern observed in Table 3. In the plots, the probability mass lying on the region in which the minimum is smaller than the scaled maximum is shown in gray. We see that, as  $n$  gets larger, more and more of the probability mass gets shifted to that region. From  $n = 10$  onwards, basically no probability mass is to be found outside that region, corresponding to the fact that when applying OR, the likelihood of finding an informative belief set is basically 1 for  $n > 10$ , as can be seen in Table 3. One gets the same pattern for other values of  $\theta$ .

**5. The stability account.** As we did for LT and OR, we use the approximate equation (AEQ) and the technique of Monte Carlo integration to determine the likelihoods of finding an informative belief set when applying HT. A visualization of the results for the three-world and four-world cases is given in Figure 6. Computational costs are considerably higher than in the cases of LT and OR. This does not come as a big surprise. In the case of LT, we just have to check, for each distribution, whether its minimum is larger than  $1 - \theta$ . For OR, there is a little more work to be done—scale the maximum, then compare the result with the minimum—but that is still no impediment to obtaining Monte Carlo results for up to thousands of possibilities. For HT, the situation is very different. Given a distribution  $\mathbf{p} = (p_1, \dots, p_n)$  with  $p_i > 0$  for all  $i$ , the fastest algorithm we can come up with to determine whether  $\mathbf{p}$  gives rise to an informative belief set in the sense of HT first checks whether the smallest probability,  $p_{(1)}$ , is smaller than  $(1-\theta)/\theta \cdot p_{(2)}$ , then whether the sum of  $p_{(1)}$  and  $p_{(2)}$  is smaller than  $(1-\theta)/\theta \cdot p_{(3)}$ , and so on. Naturally, once the algorithm finds that  $\sum_i p_{(i)} < (1-\theta)/\theta \cdot p_{(i+1)}$  for some  $i < n - 1$ , it can terminate. But, as will be seen, often there is no such stopping point.

We can still calculate values for up to 1,000 worlds; see Table 4. For up to  $n = 50$ , values are again based on a sampling of 1,000,000 probability distributions; in view of the curse of dimensionality, values for  $n \in \{100, 250, 500\}$  are based on a sampling of 5,000,000 probability distributions, and for  $n = 1,000$  even 10,000,000 distributions were used. Comparing the results for the three-world and four-world cases with the analytical results for those cases reported in Table 1 shows that here, too, we obtain excellent approximations. Even with 1,000 worlds, we may be far removed from the number of worlds that, realistically speaking, is required to represent people’s doxastic states. But at least we can discern a

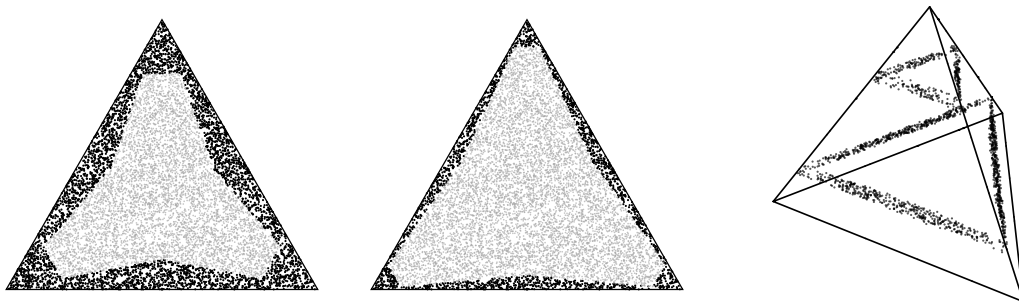


Figure 6: Same as Figures 3 and 4 but now applying the Humean thesis HT. (For explanation, see the caption of Figure 3.)

Table 4: Likelihood of a distribution giving rise to an informative belief set for various values of  $\theta$ , determined via Monte Carlo integration, applying the Humean thesis HT.

no. of worlds	$\theta$					
	.6	.7	.8	.9	.95	.99
3	.8457	.6368	.4050	.1823	.0838	.0153
4	.8418	.6143	.3709	.1593	.0730	.0137
5	.8250	.5820	.3419	.1468	.0679	.0125
6	.8059	.5559	.3248	.1404	.0650	.0122
7	.7884	.5378	.3143	.1359	.0632	.0120
8	.7766	.5256	.3074	.1328	.0617	.0118
9	.7673	.5170	.3017	.1309	.0609	.0112
10	.7603	.5117	.2989	.1295	.0601	.0113
12	.7528	.5033	.2932	.1277	.0585	.0112
14	.7471	.4986	.2891	.1252	.0576	.0109
16	.7434	.4951	.2871	.1240	.0575	.0109
18	.7404	.4915	.2848	.1226	.0569	.0107
20	.7387	.4894	.2835	.1223	.0565	.0107
25	.7350	.4855	.2808	.1213	.0562	.0105
30	.7322	.4827	.2793	.1200	.0555	.0104
50	.7269	.4787	.2763	.1191	.0549	.0103
100	.7241	.4742	.2727	.1173	.0544	.0103
250	.7215	.4723	.2713	.1166	.0541	.0102
500	.7214	.4715	.2709	.1165	.0540	.0102
1000	.7207	.4711	.2707	.1164	.0538	.0102
$(1-\theta)/\theta$	.6667	.4286	.2500	.1111	.0526	.0101

pattern now, and this may allow us to make a reliable guess as to what we can expect to see when more worlds are taken into account. Most significantly, we see that, in contrast to what was the case for LT and OR, the likelihoods are going *down* as more worlds are taken into consideration.<sup>12</sup> A pressing question then is whether they go down all the way to 0, effectively making HT, at least when we leave the toy models behind, a skeptical proposal.

That this is not the case follows from the following

**Theorem:** *Given HT with threshold  $\theta$ , where  $\mathbf{p} = (p_1, \dots, p_n) \sim \text{Dir}(\mathbf{1})$  such that  $p_i > 0$  for all  $i$ ,  $(1-\theta)/\theta$  is an approximate lower bound on the likelihood that  $\mathbf{p}$  gives rise to an informative belief set.*

A proof is given in the Appendix.

