Regularity and Hyperreal Credences

Kenny Easwaran
University of Southern California

1. Introduction

It has been widely argued that belief is not just an all-or-nothing attitude—there is also a notion of belief that comes in degrees. Defenders of this position generally also argue that these degrees of belief, or "credences," obey something like the following principles:

- There is a set Ω of doxastic possibilities for each agent, propositions correspond to subsets of Ω, and the collection $\mathcal{F}$ of propositions in which the agent has credences is an algebra. (That is, $\mathcal{F}$ is nonempty, if a proposition is in $\mathcal{F}$, then so is its complement, and if two propositions are in $\mathcal{F}$, then so is their intersection.)[1]
- A rational agent's credences are given by a probability function P. (That is, $P(p) \geq 0$ for all propositions p, $P(\Omega) = 1$, and $P(p \cup q) = P(p) + P(q)$ whenever p and q are disjoint subsets of Ω.)

---

1. Some theorists prefer to think of the objects of credence as something more sentential, rather than as sets of possibilities. The set-theoretic notation I use throughout will have to be replaced by the corresponding syntactic notation: negation in place of complement, conjunction in place of intersection, and so forth. The only significant effect this will have on my argument is that at the end of section 4, when I discuss one option that makes use of this set Ω, such a theorist will have to take the other option, which uses the conditional credence function.

- A rational agent's conditional credences satisfy the relation $P(p|q)P(q) = P(p \cap q)$.[2]

The set of doxastic possibilities represents an agent's certainties and uncertainties. Its elements may be thought of as something like possible worlds, except that they may satisfy propositions that are metaphysically impossible, or possibly even contradictory. Any proposition that the agent is not certain of must be false at some doxastic possibility. Some authors might argue that belief just is truth at all doxastic possibilities, but I suspect that many proper subsets of Ω will correspond to beliefs as well. Belief does not entail certainty, the way such a proposal would suggest.

One straightforward consequence of these principles is that if a proposition corresponds to the empty set, $\varnothing$, then a rational agent has a credence of 0 in it.[3] I will call such a proposition "doxastically impossible" because it is not true in any doxastic possibility. Many philosophers also endorse the converse:

> **Regularity:** A rational agent has credence 0 in a proposition only if it is doxastically impossible for her. Equivalently, a rational agent has credence 1 in a proposition only if it is certain for her.[4]

Some philosophers instead state a version with some sort of nondoxastic modality, especially if they think of "doxastic possibilities" as having to be logically or metaphysically possible. However, I take it that these authors are generally committed to **Regularity** as just stated, as well as some version of:

> **X-Y Transmodal Connection:** Any X'ly possible proposition is Y'ly possible.

If "Y" is interpreted as doxastic possibility for a rational agent, then this, in combination with **Regularity**, entails that rational agents are only cer-

---

2. The most traditional understanding of conditional credence in fact *defines* $P(p|q)$ to equal $P(p \cap q)/P(q)$, but if we allow for the possibility that $P(q) = 0$, as is my goal in this essay, then this needs modification. I will discuss some options in sections 1 and 2, but all major proposals agree on the multiplicative formula given here.

3. Proof: the empty set is in $\mathcal{F}$ because there is some set in $\mathcal{F}$, and the intersection of this set with its complement is the empty set. $\varnothing$ is disjoint from any proposition p, so $P(p \cup \varnothing) = P(p) + P(\varnothing)$. Since $p \cup \varnothing$ just is p, this means that $P(\varnothing) = 0$.

4. Some authors use the term "regularity" descriptively rather than normatively. That is, for them, regularity is a property of probability functions, and there is a separate requirement for rational agents to have regular probability functions. But I will use the term "**Regularity**" to refer to the normative principle instead.

tain of logical necessities, or metaphysical necessities, or something similar.[5] Conversely, if "X" is interpreted as doxastic possibility for a rational agent, then we can interpret doxastic possibilities as possible worlds, or logical models, or something else more familiar. However, I will ignore these transmodal connections and focus instead on the principle I have called **Regularity**.

Regularity is in tension with the fact that there are specific propositions that an agent can't rule out, but for which any positive real number is clearly too high a value for the credence. I will call such propositions, as well as the numerical values (if any) of their credences, "minuscule" to avoid prejudging the question of whether their credence is 0, or if they are represented in some other way. (Arguments for the existence of minuscule propositions will be given in section 3.)

Skyrms (1980) (in a brief appendix) and Lewis (1980) (in two quick paragraphs) try to resolve this tension by suggesting that credences should not have to be real valued, but should instead be allowed to take on "infinitesimal" values as well. They point out that in the 1960s, Abraham Robinson showed the existence of mathematical structures, called "hyperreals," that behave very much like the real numbers, but include elements that are positive but smaller than any positive real number. (For instance, see Robinson 1996.) Skyrms and Lewis suggest that this theory, especially as developed by Bernstein and Wattenberg (1969), can be used to save **Regularity**, and this response has been generally accepted by philosophers working in the area for the past several decades.[6]

I think that this situation is largely based on a mistake about the role of numbers in mathematical representations. Probabilism uses a set

---

5. Some authors use these principles to argue that agents should use Jeffrey's (2004) alternative to the standard update method of conditionalization since it results in one having credence 1 in one's evidence, which is generally logically and metaphysically contingent. However, if one takes standard conditionalization to produce doxastic necessity and not just credence 1, then it is compatible with the principle that I call **Regularity**, so **Regularity** itself can't be dismissed just on the trivial grounds of incompatibility with a standard update rule.

6. As examples, see Lewis 1996, 303; Swinburne 2001, 244; Holder 2002, 296; and Norton 2007, 162. Note that Bartha and Hitchcock 1999 is *not* an instance of this sort of use of hyperreals. As they say, "we are not committed to the existence of infinitesimal degrees of belief or anything of that sort. Just as imaginary numbers can be used to facilitate the proving of theorems that exclusively concern real numbers, our use of [hyperreals] will be used to facilitate and motivate the construction of purely real-valued measures." Bartha and Hitchcock 1999, 416. However, the previously listed authors, and others following them, do suggest that agents can or must have hyperreal credences.

of possibilities to represent propositions, and real numbers to represent credences. Because the real numbers are not fine grained enough to capture all the distinctions in these doxastic states, Skyrms and Lewis argue that we should use hyperreals instead. They focus on the numerical aspect of a probabilistic representation and seek to expand it so that it can represent all the relevant distinctions. But as I presented it above, probabilism uses a *set* together with some numbers, and a *conditional* credence function as well as an unconditional one. Both of these tools describe important features of credence that shouldn't be overlooked.

In section 2, I give what I take to be the four main arguments for **Regularity** and show that these tools provide responses to all of them. In section 3, I present the problem of minuscule propositions. In section 4, I explain the hyperreals used by Lewis and Skyrms to respond to this problem, and in section 5, I show that they have too much structure to properly represent credences in ordinary propositions. Although one might think that the purely numerical representation with hyperreals is relatively simple, it turns out to have complexities far beyond those that arise from the consideration of the nonnumerical aspects of the standard representation.

This is not a definitive argument against **Regularity**, and in the appendix, I give quick overviews of a few other systems that might be used to achieve the goals that motivate it. I think pursuing probability theories based on any of these systems may be a valuable project and may help with our understanding of credence. And in fact, the hyperreals may also help, as long as we understand that they do not tell us the precise structure of credences and that not all distinctions they make should be taken to be significant. But for now, I claim that there is no reason to think credences have structure beyond that given in the opening paragraph of this introduction, with a set of doxastic possibilities, a standard real-valued probability function (which may assign 0 to doxastically possible propositions), and a standard real-valued conditional probability function.

2. Arguments for Regularity

2.1. Learning Probability 0

The first argument for **Regularity** is based on conditional credence. Lewis (1980, 267) says:

> I should like to assume that it makes sense to conditionalize on any but the empty proposition. Therefore I require that C is *regular*: $C(B)$ is zero,

and $C(A/B)$ is undefined, only if B is the empty proposition, true at no worlds.

The "C" Lewis refers to is the hypothetical initial credence function of a rational agent with no a posteriori information about the world. Lewis and other Bayesians suggest that the appropriate way for a rational agent to update her credences as she gains new information is to conditionalize—that is, the credence P_1 after the learning should be related to the initial credences P_0 by $P_1(A) = P_0(A|B)$, where A is any proposition in $\mathcal{F}$, and B is the proposition learned. Many philosophers follow chapter 1 of Kolmogorov 1950, where it is stipulated that $P(A|B) = \frac{P(A \cap B)}{P(B)}$, which is undefined if $P(B) = 0$. But since an agent with no a posteriori information should be able to learn *any* nonempty proposition, either every nonempty proposition must have nonzero probability (as Lewis claims), or there must be a way to update that goes beyond this standard notion of conditionalization and the standard definition of conditional probability.[7]

I will formalize the relevant version of the argument thus:

1. Any doxastically possible proposition can be learned.
2. When a rational agent learns B, she replaces her credence $P(A)$ with $P(A|B)$ for every proposition A.
3. $P(A|B)$ is defined as $P(A \cap B)/P(B)$, and thus is undefined when $P(B) = 0$.
4. For a rational agent, learning can't leave all credences undefined.
5. Therefore, a rational agent doesn't have credence 0 in any doxastically possible proposition.

Premise 1 seems straightforward.[8] For premise 2, consider what Skyrms (1980, 74) says on his version of this argument:

7. Both this and the next argument assume that updating proceeds by conditionalization. There is a commonly proposed alternative due to Richard Jeffrey, on which no single proposition needs to be learned with certainty, so that the update is compatible with maintaining uncertainty in the learned proposition, and thus is compatible with certain transmodal connections. However, this alternative still relies on $P(A|B)$, and is thus undefined if $P(A|B)$ is. Thus, using Jeffrey conditionalization instead of standard conditionalization makes no relevant difference to either of these arguments.

8. The converse of this claim is perhaps more interesting—is it the case that everything that can be learned must be doxastically possible? It seems plausible to me that we ought to treat revisions where we give up a previous certainty as the kind that motivate an alternative to conditionalization. This situation may be more usefully studied by tech-

> How do we assimilate new knowledge of a proposition with a prior probability of zero? . . . [P]erhaps at any rate we will need external rules for some cases of belief-change not properly treated by conditionalization. But the choice should be dictated by epistemological considerations, not by the mathematics of the probability representation.

As I see it, Skyrms's point is that although premise 2 may have some problem cases, they will be epistemologically special update situations, and not the ordinary ones we normally consider, so a relevant revision of premise 2 will still leave some instances of this argument intact. Premise 4 also seems unproblematic. Thus we should focus on premise 3.

And indeed, premise 3 has serious problems. There is no need for conditional probability to be understood in terms of Kolmogorov's ratio. Many other accounts of conditional probability have been proposed that extend this account to cases where $P(B) = 0$. Perhaps the simplest modification is described in Popper 1955, according to which conditional probability is a primitive two-place function not defined in terms of unconditional probability, but freestanding. Popper's axioms guarantee that whenever $P(B) \neq 0$, the standard relations still hold, but just add the claim that $P(A|B)$ is always defined. Instead of having $P(A|B) = P(A \cap B)/P(B)$, we just have $P(A|B)P(B) = P(A \cap B)$, as I originally stated in the introduction, which can hold even when $P(B) = 0$. This sort of account is argued for by Hájek (2003), among others. Another account, quite similar to Popper's, is discussed in Rényi 1970. And in fact, although Kolmogorov (1950) stipulates the definition used in this argument in chapter 1, in chapter 5 he presents another alternative, different from the ones due to Popper and Rényi. (I argue for this account in Easwaran 2008a.) Two such accounts are compared by Seidenfeld, Schervish, and Kadane (2013).

There are a variety of options available, so there is no reason for the notion of conditional probability and its role in updating to demand **Regularity**. The ratio account is popular because it allows conditional probabilities to be defined entirely in terms of the unconditional credence function—the alternatives that I mention require in addition some sort of primitive *conditional* credence function. But Hájek argues

---

niques related to the AGM model of belief revision (as introduced in Alchourròn, Gärdenfors, and Makinson 1985), or one of its competitors. This gives us some epistemological considerations in favor of modifying premise 2—perhaps it ought to be prefaced with, "In any learning experience that doesn't involve giving up any doxastic certainties, . . . " But this modification is of no relevance to the argument for **Regularity**.

at length that the concept of conditional credence is at least as fundamental as that of unconditional credence, so that this mathematical parity is epistemologically significant and not just a quirk of the formalism.

Defenders of **Regularity** point out that some of these alternatives (and in particular, Popper functions) have a close connection to hyperreals, and thus suggest that they aren't really alternatives. For instance, Vann McGee (1994, 180) says,

> One approach, developed by Skyrms (1980) and Lewis (1980) is to use a nonstandard [hyperreal-valued] probability assignment in which those epistemically possible propositions that would ordinarily be assigned 0 probability are instead assigned infinitesimal probabilities. . . . The other approach, developed by Karl Popper, is more direct. . . . We shall see that these two approaches come to the same thing.

He then demonstrates that for every nonstandard-valued probability assignment, the restriction of the conditional and unconditional probability values to their "standard parts" gives a Popper function, and that every Popper function can be achieved in this way. However, the nonstandard-valued probability assignment corresponding to a given Popper function is highly nonunique—the hyperreal representation of an agent's doxastic state is far more fine-grained than the Popper-function representation corresponding to it.[9] Thus, although there is a connection between these two representations, my arguments from section 5.4 will suggest that this extra level of fine structure in the nonstandard-valued probability assignment isn't real. It can be used for a purely mathematical description of the Popper function, but one shouldn't read this extra representational power as meaning anything about the actual credences. The connection between these two options is not as tight as McGee initially claimed.

Given that there are many available accounts of conditional probability that allow for conditionalization on propositions with credence 0, for this argument to work, the defender of **Regularity** must give a non-question-begging argument in favor of analyzing conditional credence *exclusively* in terms of Kolmogorov's ratio. Absent any such argument, the most this line of reasoning can show is that there should be some way to coherently update on any doxastically possible information. If we assume

---

9. In fact, there are distinct hyperreal probability assignments that correspond to the same Popper function and yet actually give rise to different decision-making behavior on the part of the agent. Halpern 2010, 168.

additionally that conditionalization is the way to update, then this tells us that the conditional credences should be well defined and should themselves form a coherent probability function. One might use the converse of premise 1, and a premise claiming that any physical, metaphysical, or logical possibility can be learned (perhaps with an exception for claims like "I do not exist"), to give an argument for some sort of transmodal connection. But none of this gives any support to **Regularity** itself.

2.2. Stubbornness

The second argument also proceeds from the rule of updating by conditionalization, but considers A rather than B in $P(A|B)$. As Lewis (1980, 268) says:

> [**Regularity**] is required as a condition of reasonableness: one who started out with an irregular credence function (and who then learned from experience by conditionalizing) would stubbornly refuse to believe some propositions no matter what the evidence in their favor.

Similarly, Skyrms (1980, 74) asks, "How can a proposition of prior probability zero come to have a posterior probability different from zero?"

As I understand the implicit argument, it starts with the mathematical fact that if $P(A) = 0$, and $P(B) \neq 0$, then $P(A|B) = 0$.[10] Thus, if an agent updates only by repeated conditionalization, and starts with $P_0(A) = 0$, then at every time t, $P_t(A) = 0$, so the agent will stubbornly refuse to believe A, no matter what the evidence. Timothy Williamson (2002, 214) gives a similar version of this argument as a reason not to accept the Bayesian picture of probability.

I will formalize the relevant argument thus:[11]

1. $P(A \cap B) = 0$ when $P(A) = 0$.
2. When an agent learns B, she replaces her credence $P(A)$ with $P(A|B)$ for every proposition A.

---

10. This is a consequence of the claim that $P(A \cap B) = P(A|B)P(B)$ together with the fact that $P(A \cap B) \leq P(A)$. On the standard ratio analysis, it takes the apparently stronger form that if $P(A) = 0$, then $P(A|B) = 0$ if it exists at all.

11. This argument makes use of the notion of "high credence." Intuitively, this should mean something like "high enough for belief," but everything about the argument, and my response to it, will work equally well if this is interpreted as "greater than 0.99999," or "greater than 0.5," or even "greater than 0.00001."

3. $P(A|B)$ is defined as $P(A \cap B)/P(B)$, and thus is 0 or unde-fined when $P(A \cap B) = 0$.
4.   Therefore, if an agent has credence 0 in A, then she will never have high credence in A no matter what evidence B she learns.
5.   For any reasonable agent, and any doxastically possible prop-osition A, there is some evidence B such that learning B would give the agent high credence in A.
6.   Therefore, a reasonable agent does not have credence 0 in any doxastically possible proposition.

As in my previous argument, I will reject premise 3. All the propos-als mentioned above on which $P(A|B)$ can be defined when $P(B) = 0$ allow it to take on any value between 0 and 1 (depending on the circum-stances), even if $P(A)$ was 0. Thus, an agent can come to have high cre-dence in A, as long as she learns some other proposition B whose initial credence was also 0.

A defender of this argument might claim that if $P(B) = 0$, then B can't be the evidence in an update. After all, most examples of minuscule propositions (to be described in section 3) involve infinite precision and may be beyond human observational capacities, so perhaps they can never constitute an agent's evidence. So if those are the only doxastically possible propositions that get probability 0, then the argument could be repaired by adding a premise that propositions with credence 0 are never learned as evidence.

But if that's right, and it's impossible for humans to learn this type of proposition as evidence, then "stubbornness" seems much less prob-lematic—if something can never be learned as evidence, then it doesn't seem stubborn to refuse to give it high credence when learning other things. So the defender of this argument faces a dilemma: either prop-ositions with credence 0 can be evidence, in which case premise 3 is false; or they can't, in which case stubbornness is reasonable, so premise 5 is false. Either way, one of the premises is false, so this is no sound argument for **Regularity**. (Even conceding the conclusion, one of these premises must be false.) A defender of infinitesimals might concede this point, but still object that it is strange that only propositions with credence 0 can provide enough evidence for an agent to have high credence in other propositions with credence 0. But consider the following more general argument:

1. $P(A \cap B)$ is minuscule when $P(A)$ is minuscule.
2.   When an agent learns B, she replaces her credence $P(A)$ with $P(A|B)$ for every proposition A.
3. $P(A|B)$ is minuscule when $P(A \cap B)$ is minuscule and $P(B)$ is not minuscule.
4.   Therefore, if A is minuscule, then the agent will never have high credence in A unless she learns some B that is also minuscule.

This argument is just a modification of the first part of the above argument, but with the notion of probability 0 generalized to the notion of being minuscule (that is, less than any positive standard real number). This argument is valid, and all the premises are accepted by Lewis, Skyrms, and other defenders of **Regularity** that appeal to hyperreals as the values of credences for minuscule propositions (and even by many defenders of alternative versions of **Regularity** that don't use Robinson's hyperreals). In particular, for premise 3 to fail, there would have to be a situation in which $P(A|B)$ and $P(B)$ are both standard positive real numbers, and yet $P(A)$ is minuscule—but this would mean that either $P(A \cap B) \neq P(A|B)P(B)$ or $P(A \cap B) > P(A)$.

   Thus, defenders of hyperreals face the same issue for minuscule propositions that the opponent of **Regularity** does with probability 0. They must offer the same sort of resolution, where only minuscule propositions can provide enough evidence for one to believe other minuscule propositions. The only way to get around this is to either reject conditionalization, or revise one of the basic laws of probability for $P(A \cap B)$, either of which would destroy this argument for **Regularity**.[12]

2.3. Dutch Books

Skyrms gives a third argument for **Regularity** that is not shared by Lewis.[13] This argument extends the basic "Dutch book" argument for probabilism. The basic argument shows that if an agent's degrees of belief fail to satisfy the probability axioms, then she is vulnerable to a "Dutch book"—

12. I thank Greg Novack and Mike Titelbaum for pressing me on this point and making me realize that I should spell out the full parallel argument for the defenders of **Regularity**.
13. Versions of this argument were also given much earlier, in Kemeny 1955, Shimony 1955, and Stalnaker 1970, where they refer to **Regularity** as "strict coherence." I stick with Skyrms and Lewis just because they are the ones referred to by contemporary philosophers who defend **Regularity** and hyperreal credences.

a set of bets such that she considers each one individually fair or favorable (because its price is less than or equal to her degree of belief in the relevant proposition), and yet the whole set collectively guarantees her a loss. Since any rational agent views a guaranteed loss as neither fair nor favorable, then (bracketing some assumptions about evaluating a combination of bets by combining the evaluations of the individual bets) there is an inconsistency in her values.[14]

Similarly, Skyrms (1980, 74) suggests that if we allow for propositions of credence 0 to be doxastically possible, then "if we interpret probability as a fair betting quotient there is a bet which we will consider fair even though we can possibly lose it but cannot possibly win it." That is, if an agent's degree of belief in *A* is 0, then she will view as fair a bet that costs $0 with a payoff of $1 if *A* is true. However, if she is not absolutely certain that *A* is false, and she sells such a bet, then she is in a situation in which she has no possibility of making money, but a possibility of losing money, which she must surely regard as an unfavorable position, rather than a fair or favorable one.

I will formalize the argument thus:

1. Any rational agent evaluates a bet on *A* at a price equal to [her credence in *A* times the stakes] as fair to buy or sell, evaluates a bet at any lower price as favorable to buy, and evaluates a bet at any higher price as favorable to sell.
2. No rational agent evaluates a bet as fair or favorable if it gives some doxastic possibility for her to lose and no possibility to gain.
3. Selling for $0 a bet on *A* with positive stakes results in losing if *A* is true and has no possibility of gaining.
4. Therefore, no rational agent has credence 0 in any doxastically possible proposition.