If  $(1-\theta)/\theta$  is an approximate lower bound on the likelihood of finding an informative belief set, given HT with threshold  $\theta$ , might it also be an approximate upper bound? The question is justified in view of the fact that, as seen in Table 4, at least for the higher threshold values, the likelihoods for  $n = 1,000$  are already very close to  $(1-\theta)/\theta$ .

To investigate this question, we register, for various values of  $n$ , what percentage of distributions that were “hits” qualify as such due to the smallest probability being smaller than  $(1-\theta)/\theta$  times the second smallest probability—distributions with informatively stable sets (i.e., propositions) of size  $n-1$ —what percentage were due to the sum of the smallest and second smallest probability being smaller than  $(1-\theta)/\theta$  times the third smallest probability—distributions with informatively stable sets of size  $n-2$ —and so on. It turns out that, once the informatively stable sets of size  $n-1$  have been taken into account, then the informatively stable sets of size  $n-2$  have only very little to add to the overall score, and once the

<sup>12</sup>This is not true for  $\theta = .5$ , where the likelihood remains at 1, just as for LT and OR.

Table 5: Percentage of contribution to overall score from distributions with informatively stable sets of size  $n-1$  /  $n-2$  /  $n-3$ .

no. of worlds	$\theta$					
	.6	.7	.8	.9	.95	.99
10	91/7/2	89/9/2	91/8/1	94/5/1	97/3/0	99/1/0
25	92/6/2	90/8/2	92/7/1	96/4/0	98/2/0	99/1/0
50	92/6/2	91/8/1	92/7/1	95/4/1	98/2/0	99/1/0
100	92/6/2	91/8/1	93/7/0	95/4/1	98/2/0	99/1/0
250	91/7/2	91/8/1	92/7/1	96/4/0	98/2/0	99/1/0
500	93/6/1	91/8/1	92/7/1	95/4/1	97/2/1	99/1/0
1000	93/6/1	91/8/1	93/7/0	96/4/0	98/2/0	99/1/0

informatively stable sets of size  $n-1$  and  $n-2$  have been taken into account, the contribution of the informatively stable sets of size  $n-i$ , for all  $i > 2$ , is negligible or even nil. We find this result to hold across the various values of  $\theta$  and to be virtually independent of the size of the support of the distributions (see Table 5 for the exact outcomes). Precisely because the percentages of contribution appear to hardly depend on the size of the support, a reasonable estimate of an upper bound on the likelihood of finding an informative belief set given, for instance,  $\theta = .8$  is  $(^{100/93})(.25) \approx .2688$ , and for  $\theta = .99$ , it is  $(^{100/99})(.0101) \approx .0102$ ; similarly for the other values of  $\theta$ .

Based on his results for the three-world and four-world cases, Rott [2017] already ventured that the stability account leads to an unacceptable skepticism.<sup>13</sup> We now see more exactly how skeptical the resulting position is. Assuming  $\theta = .9$ —an assumption that, as we pointed out above, many find reasonable—the chance of finding an informative belief set is barely above 10 percent.

It further follows from the above results that, for all practical purposes, the stability account boils down to a condition on the three, at most four, least probable worlds, however many worlds we are taking into account.<sup>14</sup> Check whether the least probable world is at least  $\theta/(1-\theta)$  times less probable than the second least probable world, and if not, check whether the sum of the probabilities of the two least probable worlds is at least  $\theta/(1-\theta)$  times smaller than the probability of the third least probable world. If both queries come back negative and you are the fussy type, you may still want to check whether the sum of the probabilities of the three least probable worlds is at least  $\theta/(1-\theta)$  times smaller than the probability of the fourth least probable world. But then you are really done. It was mentioned at the outset that, by focusing on toy models, we run the risk of missing some real-world complexity; in the present case, focus on the toy models rather seems to have hidden simplicity.

Note that this finding has a silver lining. We saw that, computationally, it is much less costly to determine whether  $L^*$  or  $O^*$  is satisfied than to determine whether  $H^*$  is satisfied. For  $L^*$  and  $O^*$ , we have to inspect only one world, or respectively two worlds. For  $H^*$ , however, we have to check whether there is some world that is  $\theta/(1-\theta)$  times more probable than all less probable worlds taken together. This would seem to imply that we may have to check *many* worlds and compare their probabilities with the sum of the probabilities of the less probable worlds (which we would have to calculate each time); and for any realistic number of worlds, that seems to be impossible for beings like us. As we saw, however, that is not really necessary. The number of comparisons HT requires one to make is, in effect, hardly greater than the number of comparisons the other two belief rules require.

<sup>13</sup>Makinson [2015] had already voiced a similar concern.

<sup>14</sup>Recall that we are ignoring probability 0 worlds, so that “least probable” means “least probable among the worlds that receive positive probability.”

It is worth comparing our results concerning the stability account with the recent results of Schurz [2017]. Schurz maintains that “an ‘ideal match’ between qualitative and quantitative belief systems . . . is possible only for very small belief systems over very coarse-grained conceptual spaces.” This seems to be at odds with our finding that there are—however few—stable belief systems of ever larger cardinality. A closer look reveals that the conflict is apparent only. First, Schurz is primarily concerned with the Lockean thesis and conjunctive closure (which in effect amounts to the Humean thesis for  $\theta = .5$ ), while we focus on the more recent work of Leitgeb in which he advocates the Humean thesis with thresholds between .5 and 1. Second, Schurz is interested in Lockean thresholds (for which there are lower bounds), while we are interested in Humean thresholds (that do not have lower bounds). These differences, however, are of minor relevance. The most important difference between Schurz’s approach and the present one is a difference in perspective: Schurz takes some Lockean threshold  $\theta$  as given and then goes on to show that stability entails that the number of “doxastic possibilities” must be low (below  $1/(1-\theta)$ ). In contrast, we have studied what happens when a (comparatively large) number  $n$  of possibilities in  $W$  is given. Table 5 makes clear that the informative stable sets of doxastic possibilities that exist have almost as many elements (mostly  $n - 1$  elements, in fact). Thus, they satisfy what Schurz calls “open-mindedness.” It is clear, however, that such sets have very high probabilities, and they can only be marked off by very high Lockean thresholds. It hence transpires that Schurz’ and our results confirm rather than contradict each other: one cannot have a reasonably high number of possibilities and a reasonably low threshold value for belief at the same time.