The standard Dutch book argument for probabilism goes as follows:

1. Any rational agent evaluates a bet on *A* at a price equal to [her credence in *A* times the stakes] as fair to buy or sell,

14. Some authors present the problem of vulnerability to Dutch books as a sort of pragmatic irrationality, involving the fact that an agent who is actually willing to accept each of these bets is practically irrational since she faces a guaranteed monetary loss. The interpretation I give in terms of inconsistency of values is suggested by Skyrms (1987) and Christensen (2001), and I think it is more compelling. But nothing depends on which interpretation is used.

evaluates a bet at any lower price as favorable to buy, evaluates a bet at any higher price as favorable to sell, and evaluates a combination of bets as fair or favorable if she evaluates each individual bet as fair or favorable.

2. No rational agent evaluates a combination of bets as fair or favorable if she is certain that they would collectively cause her to lose.

3. An agent's credences satisfy the probability axioms iff there is no finite collection of bets with fair or favorable prices such that she is certain they would collectively cause her to lose.

4. Therefore, a rational agent's credences satisfy the probability axioms.

There are many well-known problems involving the first premise of these Dutch book arguments (Hájek 2005, 2008). Thus, the opponent of **Regularity** could just reject this argument along with the standard Dutch book argument, by just rejecting anything resembling the first premise of either argument. But I will not take this route—I will respond to this argument in a way that is open for defenders of Dutch book arguments.

The first possibility for rejecting the argument for **Regularity** while keeping the standard Dutch book argument is to look at the difference between the second premises—in the standard Dutch book argument, there is a doxastic necessity of loss, while in the one for **Regularity**, there is only a doxastic possibility of loss. This allows room for saying that necessary loss is problematic in a way that the possible loss is not. But it seems to me that this is a bullet-biting response—it would say that a rational agent can accept a possibility of loss with no offsetting possibility of gain.

Instead, I will reject the first premise in each of these arguments and accept only a weaker premise about favorability, rather than fairness:

1′. Any rational agent evaluates a bet on A at any price lower than [her credence in A times the stakes] as favorable to buy, evaluates a bet at any higher price as favorable to sell, and evaluates a combination of bets as favorable if she evaluates each individual bet as favorable.

In this version of the premise, I have made no assumption at all about whether an agent evaluates a bet at *exactly* her credence times the stakes as fair, favorable, or unfavorable.[15] With this modification, if the first argu-

15. In fact, some have suggested that one must evaluate bets at precisely this price as

ment is to be valid, its conclusion must be weakened to "Therefore, no rational agent has credence *less than* 0 in a doxastic possibility." This is no longer **Regularity** itself but rather a trivial consequence of the probability axioms.

However, the standard Dutch book argument can be made valid by appealing to the slightly stronger theorem that is also true:

> $3'$. An agent's credences satisfy the probability axioms iff there is no finite collection of bets with favorable prices such that she is certain they would collectively cause her to lose.

For any collection of bets with fair prices such that the agent is certain they would collectively cause her to lose some positive amount, we can alter the prices by a tiny fraction of this amount, to give a collection of bets with favorable prices that have the same Dutch book property. Thus, replacing premise 1 by $1'$ doesn't jeopardize the standard Dutch book arguments, so an opponent of **Regularity** can preserve the standard Dutch book argument if she is so inclined.[16]

----

unfavorable. Smith (1961, 5) is an early example. I will be agnostic on this point and leave open the possibility that something beyond the numerical value of one's credences is used to evaluate bets at exactly this price, so that some count as fair, some count as favorable, and some count as unfavorable. Giving a full decision theory for cases with expected value of 0 is beyond the scope of this essay.

16. As it turns out, I think there is some motivation for defenders of countable additivity to make this modification of the argument. For any finite or infinite cardinality κ, there is a collection of κ-many bets at *fair* prices that collectively make it doxastically necessary that the agent will lose if his or her credences do not satisfy κ-additivity. Thus, with the "fair or favorable" version of the argument, we seem to get an argument for additivity of arbitrary collections of propositions. But while some probability theorists support additivity for countably infinite collections of propositions, they don't generally support additivity for uncountably infinite collections of propositions since this would rule out uniform distributions on uncountable sets, just as countable additivity rules out uniform distributions on countable sets. Thus, the defender of the "fair or favorable" version of the argument needs to either distinguish between finite and countable collections of bets (for finite additivity) or between countable and uncountable collections of bets (for standard countably-additive probablism).

However, if we don't assume that agents will evaluate bets *exactly* at the specified price as fair, as in my modification, then we get a nonarbitrary reason to support countable additivity but not uncountable additivity. The reason there is no support for uncountable additivity is that any favorable price for selling must be positive, and the sum of uncountably many positive numbers is always infinite. Thus, selling uncountably many favorable bets on pairwise incompatible propositions never results in a guaranteed loss. But for the countable case, the Dutch book still works. If the agent buys a bet on the union of a sequence of propositions for ε less than his or her fair price and sells each bet on the nth

The defender of this argument for **Regularity** thus has to argue for the stronger premise 1 rather than the weaker 1′. One motivation would be to say that there *must* be some price at which a bet is evaluated as fair—not every price should be one that is favorable for buying or favorable for selling. But this assumption is not available to the defender of **Regularity** if the bets are monetary—money (and, by the argument I will give in section 5.4, utility too) comes only in real gradations, so any positive price for the bet is higher than the credence in a minuscule proposition and is favorable for selling, while a price of 0 is favorable for buying, and no real price is exactly fair. Only if the prices of bets themselves can have numerical values that are not standard real numbers can one maintain that every bet has a price that is exactly fair.

But this brings us to the first instance of the "numerical fallacy." When a bet has a positive real expected value, these premises say it should be evaluated as favorable. Premise 1 goes further and says that a bet with expected value 0 should be evaluated as fair. This seems plausible if we assume that the numerical expected value of a bet alone must tell us whether it is fair, favorable, or unfavorable. But if we allow that nonnumerical features of the mathematical representation of an agent's doxastic state might matter as well, then 1′ looks better. In cases where the expected value of a bet is exactly the same as the status quo, some nonnumerical feature may serve as a tiebreaker. For any proposition in which an agent has credence 0, the expected value of a bet at price 0 is exactly the same as the status quo, no matter whether the bet is bought or sold. However, the fact that in one case the agent has a possibility of winning with none of losing, and that in the other case the agent has a possibility of losing but none of winning, allows the agent to determine that one is favorable and the other is unfavorable, and neither is precisely fair.

And in fact, there are other motivations for thinking that actions might be evaluated by using some tool beyond numerical expected value. For actions with infinitely many possible outcomes, some expected values are infinite or undefined, which means that something other than numerical comparison is necessary to evaluate which are better or worse (Nover and Hájek 2004; Colyvan 2008; Easwaran 2008b). Similar issues arise if

---

proposition in the sequence for $\varepsilon / 2^n$ more than his or her fair price, then the total result will be exactly as if he or she had bought and sold the bets exactly at his or her fair price—which would result in a Dutch book if his or her fair prices aren't countably additive. This Dutch book parallels the one Jon Williamson (1999) gives. Exactly this point about countable versus uncountable additivity is made by Skyrms (1992, 218).

credences can be imprecise. (Adam Elga [2010] argues that there is *no* reasonable decision theory for imprecise credences, but any appropriate response to his argument will have to involve more than just single numerical expected values.) We can keep the assumption that having a greater expected value is sufficient for being preferable, but these cases already show that it is not necessary. Thus, we should reject premise 1 in both arguments and replace it by 1′. The Dutch book argument for probabilism can be saved by replacing its premise 3 by 3′. But the Dutch book argument for **Regularity** can't be saved without weakening its conclusion to a triviality.

2.4. *"What 0 Means"*

The final argument I will consider is rarely given explicitly, but I suspect that it is the intuitive motivation that most defenders of **Regularity** have for believing it. However, I will show that it too is an instance of the numerical fallacy. The basic idea is related to the faithfulness of mathematical representations. A statement of the idea is given in Williamson 2002, 213: "For subjective Bayesians, probability 1 is the highest possible degree of belief, which presumably is absolute certainty."[17]

   The main idea is that the degree of belief function is a measure of the agent's doxastic state with respect to a proposition. This function measures propositions on a scale from 0 to 1 and assigns the value 1 to doxastic necessities and 0 to doxastic impossibilities. If the function were to assign the value 1 to some proposition other than a certainty, or 0 to some proposition other than a doxastic impossibility, then this function would not properly represent the agent's attitudes because it would falsely represent her as equally confident in two propositions that she is not

17. Williamson follows this with a dramatic argument that an agent with such a high credence should be willing to sell for a penny a bet where the agent is tortured if the proposition comes out false. This example is related to the previous argument, but it also seems to prove too much—not only would it rule out having credence 1 in any proposition short of certainty, but it would also rule out credences of $1 - \varepsilon$ for small enough ε.

Because of the well-known phenomenon of risk-aversion, it seems plausible that bets with extremely large payoffs, either positive or negative, are evaluated at least partly by some means other than expected utility. Thus, methodologically, we should limit consideration to bets with small payoffs when intuitively judging the rationality of accepting certain bets. If the possible loss is held fixed at a moderate value, while possible gain becomes extremely small, then the conclusion doesn't seem implausible. The dramatization in terms of torture is a distraction—only the claim I quoted above, about the "highest possible degree of belief," is important. I thank Lina Eriksson for this point.

equally confident in (namely, one that is doxastically contingent and one that is doxastically necessary, or impossible). Thus, if a degree of belief function properly represents an agent's attitudes, then it must satisfy **Regularity**.

I will formalize the argument as follows:

1. A doxastically possible proposition is more likely for an agent than a contradiction.
2. If p is more likely than q for a rational agent, then $P(p) > P(q)$.
3. If q is a contradiction, then $P(q) = 0$.
4. Therefore, for a rational agent, if p is doxastically possible, then $P(p) > 0$.

It is clear given my previous discussion that I will reject premise 2 as an instance of the numerical fallacy. If $P(p)$ were the complete mathematical representation of how likely p is for an agent, then this would be reasonable. But it isn't. What we need is some mathematical relation $p > q$ that says when p is more likely than q. But this relation can depend on mathematical facts beyond $P(p)$ and $P(q)$. As described in the opening paragraph of the introduction, standard probabilism gives two further mathematical features that might be relevant—the conditional probability function $P(\cdot | \cdot)$, and the set Ω of doxastic possibilities.