Taking stock, we have seen that LT and OR improve *greatly* on the skeptical proposal that nothing short of a probability of 1 suffices for belief, by raising very quickly the likelihood of finding an informative belief set to essentially 1.<sup>15</sup> In contrast, HT raises that likelihood *just slightly*, from 0 to around 10 or maybe 20 percent (assuming one of the more common values for  $\theta$ ). For instance, HT is about nine times more skeptical than either LT or OR if we assume  $\theta = .9$ , and even more than 99 times more skeptical if we assume  $\theta = .99$ .

**6. Anticipated objections.** We have looked at how three belief rules fare with respect to their ability to offer a non-skeptical account of the relationship between graded and categorical beliefs. We saw that LT and OR do much better in this respect than does HT. Now we discuss two possible objections that advocates of HT might want to level at our comparison.

*6.1. The insistence on large numbers of worlds.* Leitgeb, Lin and Kelly, and Rott use only very small numbers of possible worlds in their examples. We argued that this was a serious limitation, which hid from view the extent to which both LT and OR differ from HT in their potential to avoid skepticism. But advocates of HT might try to argue that restricting the discussion to a limited number of possible worlds—the kind of case for which the likelihoods reported in Tables 2-4 are still relatively close—is actually justified.

Leitgeb [2014, 152-160; 2017, 33-41, 137-147] is in fact explicit that he views possible worlds not as maximally specific worlds, but rather as cells in a partition of the space of all possibilities, made according to the questions that one is interested in, or that one’s attention is directed to, at a particular moment. Such partitions may change from one instant to another, or more generally from one context to another. As Leitgeb puts it:

---

<sup>15</sup>It might be thought that a downside of LT and OR is that, by guaranteeing that there will be informative belief sets, these rules effectively preclude radical agnosticism; and although perhaps not an attitude many have, it would seem wrong to rule it out a priori. The thought is mistaken, however, since agnosticism is not really ruled out. As mentioned in note 8, the flat probability distribution does not give rise to an informative belief set, and its existence is consistent with the claim that the likelihood of having an informative belief set, when LT and OR are applied, is essentially 1.

The context must include or determine a partition of the 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. ([2014, 159f])

[I]n 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 whether the three propositions *A*, *B*, *C* are the case or not. ([2014, 133])

To model a situation in which we are interested in only three propositions, we need no more than eight worlds. Leaving aside the fact that already for eight worlds HT is considerably more skeptical than OR, let alone LT, it should be noted that if HT is to attain any level of generality, then this argument succeeds as a justification of limiting the discussion to small numbers of worlds only if we *in general* consider only small numbers of propositions—and that is patently false.<sup>16</sup>

Independently, Leitgeb’s interpretation of “possible worlds” might be fine if all that mattered to rational belief is what we are currently focusing on, or what is present in our working memory. But there are arguably many things stored in our retentive memory that are *always* relevant to what beliefs we can adopt. It would seem absurd to say that it is fine for a person to believe things that are inconsistent with some of her firmly held beliefs, as long as the latter are only in the back of her head, slumbering in retentive memory.

Hence, we reject the suggested interpretation because we believe that persons like to be consistent across varying contexts. In fact, we would deem anyone spineless who tried to excuse herself upon being caught to have contradicted what she asserted earlier merely by pointing out that her contradictory statements were made in different contexts. Everybody has some firmly held convictions that she takes to partly “define” her as a person, even though these convictions will often not be in the front of her mind. It is necessary to distinguish between a person’s explicit beliefs and her implicit beliefs, and any representation of a belief state that leaves out implicit beliefs must be considered seriously incomplete.

In line with these intuitions, we prefer to conceive of belief-sets-cum-probabilities as holistic representations of fully-fledged belief states which change only as a result of some sort of doxastic revision. Of course, one’s receiving new evidence may well be a perfect reason for contradicting a statement made on a previous occasion. While clearly subject to motivated changes, belief states—of which probabilities are parts—should be conceived as anchored in long-term memory and thus as less ephemeral entities than those envisioned by Leitgeb. Belief states are not only concerned with the specific question that a reasoner happens to be interested in at a particular point of time or in a particular context; this would make belief states blow with the wind.

*6.2. The Lottery Paradox and other dimensions of comparison.* Our main focus in this paper has been the question of skepticism, specifically, of how LT, OR, and HT improve upon the skeptical probability 1 requirement on categorical belief. Arguably, however, skepticism is not the only dimension along which LT, OR, and HT are to be assessed. It was already mentioned in the introduction that, in contrast to OR and HT, LT clashes with the thought that our beliefs ought to be closed under logical derivability. That may constitute an important reason for preferring OR and HT over LT.

The failure of LT to respect closure is usually brought into relief by means of Kyburg’s [1961] Lottery Paradox. For any admissible threshold value  $\theta$ , we can think of a fair lottery so

<sup>16</sup>If this needs an argument, consider a *real* example: Hessler et al. [2016] tried to predict dementia risk on the basis of six factors (including smoking and physical activity), where each factor had three levels (ideal/moderate/poor). To represent semantically all combinations of factor levels that might be associated with an increased risk of dementia, together with the two levels of the dependent variable (dementia/no dementia), takes  $2 \times 3^6 = 1,458$  possible worlds. And six predictors is an extremely modest number in the age of Big Data (Efron and Hastie [2016]).



large that, for every ticket in the lottery, the probability that it will lose exceeds  $\theta$ . Applying LT, we would thereby believe of each ticket that it will lose, and that is so even if at the same time we know that the lottery will have a winner.

It merits emphasis that it is not *generally* agreed that the Lottery Paradox dooms LT. Kyburg and others were prepared to abandon the closure of belief under logical derivability.<sup>17</sup> On the other hand, it is fair to say that nowadays a majority of philosophers reject this solution to the paradox. Many of them have tried to solve the paradox by attaching some sort of disclaimer to LT meant to prevent so-called lottery propositions—propositions stating that a given ticket in a fair and large enough lottery will lose—from qualifying as rationally believable (e.g., Pollock [1990], Ryan [1996], Nelkin [2000], Douven [2002]), but so far such attempts have been largely unsuccessful (Douven and Williamson [2006]). Lin and Kelly as well as Leitgeb propose to solve the Lottery Paradox in a different manner. We already saw that in Leitgeb's theory, context plays a key role, and the same is true for Lin and Kelly's theory (see Lin and Kelly [2012a, 567–572]). Their contextualism allows these authors to circumvent the Lottery Paradox by holding that, in some contexts, one can rationally believe of a given ticket that it will lose, while in others, one cannot believe so—all depending on which space of propositions one attends to.

Above, we expressed our concerns about the contextualism that Leitgeb endorses, and these concerns extend to Lin and Kelly's similar proposal. Nonetheless, it might be that contextualism is the price we must pay for solving the Lottery Paradox while maintaining logical closure. If so, it is to be noted that with OR we get much better value for our money than with HT, given that the latter is only slightly less skeptical than the probability 1 requirement, which blocks the Lottery Paradox at no extra cost at all.

Advocates of HT might respond that, in contrast to OR, (i) HT solves the Preface Paradox (Makinson [1965]), and (ii) it satisfies the *monotonicity principle* stating that if a proposition  $A$  is believed and another proposition  $B$  is at least as probable as  $A$ , then  $B$  is believed as well (Leitgeb [2017, 120]). As to (i), while OR does not solve the Preface Paradox on its own, there are many solutions to that paradox on the market which Lin and Kelly could appropriate.<sup>18</sup> As to (ii), many may regard the monotonicity principle as a bug of the stability account rather than a selling point. That will be true of all those who have responded to the Lottery Paradox by adding a disclaimer to LT. For example, Nelkin [2000] argues that lottery propositions are special in that the support we have for them is of an exclusively statistical nature and that therefore they cannot be rationally believed.<sup>19</sup> This is so even if the relevant lottery is so large that the lottery propositions it gives rise to are more probable than *any* contingent non-lottery proposition we happen to believe.

**7. Conclusion.** We have looked at three proposals for how to conceive of the connection between (categorical) beliefs and probabilities, focusing on how well they do with respect to avoiding the kind of skepticism associated with the most straightforward connection, which makes probability 1 a requirement for belief. We compared two recent proposals with one another and also with the older Lockean thesis, which uses high but non-maximal probability as a criterion for belief. This was a continuation of work begun in Rott [2017], which however compared the accounts in the context of toy models of rational belief featuring only three or four possible worlds. We have argued that this is not sufficient for saying with any confidence

---

<sup>17</sup>See, e.g., Klein [1985], Foley [1992a], Christensen [2004], and Kroedel [2012].

<sup>18</sup>For instance, Kaplan [1995], Douven and Uffink [2003], Cevolani and Schurz [2017], and Kim and Vasudevan [2017, Sect. 5]. D'Alfonso [to appear] shows how Cevolani and Schurz's solution to the Preface Paradox can be extended to obtain a solution to the Lottery Paradox.

<sup>19</sup>See Douven [2003] for critical discussion of Nelkin's argument, which however does not invalidate the intuition that evidence of a strictly statistical nature is insufficient to warrant belief.

how we may expect the accounts to fare in more realistic settings, when many more possible worlds are taken into consideration.

We showed that the two new accounts, which might easily look like minor variants of one another, in fact differ quite dramatically. Specifically, it has turned out that Lin and Kelly’s account improves greatly over the skepticism of the probability 1 requirement, and the same is true of the more traditional Lockean thesis. Depending on whether one thinks context-sensitivity (manifest as partition-sensitivity) is a price worth paying for the logical closure of categorical beliefs, one may then prefer one or the other. The prospects for Leitgeb’s stability account were seen to be bleaker. Supposing the threshold for belief is considerably higher than .5—as many have argued it should be—the chances that an agent believes categorically any proposition beyond those that she believes with probability 1 are disappointingly low. Those chances are not nil, but it is doubtful whether the small improvement over the skeptical account outweighs the greater complexity of the stability account, or the contextualism that comes with it.<sup>20</sup>

## Appendix

In this appendix, we prove the theorem, stated in Section 5, about the likelihood of finding an informative belief set having  $(1-\theta)/\theta$  as an approximate lower bound, applying HT with threshold  $\theta$ . The proof uses insights from order statistics, which also surfaced in Section 4. As mentioned there, virtually all results from order statistics concern independent and identically distributed samples of random variables. We could still make use of order statistics in that section, given that, while  $\mathbf{p} = (p_1, \dots, p_n) \sim \text{Dir}(\mathbf{1})$  is not *exactly* modeled by  $\mathbf{p}^* = (p_1^*, \dots, p_n^*)$  with  $p_i^* \sim \text{Beta}(1, n-1)$ , the latter can still be regarded as offering a good approximation of the former. Importantly for the proof to come, however, in the limit of  $n$  going to infinity,  $\mathbf{p}^* = (p_1^*, \dots, p_n^*)$  with  $p_i^* \sim \text{Beta}(1, n-1)$  does *exactly* model  $\mathbf{p} = (p_1, \dots, p_n) \sim \text{Dir}(\mathbf{1})$ .

To see this, note that the only difference between  $\mathbf{p}$  and  $\mathbf{p}^*$ , given finite  $n$ , is that the former adds up to 1 while the latter does not necessarily do so. But in the limit this difference does not exist. Where  $X_1, \dots, X_n$  are drawn from a random variable with mean  $\mu_X$  and standard deviation  $\sigma_X$ , it follows by the Central Limit Theorem that  $\sum_{i=1}^n X_i \sim \mathcal{N}(n \cdot \mu_X, \sqrt{n} \cdot \sigma_X)$  as  $n$  goes to infinity. For  $X \sim \text{Beta}(1, n-1)$ , it is known that

$$\mu_X = \frac{1}{n} \quad \text{and} \quad \sigma_X = \sqrt{\frac{n-1}{n^2(n+1)}}.$$

Hence, where  $p_i \sim \text{Beta}(1, n-1)$ , we have that, for  $n \rightarrow \infty$ ,

$$\sum_{i=1}^n p_i \sim \mathcal{N}\left(n \cdot \frac{1}{n}, \sqrt{n} \cdot \sqrt{\frac{n-1}{n^2(n+1)}}\right) =$$

Because

$$\lim_{n \rightarrow \infty} \left( \sqrt{\frac{n-1}{n(n+1)}} \right) = 0,$$

we find that in the limit of  $n$  going to infinity,  $\mathbf{p}^*$  does sum to 1, just like  $\mathbf{p}$  does.