If we use one of the alternative accounts of conditional probability mentioned in section 2, then there will be distinctions between propositions the agent regards as certainly false, and propositions she merely has credence 0 in. For instance, on Popper's account, if $\perp$ is a contradiction, then $P(p|\perp) = 1$ for any proposition p. (In particular, $P(\neg\perp|\perp) = 1$!) It is natural to extend this behavior to other doxastic impossibilities.[18] However, Popper's axioms allow for this to fail for doxastically possible propositions whose unconditional probability is 0. Another natural picture might suggest that $P(p|q)$ is undefined when q is doxastically impossible (perhaps because an indicative-type conditional, as conditional probability is normally taken to be, makes no

---

18. On a set-theoretic formulation, this is trivial because a doxastic impossibility and a contradiction are both represented by the same empty set. But even on a sentential formulation, we can prove that the behavior does extend this way if we assume that $P(p|q) = P(p|q \wedge \top)$ whenever $\top$ is a doxastic necessity. If q is a doxastic impossibility, then $\neg q$ is a doxastic necessity, so $P(p|q) = P(p|q \wedge \neg q) = P(p|\varnothing) = 1$.

sense when the antecedent is impossible), but is defined in all other situations.

On both accounts, if q is doxastically impossible, but p isn't, we will have $P(p|p \cup q) = 1$ and $P(q|p \cup q) = 0$. Thus, we can define $p > q$ as meaning that $P(p|p \cup q) > P(q|p \cup q)$ and get an ordering that validates premise 1 while falsifying premise 2. The relevant distinction can be captured in the conditional probability function rather than in the values of the unconditional probabilities. On this approach, conditional credence would turn out to be more fundamental than the $>$ relation. Such a view has been argued for by Hájek (2003). (Note that some, but not all, of the arguments there presuppose the failure of **Regularity**.)

On an approach where propositions are represented by sets of possibilities, the distinction can also be captured in the set structure of the propositions. Recall that the complete representation of the agent's credal state is the triple $(\Omega, \mathcal{F}, P)$ and not just P by itself. With this representation, we can draw the distinction between doxastically impossible propositions (which correspond to the empty set) and others (which are nonempty, even though their probability may be 0). There are many situations where it is sufficient to consider the numerical values of P and ignore the mathematical information contained in Ω and $\mathcal{F}$, but the argument under consideration only works if we assume that P is *always* sufficient. On this picture, we might say that $p > q$ iff ($P(p) > P(q)$ or $q \subset p$). On this account, $>$ is not a total ordering, but again it validates premise 1 and falsifies premise 2.

In either case, the argument fails because premise 2 is false. Both proposals for $>$ accept the converse of premise 2 (if $P(p) > P(q)$, then $p > q$). There would be a certain elegance to accepting premise 2 as well. But it is certainly not essential to a proper mathematical theory of $>$, once one considers the nonnumerical aspects of the mathematical representation.

In fact, I will show in section 5 that the use of hyperreals to defend **Regularity** leads to problems here. Although it can save the claim that if $p > q$, then $P(p) > P(q)$, it violates the converse—there are propositions where $P(p) > P(q)$, and yet intuitively, $p \not> q$! The new numerical representation overshoots the mark, and thus equally fails to faithfully represent the agent's credences. There may be purposes for which extraneous structure is no problem, just as there may be other purposes for which some missing structure is no problem. But if a mathematical representation of credence is not to be a misrepresentation, then missing numerical structure can be made up by considering nonnumer-

ical mathematical structure, while extra numerical structure poses a more serious problem.

However, before I can demonstrate this extra structure in the hyperreals, I must present the problem of minuscule propositions and explain the hyperreals that Lewis and Skyrms use to respond to them.

3. Minuscule Propositions

Defenders of **Regularity** have been forced to concede that some doxastically possible propositions have credence less than $1/n$ for any natural number n. It would be an interestingly bold position to deny that rational agents have credences in the propositions I will discuss in this section, or to deny that any agent may rationally treat them as doxastic possibilities, as a defender of **Regularity** without hyperreals, or another theory of infinitesimals, must do. To make clear that these propositions must have an extremely small probability (whether 0 or infinitesimal, or perhaps otherwise described), I will call such propositions "minuscule." For convenience, I will also use the term "minuscule" as a term for numbers that are less than $1/n$ for any natural number n, which are the probability values of minuscule propositions. (I will reserve the term "infinitesimal" for minuscule values that are nonzero, although some authors include 0 as an infinitesimal.)

As an example, consider a situation in which a dart is being thrown at a dart board, and consider the proposition that the center of the dart lands on the vertical line that precisely bisects the board. I claim that this proposition is a minuscule one if the agent treats the throwing of the dart as uniform, so that the probability that it lands in any given region is proportional to the area of that region.

Consider the strip around the central vertical line that is exactly $1/n$ as wide as the board is—the probability that the dart lands in this region is $1/n$, and this region entirely contains the central vertical line. Thus, the probability that the center of the dart hits the center line must be less than $1/n$ for every n. But on the other hand, it seems clear that this *could* happen, and so it seems like it should be doxastically possible—after all, nothing is special about this line to prevent the dart from hitting it, and every vertical line should be treated equally. Of course, one might worry about infinitely precise centers of darts, and the requirement that the agent distribute his credence uniformly over the board for the positions that it might hit. But as Hájek (2003) repeatedly points out, as long

as a rational agent could possibly have positive credence in this setup, our account of credence should allow for it.

For another example, consider a fair coin that will be flipped infinitely many times, and consider the proposition that this coin comes up heads on every single flip. On the one hand, the probability of this proposition must be no more than $1/2^n$ for any n because that is the probability that the first n flips come up heads, which is entailed by this proposition. But on the other hand, it seems that this proposition really does describe a doxastically possible outcome of the sequence of coin flips. By the symmetry of the situation, any two sequences of coin flips should be treated similarly—there is no reason based on this setup to treat some sequences as possible and others as impossible. Of course, one might still have doubts about this situation because of the requirement that the agent believe the infinite sequence of coin flips has some possibility of actually occurring—but a denier of minuscule propositions must say that such things are not just nonactual, but doxastically impossible for every rational agent.

For a more realistic example, consider the proposition that the speed of light is exactly $2.998 \cdot 10^8$ m/s.[19] Although our measurements may have made us absolutely certain that the speed of light is not $2.997 \cdot 10^8$ m/s, or $2.999 \cdot 10^8$ m/s, there is at least some range of values that have not been ruled out by any of our experimental observations. And it seems that there is in fact some precise fact of the matter as to what this speed is.[20] However, for any n, we can surely come up with n disjoint intervals (not necessarily of equal width), such that a rational agent could regard it as equally likely (or almost equally likely) that the true value of

19. I have been told that the speed of light actually has a stipulated value that is used as part of the definition of the meter and the second. Thus, properly speaking, I should substitute some other physical constant (like the fine-structure constant, or the exponent in some gravitational law) that has a value independent of our conventional choice of units. Further, if our theories suggest that such "constants" can actually change in value over time, then consider instead of the theory that it has a specific value, the theory that it evolves according to a particular function over time.

Maher (1990, 387–88) gives a version of this argument together with a historical claim that this accurately describes Cavendish's opinions regarding the exponent in the law of electrostatic attraction.

20. Even if some physical theories might suggest that space-time is discrete, in a way that means there can be no such infinitely precise fact of the matter, surely we are not *completely* certain that some such theory is true. Or we can consider the beliefs of some scientist from a previous century that couldn't rationally rule out theories according to which a precise value exists.

the speed of light is somewhere in one of those intervals. Each of these intervals must have credence $1/n$, so the credence for any *particular* value contained in one of the intervals must be no greater than $1/n$. Thus, any precise specification of the value appears to be a physically realistic proposition that is doxastically possible, but for which the probability must be less than $1/n$ for any n.

Based on these three examples, and the ease of generating more like them, we should agree that there are minuscule propositions.

Defenders of **Regularity** claim that minuscule propositions must not be assigned probability 0, so if they want numerical values for the probability function, then they need some theory of infinitesimals. For Skyrms, Lewis, and their followers, Robinson's hyperreals play this role. If one instead rejects **Regularity**, one can just say that minuscule propositions have credence 0 and use objects other than the numerical probabilities, like the set of doxastic possibilities, or the conditional credence function, to represent the relevant differences.

4. What Are Robinson's Hyperreals?

In order to discuss the reasons I think that hyperreals won't be able to do the work that is demanded of them, it will be important to be clear about how they work mathematically. Skyrms and Lewis cite Bernstein and Wattenberg (1969) for a mathematically sophisticated account of how this could work, but they don't consider the details explicitly themselves. However, these details give rise to the problems I will discuss later, so I rehearse them here. (More thorough discussions are given in Luxemburg 1973 and Robinson 1996.)

Robinson's hyperreals form a mathematical structure that satisfies the complete first-order theory of the real numbers and includes a copy of the standard real numbers, together with some infinitesimals—positive elements that are smaller than any positive standard real number. Because these structures satisfy the complete first-order theory of the real numbers, much of our standard reasoning carries over to them. But it is important to note that this is only the *first-order* theory—we must be careful about statements involving *sets* of real numbers.

The proof that such structures exist is not especially complicated. It relies on a familiar result from first-order logic known as the Compactness Theorem. This result states that if Γ is a set of sentences in a first-order language, and if every finite subset of Γ has a model, then Γ has a model. There are two standard proofs of this result—importantly, both

make use of nonconstructive methods, based on the Axiom of Choice. The first proof appeals to Gödel's Completeness Theorem, which states (nonconstructively) that a set of sentences has a model iff it is impossible to derive a contradiction from these sentences. Thus if Γ didn't have a model, then it would be possible to derive a contradiction from it. But since any derivation uses only finitely many sentences, this derivation would use only some finite subset Γ_0—so this finite subset Γ_0 would have no model. The second proof is preferred by model theorists, who try to avoid reference to syntactic derivations whenever possible. On this proof, the model for Γ is constructed directly from the models of its finite subsets by means of an "ultraproduct" construction, which relies on the Axiom of Choice to (nonconstructively) provide a suitable "ultrafilter." (See Chang and Keisler 1990, chapter 4, or any other model theory textbook, for details.)

Given the Compactness Theorem, Robinson's result is fairly straightforward. Let $\mathcal{L}$ be a first-order language for talking about the real numbers that includes a name for each real number, and add to it a new constant c. Let Γ be the set of all sentences in $\mathcal{L}$ that are true about the real numbers (including particular sentences like "$2 < 5$" and general ones like "$\forall x \, (x = 0 \lor \exists y \, (x \cdot y = 1))$"), together with the sentences "$c > 0$" and "$c < K$" for each K that names a positive real number. Now it is clear that every finite subset of Γ has a model—one such model will just interpret all of $\mathcal{L}$ in the standard way and interpret c as a positive real number that is smaller than any positive real number whose name is mentioned in this finite subset. But then the Compactness Theorem guarantees that Γ itself must have a model.

Because Γ includes all sentences of $\mathcal{L}$ true in the standard real numbers, the model satisfies the complete first-order theory of the real numbers. Because $\mathcal{L}$ includes names for each real number, the model includes a copy of the standard real numbers. And this model must have an interpretation for "c," which must be positive (because Γ contains "$c > 0$") and smaller than any positive standard real number (because Γ contains each $\ulcorner c < K \urcorner$). Thus, the model contains at least one infinitesimal. (Of course, c is not the only such infinitesimal—for example, $2c$, $c/5$, and c^2 will be among the infinitely many others. There will also be "infinitely large" numbers like $1/c$.)

A point that will become important later about this proof is that it is nonconstructive—both proofs of the Compactness Theorem make use of nonconstructive methods that go beyond Zermelo-Fraenkel set theory, in the completeness case to give a maximal consistent set of sentences

extending a given consistent set, and in the ultraproduct case to give a nonprincipal ultrafilter to use as the base for the ultraproduct.[21] In fact, *no* constructive proof (either of the Compactness Theorem or of the existence of hyperreal structures) is possible—there is no way to exhibit a specific structure that provably shares the first-order properties of the reals and contains infinitesimals.[22]

First-order equivalence is sufficient for the basic theory of probability because it means that the standard results about addition, multiplication, and ordering still apply, including things like commutativity, associativity, existence of multiplicative inverses, and the preservation of order under multiplication or division by positive numbers. However, for some more advanced results in probability, we need second-order and higher-order expressive power, to talk about sequences, limits, and notions like topology and measurability for sets of reals.

Fortunately, even though the Compactness Theorem only applies to first-order theories, much of this higher-order work can be expressed in a first-order set theory, so that the Compactness Theorem can still be applied. One theory that will suffice is full Zermelo-Fraenkel set theory with the Axiom of Choice, but there are far weaker theories that will also suffice, such as a sort of Russellian theory of types built up off the real numbers. (See Burgess 2005 for discussions of some such systems. Section 4.4 of Chang and Keisler 1990 explicitly discusses the construction of models that include the real numbers, infinitesimals, and a theory of sets.) At any rate, we can let Γ be the set of all true first-order sentences in this much larger theory, together with sentences about some constant c that entail that it must be an infinitesimal, and the result will again be a model of this large theory that manages to include infinitesimals, while still making sense of all the constructions the original theory could talk about.

But because we are dealing only with a *first-order* theory, and not a true second-order theory, there will be some oddities with this model—for instance, not every subset of the domain will be represented by one of the objects that this model calls a "set." A true second-order theory would

21. The Compactness Theorem and the existence of the relevant ultrafilters are both equivalent to the Boolean prime ideal theorem, which is weaker than the Axiom of Choice, but still independent of Zermelo-Fraenkel set theory (Moore 1982).

22. Kanovei and Shelah (2004) prove that given the Axiom of Choice, there is a sentence that defines a particular hyperreal structure. But they also point out, as I will in section 5, that given ZF set theory without the Axiom of Choice, it is consistent that no hyperreal structure exists.

quantify over *all* subsets of the domain, while a first-order theory for talking about sets has a special domain of objects that play the subset role. Nothing in a first-order theory can guarantee that all subsets are represented there. Thus, there will be a distinction between the "internal" sets that the model represents with these objects and "external" sets that aren't represented by anything in the relevant model.[23] This distinction will become important later. (For more on the distinction between things that can be properly expressed in a first-order theory of sets and things that require true second-order logic, see, for instance, Boolos 1984.)

At any rate, the construction gives a model that behaves like the real numbers, includes infinitesimals, and can talk about sets and sequences. Thus, the model has the expressive power needed for probability theory. Skyrms, Lewis, and their followers hope that by using one of these models, rather than the standard real numbers, we can save **Regularity** by applying the infinitesimal values to minuscule propositions. However, the worries about external sets and sequences give some cause for concern, and I will eventually show that they doom the approach.

5. There Are No Hyperreal Credences

An important recent argument against this use of infinitesimals is Williamson 2007. In this paper, Timothy Williamson argues that infinitesimals can't be used for the case of the minuscule proposition of a fair coin coming up heads on all of its infinitely many flips. Williamson (2007, 4) says that "infinitesimal probabilities may be fine in other cases, but they do not solve the present problem." Williamson's argument rules out any such use of infinitesimals, given some weak ordering assumptions and some intuitions about the comparative probabilities of certain minuscule propositions.

I present Williamson's argument in the first subsection of this section and suggest a response in the second subsection. In the third subsection, I analyze what goes wrong with this response, and use it to show that no calculation will yield a hyperreal credence for the kind of proposition involved in this case. The fourth subsection gives the final

---

23. An example of such an "external" set is the set F of all finite numbers. If this set existed in the model, then it would satisfy the following three first-order properties: $\forall x(x < 1 \rightarrow x \in F)$, $\forall x \forall y((x \in F \wedge y \in F) \rightarrow (x + y) \in F)$, $\exists x(x \notin F)$. However, in the standard model, it is clear that no such set exists. Thus, since the hyperreal model satisfies all the same first-order formulas as the standard model, it must not include such a set F.

argument that shows that credences in these propositions can't be hyper-real even if they are assigned in some noncalculational way. Unlike Williamson's argument, my argument doesn't rely on intuitions about equiprobability, but only on the supervenience of credences on the physical world. My conclusion applies only to Robinson's hyperreals, rather than other theories of infinitesimals, but it shows that hyperreals can't be the credences of *any* ordinary proposition (that is, a proposition that doesn't itself explicitly mention hyperreals, or similarly complicated mathematical objects), not just the one about infinitely many coin tosses.

5.1. *Williamson's Argument*

Williamson's argument proceeds as follows. Consider two fair coins that will be flipped countably many times—for definiteness, say that they will be flipped once per second, assuming that seconds from now into the future can be numbered with the natural numbers. Let the first coin be flipped starting at second 1, while the other coin is flipped starting at second 2. Let A_1 be the event that the first coin comes up heads on every single flip, A_2 be the event that the first coin comes up heads on every flip after the first, and B_1 be the event that the second coin comes up heads on every flip. By the symmetry of the situation, we might judge that $P(A_1) = P(B_1)$ because it shouldn't matter when exactly the flips occur, if they occur in the same sort of sequence. However, we might also judge that $P(B_1) = P(A_2)$ because these are corresponding sequences of flips that happen at the same moment. But $P(A_2) = 2P(A_1)$ because A_2 is independent of the first coin coming up heads on its first flip, which has probability $1/2$. So $2P(A_1) = P(A_2) = P(B_1) = P(A_1)$. Subtracting $P(A_1)$ from both sides, we get that $P(A_1) = 0$. This argument works in the hyperreals because the calculation was expressed entirely in the language of first-order arithmetic.

As mentioned at the end of section 4, this is a case where the advocate of hyperreals gets too much structure. We seem to have the intuitions that $A_2 > A_1$, and yet $A_2 \not> B_1$ and $B_1 \not> A_1$. There is no way to preserve these intuitions if $>$ must correspond directly to something numerical, which presumably must give a linear ordering. No matter what values we have for $P(A_1), P(A_2), P(B_1)$, as long as $P(A_2) > P(A_1)$, it must either be the case that $P(A_2) > P(B_1)$, or (as defended by Weintraub [2008]) $P(B_1) > P(A_1)$. However, on either suggestion made at the end of section 4, both of which may allow for $>$ to be a partial ordering rather

than a total ordering, the intuitions are preserved.[24] By increasing the fine-grainedness of the numerical values available, the advocate of hyper-reals (or in fact *any* purely numerical representation) has made too *many* distinctions in the probability values, rather than too few. They must thus deny at least one of the intuitions in this case, in order to get $P(A_1) > 0$.

5.2. The Response?

In fact, a defender of hyperreals seems to have an argument that the probability of an infinite sequence of heads *must* be nonzero—we seem to be able to exhibit an infinitesimal that must give a lower bound on $P(A_1)$. However, it will turn out that this response proves too much and shows that *every* infinitesimal is a lower bound, so *no* value, infinitesimal or not, could possibly be the correct value. Instead of solving the problem, this attempted response makes things worse for hyperreals. But it will help demonstrate the relevance of external sets for the applications of the hyperreals, which will show that the hyperreals can't serve the pur-pose of aiding calculation.

The argument proceeds as a sort of dual to the argument that A_1 was a minuscule proposition. Recall that in section 3, we considered the proposition that the first n flips came up heads and showed that this proposition has probability $1/2^n$, and since this proposition is entailed by A_1, the probability of A_1 must be lower.

But imagine now that the coins will be flipped not just on every second corresponding to a natural number, but also for all the seconds corresponding to the additional infinitely large "hypernatural numbers" in some specific hyperreal structure.[25] (Ignore for the moment that these

---

24. Defining $p > q$ iff ($P(p) > P(q)$ or $q \subset p$), this works because $P(A_1) = P(B_1) = P(A_2) = 0$ and $A_1 \subsetneq A_2$, while B_1 is neither a subset nor a superset of either of the other two events. Defining $p > q$ iff $P(p|p \cup q) > P(q|p \cup q)$, we have to be a bit more careful. As long as $P(A_1|A_1 \cup B_1)$ and $P(A_2|A_2 \cup B_1)$ are both undefined, this suggestion will work as well.

Williamson claims that we can't have $A_1 > \varnothing$, but his argument assumes that $B_1 \succeq A_2$. Both of the models just given show that if B_1 and A_2 can be incomparable, rather than equally likely, then it can be the case that $A_1 > \varnothing$. Williamson claims that we have an intuition that B_1 is equiprobable with A_2, but I claim that our intuition is just that B_1 is neither more nor less probable than A_2 and that we can't reliably tell the difference between these types of intuition. At any rate, the subtle differences between equiprob-ability and incomparable probabilities with the same numerical value (Williamson and I agree that these events all have probability 0) make such intuition-based arguments more difficult.

25. To show that these infinitely large natural numbers must exist, recall that the

additional flips change the case and may thus change the relevant probabilities.) Now consider the claim that the coin comes up heads on every flip up to some hypernatural N, and not just on the flips corresponding to standard natural numbers. This proposition entails that every flip in the original infinite sequence comes up heads (since the sequence up to N includes all the standard natural numbers and more), and thus $P(A_1)$ must be at least as large as its probability. But the probability of this claim seems to be $1/2^N$, which is a nonzero infinitesimal. Thus, it appears that $P(A_1)$ must be larger than some infinitesimal, and not equal to 0 as Williamson's argument suggested!

However, this argument turns out to be too powerful. Let ε be any positive infinitesimal hyperreal. Then a version of this argument will show that $P(A_1) > \varepsilon$. Since ε is infinitesimal, $1/\varepsilon$ is larger than every natural number. For any real number $x > 0$, there is an integer power of 2 between x and $x/2$. Since this is a first-order claim, the nonstandard model must satisfy it as well—when x is $1/\varepsilon$, call the relevant number 2^N. N must be a hypernatural number since otherwise 2^{N+1} would be a standard natural number larger than $1/\varepsilon$. But now consider the claim that every flip up to N comes up heads. This proposition still entails A_1, but it has probability $1/2^N$, which is greater than ε.