---

<sup>20</sup>We are grateful to Patricia Mirabile, Gerhard Schurz, Stefan Solbrig, Christopher von Bülow, Sylvia Wenmackers, and three referees of this journal for valuable comments on previous versions of this paper. We also thank audiences at the University of Leuven and TU Dortmund for stimulating questions and comments.

In preparation for the proof, we record the following further facts: First, where  $X$  is a random variable with pdf  $f(x)$  and cdf  $F(x)$ , and where  $\mathbf{X} = (X_1, \dots, X_n)$  is any random sample from any distribution of  $X$ ,  $X_{(i)}$  denotes the  $i$ -th order statistic of  $\mathbf{X}$ , that is, the  $i$ -th smallest value of  $\mathbf{X}$ . Given  $i < j$ , the joint pdf for  $X_{(i)}$  and  $X_{(j)}$  is

$$f_{i,j}(x, y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(x)^{i-1} (F(y) - F(x))^{j-i-1} (1 - F(y))^{n-j} f(x) f(y), \quad (1)$$

with  $x, y \in \mathbb{R}$  and  $x < y$  (Arnold, Balakrishnan, and Nagaraja [2008, 16]). Second, supposing  $X \sim \text{Beta}(1, n-1)$  and noting that  $B(1, n-1) = 1/(n-1)$ , we get that

$$f(x) = \frac{(1-x)^{n-2}}{B(1, n-1)} = (n-1)(1-x)^{n-2} \quad \text{and} \quad F(x) = 1 - (1-x)^{n-1}. \quad (2)$$

Finally, for convenience of notation, we define  $r_\theta := (1-\theta)/\theta$ . We are now ready to prove

**Theorem:** *Given HT with threshold  $\theta$ , where  $\mathbf{p} = (p_1, \dots, p_n) \sim \text{Dir}(\mathbf{1})$  such that  $p_i > 0$  for all  $i$ ,  $r_\theta$  is an approximate lower bound on the likelihood that  $\mathbf{p}$  gives rise to an informative belief set.*

*Proof:* The proof is in two parts. First we show that, applying HT with threshold  $\theta$ , the likelihood of finding an informative belief set converges to a limit that is greater than or equal to  $r_\theta$ , and then we show that this convergence is at least approximately monotonic.

1. Given  $\mathbf{H}^*$ , the likelihood that  $\mathbf{p}$  gives rise to an informative belief set is at least as great as the likelihood that  $p_{(1)} < r_\theta \cdot p_{(2)}$ . We prove that the latter likelihood goes to  $r_\theta$  as  $n \rightarrow \infty$ . From (1) we derive that, for any given random variable  $X$ , the joint pdf of the two smallest order statistics,  $X_{(1)}$  and  $X_{(2)}$ , is given by

$$f_{1,2}(x, y) = \frac{n!}{(n-2)!} f(x) f(y) (1 - F(y))^{n-2}. \quad (3)$$

Given that  $p_i \sim \text{Beta}(1, n-1)$ , we further derive from (2) that the joint pdf of  $p_{(1)}$  and  $p_{(2)}$  is

$$g_{1,2}(x, y) = n(n-1)^3 (1-x)^{n-2} (1-y)^{n^2-2n}. \quad (4)$$

We want to determine the likelihood that  $p_{(1)} < r_\theta \cdot p_{(2)}$ , so instead of (4) we must consider the following transformed joint pdf of the two smallest order statistics, where the second order statistic is scaled by  $r_\theta$  (see Rice [2007, Ch. 3.6] on how to work out the transformation):

$$h(x, y) = \frac{n}{r_\theta} (n-1)^3 (1-x)^{n-2} \left(1 - \frac{y}{r_\theta}\right)^{n^2-2n},$$

where now  $y \in (0, r_\theta)$ . To determine the likelihood we are after, we must integrate  $h$  over the region of the domain of integration where  $x < y$ , which is the region defined by  $\{(x, y) : 0 \leq x < y < r_\theta\}$ . To facilitate the integration, we first pull out

$$\frac{n}{r_\theta} (n-1)^3, \quad (5)$$

which is a constant as far as the integration is concerned. Then we evaluate

$$\int_0^{r_\theta} \int_x^{r_\theta} (1-x)^{n-2} \left(1 - \frac{y}{r_\theta}\right)^{n^2-2n} dy dx. \quad (6)$$

Using substitution, we find that

$$\int (1-x)^{n-2} \left(1 - \frac{y}{r_\theta}\right)^{n^2-2n} dy = -\frac{r_\theta(1-x)^{n-2} \left(1 - \frac{y}{r_\theta}\right)^{(n-1)^2}}{(1-n)^2}.$$

The resulting expression vanishes for  $y = r_\theta$ , so the inner integral in (6) evaluates to

$$\frac{r_\theta(1-x)^{n-2} \left(1 - \frac{x}{r_\theta}\right)^{(n-1)^2}}{(1-n)^2}.$$