Thus, we see that $P(A_1) > \varepsilon$, as claimed. Since this is the case for *every* ε, this means that although any positive real number is too large to be $P(A_1)$, every infinitesimal is too small—but by definition, there is nothing smaller than every positive real number except for these infinitesimals. So *no* value is possible.

5.3. Calculations with Internal and External Sets

The problem with these arguments is that we are trying to use a *nonstandard* model to calculate the probability that every *standard* flip comes up heads. If we are using a nonstandard model that can talk about sets of numbers as well as numbers, then it turns out that the set of all standard natural numbers is an "external" set that this model can't talk about—

---

standard model satisfies the claim that for every x there is a natural number between x and $x + 1$, and also the claim that every real number has a multiplicative inverse. Since these are first-order claims, the nonstandard model must satisfy them as well. If ε is some infinitesimal, then $1/\varepsilon$ must be infinitely large—ε is less than $1/n$ for every standard natural number, so $1/\varepsilon$ must be larger than each n. But any "natural number" N between $1/\varepsilon$ and $1/\varepsilon + 1$ must be an infinitely large natural number, which we can call a "hyper-natural number."

therefore, it should be no surprise that this model can't be used to calculate a specific probability for events defined in terms of this set.

To show that the set of standard natural numbers is external, consider the normal argument that the probability that the first N flips all come up heads is $1/2^N$. This argument works by induction. If $N = 1$, then the probability that the first N flips all come up heads is clearly $1/2^1$. Now, we assume that the claim is true for N and show that it is true for $N + 1$. The next flip of the coin is fair, and thus has probability $1/2$ of coming up heads. The first N flips and the next flip are independent, and so the probability that the first N flips come up heads *and* the next one does is the product of their two probabilities, which is $1/2^N \cdot 1/2 = 1/2^{N+1}$. Thus, by induction, this must be true for all N.

But induction is a second-order principle. It says that for any *set* of natural numbers, if the set contains 1, and contains $N + 1$ whenever it contains N, then the set contains all positive natural numbers. But notice that in a hyperreal model, the set of *standard* natural numbers violates this principle since it leaves out the hypernaturals. If the language and logic used for calculations with infinitesimals (and other nonstandard numbers) has an induction principle that holds for all sets that it recognizes, then the set of standard natural numbers is not a set internal to this model, so it can't tell us anything about the probability of an event essentially involving the set of standard natural numbers, like the one Williamson is interested in. Conversely, if the model does give a way to calculate the probability of this event, then it doesn't satisfy the full induction principle, and there is no way to calculate the probability of N flips all coming up heads. Either way, the attempted response to Williamson's argument fails.

And this holds more generally, not just in the example that Williamson considers. If we use the hyperreals to calculate the probability of a proposition, then there are three possibilities. The proposition might be an "ordinary" proposition, which the language can express using only standard first-order vocabulary (such as the proposition that the first 739 flips come up heads, or that the dart falls exactly on the center line of the board). The proposition might be one that the language can express, but only using vocabulary that refers to particular nonstandard elements of the hyperreal model (such as the proposition that the first N flips come up heads, or that the dart falls within $1/N$ of the center line of the board, where N is a particular hypernatural number). Or the proposition might be one that the language can't express at all (like the proposition that every standard flip comes up heads).

In the first case, since the calculation is first-order and uses only standard vocabulary, the first-order equivalence between the hyperreals and the standard reals means that the result must be the same as if we calculated with the standard model—so the result can't be infinitesimal. In the third case, we just can't use the model to do the calculation—we need some extended technique. Only in the second case can this method assign an infinitesimal value. But these cases can't provide an argument for the use of hyperreals in describing mental states since they already presuppose that propositions involving hyperreals get credences. At any rate, the *ordinary* minuscule propositions discussed in section 3 must get probability 0. And this would mean giving up **Regularity**, which was a primary motivation for using the hyperreals in the first place.

To sum up: the argument against Williamson's assignment of probability 0 to an infinite sequence of heads failed because it tried to do a calculation on a set external to the language. And this is a general problem for the hyperreals—any proposition expressible in standard vocabulary whose probability is calculated in a hyperreal model must get a standard probability.[26]

5.4. *The Complexity Argument*

In response to these earlier points, a defender of **Regularity** might suggest that hyperreal probabilities are assigned in some language-external way that doesn't involve any calculation within the model.[27] In this section, I will show that this sort of response can't work—at least, any such assignment of hyperreal values to credences in ordinary propositions (ones that can be expressed using only standard vocabulary) will impute some structure that actual credences of physical agents themselves can't have.

Although Bayesianism concerns itself with idealized rational agents, and not the imperfect physical beings we encounter in our daily life, I claim that the essentially nonphysical nature of agents with hyperreal credences makes them irrelevant for the epistemology of physical agents. The other idealizations, of logical omniscience and the like, are

---

26. A similar argument against the possibility of infinitesimal *chances* rather than credences is given in Barrett 2010.

27. In effect, this is how the proposal in Bernstein and Wattenberg 1969 works, which is cited as a model by both Skyrms and Lewis. The *hyperreal* interval [0,1] is broken up into *N* segments, where *N* is some particular infinitely large hypernatural number, and this division is used to assign probabilities to various subsets of the *standard* interval [0,1] (without hyperreals) so that every singleton has nonzero probability.

not physically impossible, and we can make sense of a way in which actual imperfect agents might become more and more like these idealized agents.[28] These idealizations are like the ones from physics involving frictionless surfaces, and infinitely deep water for waves to travel on. But where these idealizations involve the *removal* of some limitation, the hyperreals involve the *addition* of nonphysical structure. Although I phrase my argument in terms of the actual credences of physical agents, it works just as well for any rational requirement on physical agents. Just as no agent could have a credence that was a particular hyperreal, no agent could have a rational requirement involving some particular hyperreal.

The premises and conclusion of the argument are as follows:

1. Credences supervene on the physical, in the sense that there is a function that takes as input a complete mathematical description of the physical world, and a specification of an agent and a proposition, and returns as output the number representing the credence of the agent in that proposition.[29]

2. The function relating credences to the physical is not so complex that its existence is independent of Zermelo-Fraenkel set theory (ZF).

3. All physical quantities can be entirely parameterized using the standard real numbers.

4. The existence of a function with standard real number inputs and hyperreal outputs is independent of ZF.

5. Therefore, credences in ordinary propositions (ones expressible without mention of hyperreals or closely related notions) do not have hyperreal values.

28. In fact, the statement of Bayesianism from the first paragraph of the introduction doesn't even involve this much idealization. Because of the use of doxastic possibilities, there may be logical necessities that the agent fails to have credence 1 in. Because there is no diachronic rule of updating, there is no requirement of perfect memory. There may still be some sort of idealization involved in the construction of the set of doxastic possibilities, but we can think of this set as being in a way implicitly defined by the entirety of the physical facts about the agent, even though no particular doxastic possibility is represented by any particular thing in the agent's brain.

29. In the sections defending **Regularity**, I was very interested in the nonnumerical aspects of credence, but the discussion here of hyperreal credences is just about the numerical representation.

The first two premises of the argument express a form of physical supervenience about credences — there couldn't be two worlds that agree on the entirety of the physical facts and yet are different in terms of the credence a particular agent has in a particular proposition, and the pattern of dependence is (in some very generalized sense) computable. Premise 3 is an assumption about the structure of the actual physical world. Together, these first three premises entail (given only standard set theory) that there is a function that takes a standard real number description of the universe as an input, together with a specification of an agent and a proposition, and gives that agent's credence in that proposition as an output. Premise 4 is a mathematical result that I will demonstrate further on, and it implies that this function can't take a proposition described entirely in terms of standard real numbers and give a hyperreal output, which is the conclusion of the argument.

While premises 1 and 3 might be controversial, it is only essential to my argument that they be at least plausible. The defender of hyperreal credences must deny at least one of these assumptions, which would entail doing serious physics, or philosophy of mind. It seems wrong to judge the answers to these questions based on an epistemological principle like **Regularity**. One should have independent grounds for rejecting these assumptions in order to reject my conclusion. (But see note 31 for a further concern about rejecting premise 3.)

Premise 2 can be motivated as a version of the Church-Turing thesis. This thesis states that all intuitively computable functions can be computed by Turing machines. Many authors have suggested stronger versions saying that in fact any mental process whatsoever can be simulated by a Turing machine. They have often defended this claim by appeal to an even stronger principle stating that any *physical* process can be simulated by a Turing machine. Since anything simulated by a Turing machine can be proven to exist within the framework of ZF set theory, without appeal to anything more complicated, this would entail premise 2. And of course, premise 2 is much weaker — there are plenty of noncomputable functions that can be perfectly well defined within ZF (for instance, Turing's original "halting function," and most other standard examples of noncomputable functions). Of course, the strong physical version of the Church-Turing thesis may be implausible, as argued by Copeland and Sylvan (1999) (as well as by many others). But proposed challenges to it only go a few levels up the Turing hierarchy, and don't come anywhere near the complexity level of ZF, much less beyond it. There's no clear motivation for thinking that the interpretation of physi-

cal processes as mental ones should introduce this particular type of logical complexity, unless one were already committed to using hyperreals or something similar.

My argument doesn't make any assumptions about what form the physical realization of credences takes. If an agent's mental state must include a concrete representation within her brain of every single proposition that she has credences in, together with a representation of the value of that credence, then I might be able to strengthen the conclusion to show that *no* proposition gets hyperreal credence. The defender of **Regularity** might use this sort of picture to argue that physical agents can't have credences in the sorts of infinitary propositions argued to be minuscule. But on most accounts, mental states can involve physical processes outside the agent's brain and can be dispositional in ways that don't require explicit representation of every proposition or doxastic possibility.

My assumptions are consistent with the following scenario. Perhaps an agent can have dispositional credences just by having a commitment to some kind of uniformity over her doxastic possibilities. The agent might be unsure whether a particular dartboard with width one meter is properly parameterized by the real numbers or by the hyperreals, and be committed to credence $1/2$ in each.[30] Her commitment to uniformity may be sufficient to fix her conditional credence in every proposition of the form "the exact center of the dart hits some point within x meters of the vertical line at the center of the board" to be $2x$, conditional on the board being parameterized by the hyperreals. If so, then for any particular hyperreal x, the agent will dispositionally have hyperreal credence in this proposition, even though he or she is unable to grasp the proposition directly. Of course, such a proposition is not an "ordinary" proposition since we need to use a hyperreal to even state it. But my argument shows that even on such a dispositional account of credences, physical agents don't have hyperreal credences in ordinary propositions.

Something like premise 3 is clearly essential for an argument like this to work. If the physical world really does involve magnitudes with the structure of the hyperreals, then it is not hard to see how agents might conceivably have hyperreal credences.[31] For instance, it

---

30. Premise 3 entails that in fact every dartboard is properly parameterized by the real numbers. But, as already mentioned, this fact is compatible with at least some reasonable agents being unsure of it.