This expression has a power series representation, which is known in virtue of a result due to Euler (Andrews, Askey, and Roy [1999, Thm. 2.2.1]). After first pulling out the constant

$$\frac{r_\theta}{(1-n)^2}, \quad (7)$$

we leave the exact calculation to *Mathematica*, which yields

$$\int_0^{r_\theta} (1-x)^{n-2} \left(1 - \frac{x}{r_\theta}\right)^{(n-1)^2} dx = \frac{r_\theta ({}_2F_1(1, 2-n; n^2-2n+3; r_\theta))}{n^2-2n+2},$$

where  ${}_2F_1$  is the Gaussian hypergeometric function, which is defined as

$${}_2F_1(a, b; c; z) := \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!},$$

with  $(s)_n = \prod_{k=0}^{n-1} (s+k)$  for  $n > 0$  and  $(s)_n = 1$  for  $n = 0$ . The product of the constants (5) and (7) that were pulled out in previous steps simplifies to  $n^2 - n$ . Thus, where  $n$  is the number of worlds, the likelihood that the least probable world has a probability less than  $r_\theta$  times the probability of the second least probable world is given by the following expression:

$$\frac{r_\theta(n^2 - n) ({}_2F_1(1, 2-n; n^2-2n+3; r_\theta))}{n^2-2n+2}. \quad (8)$$

Our goal is to show that this expression approaches  $r_\theta$  as  $n$  gets larger and larger. We first observe that the limit we want to determine can be written as

$$r_\theta \lim_{n \rightarrow \infty} \left( \frac{n^2 - n}{n^2 - 2n + 2} \right) \lim_{n \rightarrow \infty} ({}_2F_1(1, 2-n; n^2-2n+3; r_\theta)).$$

It is easily seen that

$$\lim_{n \rightarrow \infty} \left( \frac{n^2 - n}{n^2 - 2n + 2} \right) = 1.$$

As for the limit of the remaining expression, note that the beginning of the series expansion of  ${}_2F_1(1, 2-n; n^2-2n+3; r_\theta)$  looks as follows:

$$1 - \frac{(n-2)r_\theta}{n^2-2n+3} + \frac{(n-2)(n-3)r_\theta^2}{(n^2-2n+3)(n^2-2n+4)} - \frac{(n-2)(n-3)(n-4)r_\theta^3}{(n^2-2n+3)(n^2-2n+4)(n^2-2n+5)} + \frac{(n-2) \cdots (n-5)r_\theta^4}{(n^2-2n+3) \cdots (n^2-2n+6)} - \cdots$$

The limits of the  $i$ -th fraction appearing in this series have the schematic form

$$r_\theta^i \cdot \lim_{n \rightarrow \infty} \left( \frac{\prod_{k=1}^i (n-k-1)}{\prod_{k=1}^i (n^2-2n+k+2)} \right).$$

To see that this expression evaluates to 0, note that, first,  $(n-k-1) < n$  and  $n^2 - 2n + k + 2 > n^2 - 2n$  and thus

$$\left( \frac{\prod_{k=1}^i (n-k-1)}{\prod_{k=1}^i (n^2 - 2n + k + 2)} \right) < \left( \frac{n}{n^2 - 2n} \right)^i = \left( \frac{1}{n-2} \right)^i,$$

and second,  $(1/n-2)^i$  goes to 0 as  $n$  goes to infinity. As a result, the limit of  ${}_2F_1(1, 2-n; n^2 - 2n + 3; r_\theta)$  as  $n \rightarrow \infty$  equals 1, and so

$$r_\theta \lim_{n \rightarrow \infty} \left( \frac{n^2 - n}{n^2 - 2n + 2} \right) \lim_{n \rightarrow \infty} ({}_2F_1(1, 2-n; n^2 - 2n + 3; r_\theta)) = r_\theta \cdot 1 \cdot 1 = r_\theta.$$

This means that, as more and more worlds are considered, the likelihood that the least probable world is smaller than  $r_\theta$  times the probability of the second least probable world goes to  $r_\theta$ .

2. The result established so far is consistent with the likelihood of finding an informative belief set, given HT with threshold  $\theta$ , dipping below  $r_\theta$  for specific values of  $n$ . To show that  $r_\theta$  is a lower bound on that likelihood, it must be shown that (8) is monotonically decreasing. A standard approach here would be to show that its first derivative is always negative, or else to show that the ratio of successive terms is never greater than 1. But taking the derivative with respect to  $n$  involves taking partial derivatives of the second and third parameter of the hypergeometric function in the numerator of (8), and while in recent years progress has been made on differentiating the hypergeometric function (Ancarani and Gasaneo [2009]) and *Mathematica* can even find the first derivative of (8), the resulting expression is so complicated that we are unable to determine its sign. (Of course we can determine its sign for any particular combination of values for  $n$  and  $r_\theta$ , but that is not giving us a proof.) For the ratio test, basically the same problem arises. Thus, rather than showing that (8) is monotonically decreasing, we show that a function closely approximating (8) is monotonically decreasing—which is why we are only able to claim that  $r_\theta$  is an *approximate* lower bound on the likelihood of finding an informative belief set.

First note that (8) can be given the following series representation:

$$\begin{aligned} \frac{(n^2 - n)r_\theta}{n^2 - 2n + 2} &= \frac{(n-2)(n^2 - n)r_\theta^2}{(n^2 - 2n + 2)(n^2 - 2n + 3)} + \\ &\quad \frac{(n-3)(n-2)(n^2 - n)r_\theta^3}{(n^2 - 2n + 2)(n^2 - 2n + 3)(n^2 - 2n + 4)} - \\ &\quad \frac{(n-4)(n-3)(n-2)(n^2 - n)r_\theta^4}{(n^2 - 2n + 2)(n^2 - 2n + 3)(n^2 - 2n + 4)(n^2 - 2n + 5)} + \cdots \quad (9) \end{aligned}$$

This is an alternating series, and one readily verifies that it satisfies the conditions of the Alternating Series Estimation Theorem (see, e.g., Stewart [2012, 754]). From this theorem it follows that the bound on the truncation error of the partial sum of the first  $k$  terms of (9) is given by the  $(k+1)$ -st term. In particular, if we add up only the first two terms of (9), then the bound on the truncation error is given by this expression:

$$\frac{(n-3)(n-2)(n^2 - n)r_\theta^3}{(n^2 - 2n + 2)(n^2 - 2n + 3)(n^2 - 2n + 4)}.$$

As  $n$  goes to infinity, this expression goes to 0, but already for  $n \geq 1000$  the truncation error is essentially 0, for any of the values of  $\theta$  considered in the paper. (Notice that we are interested in large values of  $n$ ; for  $n$  up to 1000, we already know, from the numerical integration results, that (8) is monotonically decreasing, for all values of  $\theta$ .) Hence, the

function

$$f_\theta(n) := \frac{(n^2 - n)r_\theta}{n^2 - 2n + 2} - \frac{(n - 2)(n^2 - n)r_\theta^2}{(n^2 - 2n + 2)(n^2 - 2n + 3)} \quad (10)$$

closely approximates (8) for large  $n$  and for all relevant values of  $\theta$ .