31. Interestingly, although hyperreal physics might allow hyperreal credences, it may

could be that credences in particular propositions are given by the precise voltage drop across some particular neuron or synapse in the agent's brain. If voltages can be hyperreal, then these sorts of credences can be too.[32] But my argument shows that if none of the fundamental physical quantities have hyperreal structure, then even a substantially more complicated realization of credences (possibly involving not just the voltage across a particular synapse, but states of the entire network of neurons, or causal connections to the external world, or even a radical version of the extended mind hypothesis [Clark and Chalmers 1998]) can't give rise to hyperreal structure in the credences. This is why premise 1 appeals to a description of the full physical world and a specification of the agent, rather than just a physical description of the agent.

Now I will argue for premise 4. (This argument is given in the first footnote of Kanovei and Shelah 2004.) There are various results due to Robert Solovay and Hugh Woodin showing that, assuming the existence of certain large cardinals, it is consistent with ZF (without the Axiom of Choice) that there are no nonmeasurable sets of real numbers (Neeman 2010). However, given a nonstandard hyperreal number, one can define a nonmeasurable set of real numbers.[33] Thus, it is compatible with ZF set theory that there are no functions that give a nonstandard hyperreal output for any standard real-valued inputs. However, ZF together with the Axiom of Choice does prove the existence of such functions. Thus, the existence of such functions is independent of ZF, which (by premise 2) means that they are too complex to properly represent the physical

---

not suffice to save **Regularity**. If we consider the dartboard example again, then we can see that an agent's credence that the center of the dart hits *precisely* the center line of the dartboard will have to be even smaller than any of the infinitesimals available from the hyperreal structure used in physics. So we will need credences to have some even finer hyperreal structure than the physics. And I suspect that a variant of this overall argument will rule out this sort of mismatch between the physical hyperreals and the ones used for credences.

32. If chances are themselves fundamental physical quantities, rather than themselves being realized by other fundamental physical quantities, then the existence of hyperreal chances (as argued for by Hofweber [forthcoming]) could be enough for there to be hyperreal credences. But as in footnote 4, this may not save **Regularity**.

33. One version of this proof is in Luxemburg 1973, 66–67. Another version is given by Terence Tao at terrytao.wordpress.com/2008/10/14/non-measurable-sets-via-non-standard-analysis/. The construction involved is actually very similar to the Bernstein and Wattenberg construction of hyperreal probabilities—although the construction gives every singleton a nonzero probability, it also shows that some more complex sets can't get *any* probability, real or hyperreal.

manifestation of credences. Thus, the credences of physical agents in ordinary propositions are not hyperreal.

This argument is in many ways just a sharpening of the argument given in section 5 of Hájek 2003. Hájek makes the argument that infinitesimal probability assignments are "defective" because they are "ineffable." That is, we have no way to pick out which infinitesimal is the one assigned to any given proposition. My claim is the more specific one that no physical facts could make one of these infinitesimals rather than another be the credences of a particular agent. Although the Axiom of Choice guarantees that such hyperreal-valued functions exist, and although these functions are quite useful to talk about in mathematical contexts, they have mathematical structure that goes beyond that of credences.

None of this rules out a certain instrumental use of hyperreals. For instance, as mentioned in note 6, Bartha and Hitchcock (1999) use hyperreals to describe a particular standard real-valued probability function. In many cases, it may be more convenient for a theorist to describe credences by using a hyperreal-valued function than to use the set of possibilities Ω, the algebra of propositions $\mathcal{F}$, the probability function P, and a conditional probability function. But the structure of the hyperreals goes beyond the physical structure of credences, while $(\Omega, \mathcal{F}, P)$ doesn't seem to. Thus, if we want our mathematical theory to faithfully represent the structure of credences, as supposed by the arguments for **Regularity**, then we should prefer the nonnumerical structure of the standard representation over the apparent convenience of the numerical structure of the hyperreals.

6. Conclusion

I have shown that the arguments in favor of **Regularity** are all unsound. The mathematical structure of probability theory (especially when supplemented with a conditional probability function) involves several features that can do the work that nonzero values are supposed to do in these arguments. There is no need for betting behavior or comparative probability to be represented purely by individual numbers in the system. Additionally, the particular numbers endorsed by Skyrms, Lewis, and their followers (namely, the Robinson-style hyperreals) have too much mathematical structure to represent anything about any physically possible agents. The advantage that the hyperreals have is that they are first-order equivalent to the standard reals. However, they are so unlike the

standard reals in terms of second-order logic (with the distinction between internal and external sets) and set-theoretic complexity that they can't provide a faithful model of credences of the sort wanted by defenders of **Regularity**. There are of course many other number systems that are simpler than the hyperreals, which may be promising for this purpose, and I canvass several such systems in the appendix. But the basic point still stands—the mathematical structure surrounding the standard real number representation of credence appears to provide an adequate representation of credences, despite giving up **Regularity**. Any extension of this system that is intended to save **Regularity** should avoid introducing extra complexities like those of the hyperreals.

A. Appendix: Alternative Theories of Infinitesimals

Although this essay argues that Robinson-style hyperreals can't be the values of credences, there are several other frameworks that have been proposed that can reasonably be called "infinitesimal probabilities." It may be that some of these systems do a better job of representing the epistemic structure of credences than the version of the Kolmogorov picture that I defend in the main text, and so they merit further study. But the question of whether they save **Regularity** comes down to the question of whether these are considered to be standard real numbers with further mathematical structure, or whether the structure as a whole constitutes a new number system. It seems to me that this is a relatively empty terminological question, and thus research on these versions of the theory should focus on the extent to which they do or don't respect the epistemology, and not on whether or not they happen to assign a "number" that looks like 0 to a doxastically possible proposition.

A.1. Carnap

Carnap was already aware of the problem of minuscule propositions in 1960, before Robinson's construction of the hyperreals.[34] Since there was at that time no known rigorous mathematics of infinitesimals, Carnap sought to outline what such a theory ought to look like, in order for infinitesimals to do the work he wanted for probability. The draft he wrote eventually appeared posthumously as Carnap 1980.

---

34. I thank Branden Fitelson for pointing out to me the papers discussed in this section.

In this draft, Carnap posed four problems whose solution would yield a notion of infinitesimal probabilities, together with partial solutions to the first and third problems. The first problem is to lay down axioms that can be used to characterize the relations explicating the notion of one set of real numbers being smaller than another, and one set being *infinitely* smaller than another. He proposed about twenty conditions that these two relations should jointly satisfy and listed some theorems that follow from them. The second problem was to give an explicit characterization of some relation on sets of real numbers that would satisfy these axioms, which he was unable to do.

The third problem is to investigate the equivalence classes of sets of real numbers under the "same size" relation characterized in the first two problems. He carried out this project to the extent of showing that these equivalence classes could be considered as themselves constituting a number system that contained infinitesimals. The fourth problem is to give an explicit characterization of this number system and a function assigning values from this number system to sets of real numbers.

Given his partial characterizations, Carnap was able to give some characterization of what the number system might look like. In particular, just as in Robinson's later system, there would be some infinite set of infinitesimal numbers ε_i, such that for any two of them, one would be infinitely smaller than the other. However, unlike in Robinson's system, smaller infinitesimals would be "absorbed" into larger ones, so that if ε_1 is infinitely smaller than ε_2, then $\varepsilon_1 + \varepsilon_2 = \varepsilon_2$ and $\varepsilon_1/\varepsilon_2 = 0$, which in Robinson's system would happen only if $\varepsilon_1 = 0$. This might have interesting implications for the relation between conditional and unconditional probability.

In the same volume where this draft was first published, Douglas Hoover (1980) published a short note using Robinson's system (in particular, with the construction given by Parikh and Parnes [1974]) to shed some light on Carnap's problems. In particular, he showed that the Parikh and Parnes system satisfies all but two of Carnap's axioms and that those two axioms were inconsistent with the others in any case, so that nothing better could be hoped for. However, the resulting number system is the Robinson-style hyperreals, which (as mentioned above) behave somewhat differently from the number system Carnap envisioned. A similar account that also uses the hyperreals is given by Benci, Horsten, and Wenmackers (2013). But perhaps some other system satisfies those two axioms while rejecting some others and behaves more like the system Carnap hoped for.

A.2. *Lexicographic Probabilities*

Another approach to minuscule propositions is the technique of "lexicographic probabilities." Versions of this approach appeared as early as Kemeny 1955. Although Kemeny generally imposes the requirement of "strict coherence" (his term for **Regularity**), on pages 270–72, he considers what happens if this requirement is removed.

In his system, probabilities are assigned to sentences from some finite language. He shows that if strict coherence is required, then the probability functions are determined by assignments of nonzero numbers to the state descriptions (maximal consistent conjunctions of atomic sentences and their negations), summing to 1. The conditional probability $P(a|b)$ (Kemeny uses the notation "$P(a, b)$") is then given by the sum of the values on the state descriptions that make both a and b true, divided by the sum of the values on the state descriptions that make b true.

If strict coherence is not required, the situation is a bit more interesting. Instead of a single assignment of numbers to the state descriptions, we need a *sequence* of such assignments, such that the values in each individual assignment add up to 1, and such that every state description gets a nonzero value on exactly one of the assignments. In this case, the conditional probability $P(a|b)$ is defined as before, except that the values used in the calculation are the values given by the *first* assignment in the sequence where some state description making b true has a nonzero value.

This construction has since been generalized by others, including van Fraassen (1995) and Halpern (2010). In the modern version, we consider an arbitrary well-ordered sequence of probability functions, with the requirement that every nonempty proposition get a nonzero value in some function in the sequence. To update on a proposition, one first removes from the sequence all functions that give this proposition the value 0, and then applies standard conditionalization to all remaining functions. At any point in time, only the first function in the sequence represent's the agent's credences, with the others only serving to encode information about conditional credences and updates.

Van Fraassen and Halpern both consider the relation between these lexicographic probabilities and Popper's functions with primitive conditional probabilities. As suggested by Kemeny, Halpern shows that there is a strong equivalence between Popper functions and these lexicographic probabilities if we impose particular relations between the assignments. That is, the countably additive lexicographic probabilities

and the countably additive Popper functions can be paired up in such a way that corresponding lexicographic probabilities and Popper functions give exactly the same conditional probabilities to every pair of propositions. However, Halpern also shows (in examples 3.2 and 3.5) that if we impose a somewhat stronger relation between different assignments in the sequence, or give up on countable additivity, then there are Popper functions to which no lexicographic probability corresponds.

Both authors also consider the relation between lexicographic probabilities and probability functions that are allowed to take hyperreal values. Van Fraassen, in appendix A4, cites McGee's (1994) result showing that there is a correspondence between Popper functions and hyperreal probabilities and goes on to argue that the Popper functions (or associated lexicographic probabilities) are superior to the hyperreal-valued functions. In particular, he cites the fact that hyperreal values are highly nonunique and that the lexicographic probabilities are much easier to construct given a sequence of conditional probability values that one wants to match.

Halpern shows that the relations between these three approaches are somewhat more subtle. Although taking the "standard part" (the real number closest to a given hyperreal) of every conditional probability in a hyperreal-valued probability function gives a Popper function, and every Popper function arises in this way, Halpern shows that on a natural way of interpreting decision theory in these two frameworks, the corresponding functions give rise to different preferences among gambles. As Halpern shows in his example 5.3, an agent who has credence $1/2 + \varepsilon$ in p and $1/2 - \varepsilon$ in $\neg p$ will prefer a payoff conditional on p to the *same* payoff conditional on $\neg p$, but will *disprefer* it to any *larger* payoff conditional on $\neg p$. Since there is no Popper function with this behavior, the Popper function corresponding to this hyperreal-valued probability function fails to properly represent it.

As it turns out, the correspondence between lexicographic probabilities and hyperreal-valued probability functions doesn't have this problem—but as mentioned above, in infinite probability spaces where countable additivity isn't required, the correspondence is only one way. There are hyperreal-valued probability functions that are not represented by any lexicographic probability.

Thus, Popper functions, lexicographic probabilities, and hyperreal-valued probability functions are very similar in behavior (much more similar than Carnap's proposal is to any of them), but there are still important differences. In particular, hyperreal values allow far more

fine-grained distinctions than either of these other options. Additionally, since Popper functions and lexicographic probabilities are both defin-able in very constructive ways, the arguments I give in section 5 against hyperreals don't cause problems for Popper functions or lexicographic probabilities.

Whether these lexicographic probabilities really represent "infini-tesimal credences" or count as a way to satisfy **Regularity** seem to be primarily terminological questions. We can say that a proposition is minus-cule if it gets the value 0 from the first function in the sequence and say that its credence is "infinitesimal" if it gets a nonzero value from some later function in the sequence. But we might also just identify credences with the value assigned by the first function in the sequence, which would interpret these lexicographic probabilities as violating **Regularity**.

A.3. Further Mathematical Theories of Infinitesimals That Could Be Applied

There are also some other mathematical theories of infinitesimals that could be used in place of Robinson's hyperreals. For instance, one could use the theory of "surreal numbers" developed by John Conway, or the techniques of "smooth infinitesimal analysis" based on the ideas of Wil-liam Lawvere. (See Conway 1976 for the former and Bell 1998 for the latter.) Smooth infinitesimal analysis doesn't seem like an especially promising formalism since it treats infinitesimals as more like "infinitely small line segments" rather than as points on a number line, and it requires intuitionist logic instead of classical logic. The surreal numbers seem more promising as a device for future philosophers of probability to use. Their construction is a simultaneous generalization of Dedekind's construction of the real numbers and von Neumann's construction of the ordinals and can be carried out in a very weak set theory. As it turns out, we can name particular surreal infinitesimals, like $1/\omega$ and $2^{-\omega}$. However, the use of surreal numbers for probability values will have to be substantially different from the way Skyrms and Lewis recommend using hyperreals because the technique they take from Bernstein and Wattenberg (1969) leads directly to the construction of nonmeasurable sets, and thus goes beyond ZF in some substantial way. Additionally, the defender of surreal probabilities will need to address the worries raised by Williamson's argu-ment about linearly ordered comparative probabilities. It would be inter-esting to see whether the use of surreal numbers could get around these worries, but the eventual theory will have to look substantially different

from the one that Skyrms and Lewis proposed and other philosophers have uncritically adopted.

References

Alchourròn, C. E., P. Gärdenfors, and D. Makinson. 1985. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *Journal of Symbolic Logic* 50: 510–30.

Barrett, M. 2010. "The Possibility of Infinitesimal Chances." In *The Place of Probability in Science,* ed. E. Eells and J. H. Fetzer, Boston Studies in the Philosophy of Science, 65–79. Dordrecht: Springer.

Bartha, P., and C. Hitchcock. 1999. "The Shooting-Room Paradox and Conditionalizing on Measurably Challenged Sets." *Synthese* 118: 403–37.

Bell, J. L. 1998. *A Primer of Infinitesimal Analysis.* New York: Cambridge University Press.

Benci, V., L. Horsten, and S. Wenmackers. 2013. "Non-Archimedean Probability." *Milan Journal of Mathematics* 81, no. 1: 121–51.

Bernstein, A. R., and F. Wattenberg. 1969. "Non-standard Measure Theory." In *Applications of Model Theory of Algebra, Analysis, and Probability,* ed. W. A. J. Luxemburg, 171–86. New York: Holt, Rinehart and Winston.

Boolos, G. 1984. "To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables)." *Journal of Philosophy* 81, no. 8: 430–49.

Burgess, J. 2005. *Fixing Frege.* Princeton: Princeton University Press.

Carnap, R. 1980. "The Problem of a More General Concept of Regularity." In *Studies in Inductive Logic and Probability*, ed. R. Jeffrey, 2:145–55. Berkeley: University of California Press.

Chang, C. C., and H. J. Keisler. 1990. *Model Theory.* New York: Elsevier.

Christensen, D. 2001. "Preference-Based Arguments for Probabilism." *Philosophy of Science* 68, no. 3: 356–76.

Clark, A., and D. Chalmers. 1998. "The Extended Mind." *Analysis* 58: 10–23.

Colyvan, M. 2008. "Relative Expectation Theory." *Journal of Philosophy* 105, no. 1: 37–44.

Conway, J. 1976. *On Numbers and Games.* Natick, MA: A K Peters.

Copeland, B. J., and R. Sylvan. 1999. "Beyond the Universal Turing Machine." *Australasian Journal of Philosophy* 77, no. 1: 46–66.

Easwaran, K. 2008a. "The Foundations of Conditional Probability." PhD diss., University of California, Berkeley.

———. 2008b. "Strong and Weak Expectations." *Mind* 117, no. 467: 633–41.

Elga, A. 2010. "Subjective Probabilities Should Be Sharp." *Philosophers' Imprint* 10, no. 5: 1–11.

Hájek, A. 2003. "What Conditional Probability Could Not Be." *Synthese* 137: 273–323.

———. 2005. "Scotching Dutch Books?" *Philosophical Perspectives* 19: 139–51.

———. 2008. "Arguments for—or against—Probabilism." *British Journal for the Philosophy of Science* 59, no. 4: 793–819.

Halpern, J. 2010. "Lexicographic Probability, Conditional Probability, and Non-standard Probability." *Games and Economic Behavior* 68, no. 1: 155–79.

Hofweber, T. Forthcoming. "Infinitesimal Chances." *Philosophers Imprint.*

Holder, R. 2002. "Fine-Tuning, Multiple Universes and Theism." *Noûs* 36, no. 2: 295–312.

Hoover, D. 1980. "A Note on Regularity." In *Studies in Inductive Logic and Probability,* ed. R. Jeffrey, 2:295–297. Berkeley: University of California Press.

Jeffrey, R. 2004. *Subjective Probability: The Real Thing.* New York: Cambridge University Press.

Kanovei, V., and S. Shelah. 2004. "A Definable Nonstandard Model of the Reals." *Journal of Symbolic Logic* 69, no. 1: 159–64.

Kemeny, J. 1955. "Fair Bets and Inductive Probabilities." *Journal of Symbolic Logic* 20, no. 3: 263–73.

Kolmogorov, A. N. 1950. *Foundations of the Theory of Probability.* New York: Chelsea.

Lewis, D. 1980. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability,* ed. R. Jeffrey, 263–93. Berkeley: University of California Press.

———. 1996. "Desire as Belief 2." *Mind* 105, no. 418: 303–13.

Luxemburg, W. A. 1973. "What Is Nonstandard Analysis?" *American Mathematical Monthly* 80, no. 6: 38–67.

Maher, P. 1990. "Acceptance without Belief." In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,* 1:381–92. Chicago: University of Chicago Press.

McGee, V. 1994. "Learning the Impossible." In *Probability and Conditionals,* ed. E. Eells and B. Skyrms, 179–99. New York: Cambridge University Press.

Moore, G. 1982. *Zermelo's Axiom of Choice: Its Origins, Development, and Influence.* Mineola, NY: Springer-Verlag.

Neeman, I. 2010. "Determinacy in $L(\mathbb{R})$." In *Handbook of Set Theory,* ed. M. Foreman and A. Kanamori, 1887–1950. Dordrecht: Springer.

Norton, J. 2007. "Probability Disassembled." *British Journal for the Philosophy of Science* 58: 141–71.

Nover, H., and A. Hájek. 2004. "Vexing Expectations." *Mind* 113: 305–17.

Parikh, R., and M. Parnes. 1974. "Conditional Probabilities and Uniform Sets." In *Victoria Symposium on Nonstandard Analysis,* ed. A. Hurd and P. Loeb, 180–94. Berlin: Springer-Verlag.

Popper, K. 1955. "Two Autonomous Axiom Systems for the Calculus of Probabilities." *British Journal for the Philosophy of Science* 6, no. 1: 51–57.

Rényi, A. 1970. *Foundations of Probability.* San Francisco: Holden-Day.

Robinson, A. 1996. *Non-standard Analysis.* Princeton: Princeton University Press.

Seidenfeld, Teddy, Mark J. Schervish, and Joseph B. Kadane. 2013. "Two Theories of Conditional Probability and Non-Conglomerability." Eighth International Symposium on Imprecise Probability: Theory and Applications, July 2–5, Compiègne, France.

Shimony, A. 1955. "Coherence and the Axioms of Confirmation." *Journal of Symbolic Logic* 20, no. 1: 1–28.

Skyrms, B. 1980. *Causal Necessity.* New Haven: Yale University Press.

———. 1987. "Coherence." In *Scientific Inquiry in Philosophical Perspective,* 225–42. Pittsburgh, PA: University of Pittsburgh Press.

———. 1992. "Coherence, Probability and Induction." *Philosophical Issues* 2: 215–26.

Smith, C. 1961. "Consistency in Statistical Inference and Decision." *Journal of the Royal Statistical Society, Series B (Methodological)* 23, no. 1: 1–37.

Stalnaker, R. 1970. "Probability and Conditionals." *Philosophy of Science* 37, no. 1: 64–80.

Swinburne, R. 2001. *Epistemic Justification.* Oxford: Oxford University Press.

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

Weintraub, R. 2008. "How Probable Is an Infinite Sequence of Heads? A Reply to Williamson." *Analysis* 68, no. 3: 247–50.

Williamson, J. 1999. "Countable Additivity and Subjective Probability." *British Journal for the Philosophy of Science* 50: 401–16.

Williamson, T. 2002. *Knowledge and Its Limits.* Oxford: Oxford University Press.

———. 2007. "How Probable Is an Infinite Sequence of Heads?" *Analysis* 67, no. 3: 173–80.