To show that  $f_\theta$  is monotonically decreasing for those values of  $\theta$ , we show that  $f_\theta(n - 1) - f_\theta(n) \geq 0$ , for all  $n$ . Writing out  $f_\theta(n - 1) - f_\theta(n)$ , we get

$$\left( \frac{(n^2 - n)r_\theta}{n^2 - 2n + 2} - \frac{(n - 2)(n^2 - n)r_\theta^2}{(n^2 - 2n + 2)(n^2 - 2n + 3)} \right) - \left( \frac{(n^2 + n)r}{n^2 + 1} - \frac{(n - 1)(n^2 + n)r_\theta^2}{(n^2 + 1)(n^2 + 2)} \right), \quad (11)$$

which after some algebraic manipulation reduces to

$$\frac{nr_\theta(n^5(1 - r_\theta) + n^4(3r_\theta - 5) + n^3(3r_\theta + 11) - n^2(11r_\theta + 19) + 2n(8r_\theta + 9) - 2(5r_\theta + 9))}{(n^2 + 1)(n^2 + 2)(n^2 - 2n + 2)(n^2 - 2n + 3)}.$$

Note that the denominator is positive for all relevant values of  $n$ , so the sign of the expression as a whole depends on the sign of the numerator. Expanding all the terms in the numerator yields

$$n^6r_\theta - n^6r_\theta^2 + 3n^5r_\theta^2 - 5n^5r_\theta + 3n^4r_\theta^2 + 11n^4r_\theta - 11n^3r_\theta^2 - 19n^3r_\theta + 16n^2r_\theta^2 + 18n^2r - 10nr_\theta^2 - 18nr_\theta.$$

To show that this expression is greater than 0 for all relevant values of  $r_\theta$ , we set it equal to 0 and solve for  $r_\theta$  in terms of  $n$ :

$$n^6r_\theta - n^6r_\theta^2 + 3n^5r_\theta^2 - 5n^5r_\theta + 3n^4r_\theta^2 + 11n^4r_\theta - 11n^3r_\theta^2 - 19n^3r_\theta + 16n^2r_\theta^2 + 18n^2r - 10nr_\theta^2 - 18nr_\theta = 0$$

if and only if (dividing by  $nr_\theta$  and rearranging terms)

$$n^5 - 5n^4 + 11n^3 - 19n^2 + 18n - 18 = n^5r_\theta - 3n^4r_\theta - 3n^3r_\theta + 11n^2r_\theta - 16nr_\theta + 10r_\theta$$

if and only if (pulling out  $r_\theta$  from the right-hand side and dividing the left-hand side by the remaining factor)

$$r_\theta = \frac{n^5 - 5n^4 + 11n^3 - 19n^2 + 18n - 18}{n^5 - 3n^4 - 3n^3 + 11n^2 - 16n + 10}. \quad (12)$$

We now find that (11) is smaller than 0 precisely if  $r_\theta$  is greater than the right-hand side of (12). Already for  $n = 1000$ , the right-hand side equals .998, and  $r_\theta > .998$  corresponds to  $\theta < .500498$ . Hence, assuming any of the thresholds considered in this paper,  $f_\theta(n - 1) > f_\theta(n)$  for all  $n \geq 1000$ . (As stated earlier, for values of  $n$  up to 1000 we already know that (8) is monotonically decreasing.) And given that, for any relevant  $\theta$ , the function  $f_\theta$  closely approximates (8), it follows that the latter is approximately monotonically decreasing as well.

Combining this with the result from the first part of the proof yields the conclusion that  $r_\theta$  is an approximate lower bound on the likelihood of finding an informative belief set, given HT with threshold  $\theta$ . ■

## References

Ancarani, L. U., and Gasaneo, G. [2009]. Derivatives of any order of the Gaussian hypergeometric function  ${}_2F_1(a, b, c; z)$  with respect to the parameters  $a$ ,  $b$  and  $c$ . *Journal of Physics A: Mathematical and Theoretical* 42(39):1-10. doi: 10.1088/1751-8113/42/39/395208

- Andrews, G. E., Askey, R., and Roy, R. [1999]. *Special functions*. Cambridge: Cambridge University Press.
- Arnold, B. C., Balakrishnan, H. N., and Nagaraja, H. N. [2008]. *A first course in order statistics*. Philadelphia: SIAM.
- Bellman, R. E. [1957]. *Dynamic programming*. Princeton, NJ: Princeton University Press.
- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. [2017]. Julia: A fresh approach to numerical computing. *SIAM Review* 1(59):65–98.
- Cevolani, G., and Schurz, G. [2017]. Probability, approximate truth, and truthlikeness: More ways out of the preface paradox. *Australasian Journal of Philosophy* 95(2):209–225.
- Chi, M. T. H., and Ohlsson, S. [2005]. Complex declarative learning. In K. J. Holyoak and R. G. Morrison (eds.), *The Cambridge handbook of thinking and reasoning* (pp. 371–399). Cambridge: Cambridge University Press.
- Christensen, D. [2004]. *Putting logic in its place: Formal constraints on rational belief*. Oxford: Oxford University Press.
- Connor, R. J., and Mosimann, J. E. [1969]. Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association* 64(325):194–206.
- D'Alfonso, S. [to appear]. Truthlikeness and the lottery paradox via the preface paradox. *Australasian Journal of Philosophy*. doi: 10.1080/00048402.2017.1372491
- David, H. A., and Nagaraja, H. N. [2003]. *Order statistics*. Hoboken, NJ: Wiley.
- de Finetti, B. [1962]. Does it make sense to speak of ‘good probability appraisers’? In I. J. Good (ed.), *The scientist speculates: An anthology of partly-baked ideas* (pp. 357–364). New York: Basic Books.
- Douven, I. [2002]. A new solution to the paradoxes of rational acceptability. *British Journal for the Philosophy of Science* 53(3):391–410.
- Douven, I. [2003]. Nelkin on the lottery paradox. *Philosophical Review* 112(3):395–404.
- Douven, I., and Uffink, J. [2003]. The preface paradox revisited. *Erkenntnis* 59(3):389–420.
- Douven, I., and Williamson, T. [2006]. Generalizing the lottery paradox. *British Journal for the Philosophy of Science* 57(4):755–779.
- Easwaran, K. [2016]. Dr. Truthlove or: How I learned to stop worrying and love Bayesian probabilities. *Noûs* 50(4):816–853.
- Efron, B., and Hastie, T. [2016]. *Computer age statistical inference: Algorithms, evidence and data science*. Cambridge: Cambridge University Press.
- Foley, R. [1992a]. The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly* 29(2):111–124.
- Foley, R. [1992b]. *Working without a net: Essays in egocentric epistemology*. New York: Oxford University Press.
- Foley, R. [2009]. Beliefs, degrees of belief, and the Lockean thesis. In F. Huber and C. Schmidt-Petri (eds.), *Degrees of belief* (pp. 37–47). Dordrecht: Springer.
- Hessler, J. B., Ander, K.-H., Brönnner, M., Etgen, T., Förstl, H., Poppert, H., ... Bickel, H. [2016]. Predicting dementia in primary care patients with a cardiovascular health metric: A prospective population-based study. *BMC Neurology* 16:116. doi: 10.1186/s12883-016-0646-8
- Johnson, N., Kotz, S., and Balakrishnan, N. [1995]. *Continuous univariate distributions* (Vol. II). New York: Wiley.
- Kagehiro, D., and Stanton, W. [1985]. Legal vs. quantified definitions of standards of proof. *Law and Human Behavior* 9(2):159–178.
- Kaplan, M. [1981]. A Bayesian theory of rational acceptance. *Journal of Philosophy* 78(6):305–330.
- Kaplan, M. [1995]. Believing the improbable. *Philosophical Studies* 77(1):117–146.
- Kim, B., and Vasudevan, A. [2017]. How to expect a surprising exam. *Synthese* 194(8):3101–

3133.

- Klein, P. [1985]. The virtues of inconsistency. *Monist* 68(1):105–135.
- Kotz, S., Balakrishnan, N., and Johnson, N. L. [2000]. *Continuous multivariate distributions* (Vol. I). New York: Wiley.
- Krauth, W. [2006]. *Statistical mechanics: Algorithms and computations*. Oxford: Oxford University Press.
- Kroedel, T. [2012]. The lottery paradox, epistemic justification and permissibility. *Analysis* 72(1):57–60.
- Kyburg, H. E. [1961]. *Probability and the logic of rational belief*. Middletown, CT: Wesleyan University Press.
- Kyburg, H. E. [1990]. *Science and reason*. Oxford: Oxford University Press.
- Leitgeb, H. [2013]. Reducing belief simpliciter to degrees of belief. *Annals of Pure and Applied Logic* 164(12):1338–1389.
- Leitgeb, H. [2014]. The stability theory of belief. *Philosophical Review* 123(2):131–171.
- Leitgeb, H. [2015]. The Humean thesis on belief. *Proceedings of the Aristotelian Society, Supplementary Volume* 89(1):143–185.
- Leitgeb, H. [2017]. *The stability of belief: How rational belief coheres with probability*. Oxford: Oxford University Press.
- Lin, H., and Kelly, K. T. [2012a]. A geo-logical solution to the lottery paradox, with applications to conditional logic. *Synthese* 186(2):531–575.
- Lin, H., and Kelly, K. T. [2012b]. Propositional reasoning that tracks probabilistic reasoning. *Journal of Philosophical Logic* 41(6):957–981.
- Magnussen, S., Eilertsen, D. E., Teigen, K. H., and Wessel, E. [2014]. The probability of guilt in criminal cases: Are people aware of being “beyond reasonable doubt”? *Applied Cognitive Psychology* 28(2):196–203.
- Makinson, D. C. [1965]. The paradox of the preface. *Analysis* 25(6):205–207.
- Makinson, D. C. [2015]. *The scarcity of stable belief sets*. (Ms. of 22 February 2015, retrieved from [sites.google.com/site/davidcmakinson/listofpublications](https://sites.google.com/site/davidcmakinson/listofpublications))
- Metropolis, N., and Ulam, S. [1949]. The Monte Carlo method. *Journal of the American Statistical Association* 44(247):335–341.
- Moser, P. K., and Tlumak, J. [1985]. Two paradoxes of rational acceptance. *Erkenntnis* 23(2):127–141.
- Nelkin, D. K. [2000]. The lottery paradox, knowledge, and rationality. *Philosophical Review* 109(3):373–409.
- Pollock, J. [1990]. *Nomic probability and the foundations of induction*. Oxford: Oxford University Press.
- Pruss, A. [2012]. The badness of being certain of a falsehood is at least  $1/(\log 4 - 1)$  times greater than the value of being certain of a truth. *Logos and Episteme* 3(2):229–238.
- Rice, J. A. [2007]. *Mathematical statistics and data analysis*. Belmont CA: Thomson Brooks/Cole.
- Rott, H. [2017]. Stability and skepticism in the modelling of doxastic states: Probabilities and plain beliefs. *Minds and Machines* 27(1):167–197.
- Ryan, S. [1996]. The epistemic virtues of consistency. *Synthese* 109(2):121–141.
- Schurz, G. [2017]. Impossibility results for rational belief. *Noûs*. doi: 10.1111/nous.12214.
- Shear, T., and Fitelson, B. [2018]. Two approaches to belief revision. *Erkenntnis*. doi: 10.1007/s10670-017-9968-1.
- Stewart, J. [2012]. *Multivariable calculus*. Belmont CA: Thomson Brooks/Cole.