Staying Regular?

Alan Hájek

> ALI G: So what is the chances that me will eventually die?
> C. EVERETT KOOP: That you will die? – 100%. I can guarantee that 100%: you will die.
> ALI G: You is being a bit of a pessimist…
> –Ali G, interviewing the
> Surgeon General, C. Everett Koop

> *To be uncertain is to be uncomfortable, but to be certain is to be ridiculous*
> —Chinese proverb

SINCE THIS IS RATHER LONG, I HAVE PUT IN GRAY FONT MATERIAL THAT COULD BE SKIPPED IF YOU ARE IN A HURRY.

1. Introduction

The scandal of orthodox Bayesianism is its permissiveness. Its constraints on rational credences may be fine as far as they go: your credences must be non-negative, normalized, and additive. But the constraints don't go nearly far enough. We may claim to embrace de Finetti's (1931) revolutionary spirit, to celebrate his freewheeling epistemology according to which rational belief is merely a matter of probabilistic 'coherence'. But in our hearts, we know that rationality is not so tolerant.

So some Bayesians are less permissive. They add further constraints on rational credences—the Principle of Indifference (Keynes 1921) or its modern incarnation in the Principle of Maximum Entropy (Jaynes 2003), the Principal Principle (Lewis 1980), the Reflection Principle (van Fraassen 1984), and so on. In this paper I canvass the fluctuating fortunes of a much-touted constraint, so-called *regularity*. It starts out as an intuitive and seemingly innocuous constraint that bridges modality and probability, although it quickly runs into difficulties in its exact formulation. I massage it, offering what I take to be its most compelling version: a constraint that bridges *doxastic* modality and *doxastic* (subjective) probability. So understood,

regularity promises to offer a welcome connection between traditional and Bayesian epistemology.

Yet I will argue that it is untenable. It will turn out that there are two different ways to rationally violate regularity: with *zero* probability assignments, and with *no* probability assignments at all (probability gaps). Both kinds of violations of regularity have serious consequences for Bayesian orthodoxy. I consider especially their ramifications for:

- the standard ratio formula for conditional probability

- conditionalization, characterized with that formula

- the standard multiplication formula for independence

- expected utility theory

2. Regularity: motivations and initial formulations

I submit that a guiding idea behind the Bayesian constraints that have been offered is the norm that *your credences should reflect your evidence*. Your attitude to a given proposition should be neither more committal, nor less committal, than your evidence bids it to be. A somewhat permissive Bayesian may insist that the only tenable *numerical* constraints on unconditional probabilities—unique *numbers* that are demanded of your credences in particular propositions—are those that follow from the probability axioms. That is, where C is a rational credence function,

If X is a tautology/necessary, then $C(X) = 1$.

If X is a contradiction/impossible, then $C(X) = 0$.

These constraints apparently accord with the norm. There is a good sense in which your evidence for any given tautology is as good as it could possibly be, and your

evidence against any given contradiction is as good as it could possibly be. The non-negativity axiom imposes only a weak inequality on rational credences:

For all X, $C(X) \geq 0$.

But even our somewhat permissive Bayesian may want to admit the smallest strengthening of this constraint if X is possible. The inequality becomes strict:

If X is possible, then $C(X) > 0$.

This is our first schematic formulation of *regularity*, but we will soon generalize it and refine it. The constraint entails the converse of the normalization axiom: *only necessary truths receive probability 1.* That is:

If $C(X) = 1$, then X is necessary.

These constraints advert to X being *possible* or *necessary*—but in what sense? This takes us from probability to modality—or better, *modalities*.

Modalities are puzzling. They involve not only, and some not even, what actually happens, but what *could* or *would* or *must* happen. Probabilities are puzzling twice over. They are not only modalities, but they are also ones that come in degrees. We start to gain a handle on both binary 'box'/'diamond' modalities and numerical probabilities when we formalize them, with various systems of modal logic for the former, and with Kolmogorov's axiomatization for the latter. We would understand both still better if we could provide bridge principles linking them.

Regularity conditions are such bridge principles—one for each pairing of a modality and a kind of probability. The terminology is remarkably unmnemonic. Schematically, the conditions have the form:

If X is possible, then the probability of X is positive.

Think of them as *open-mindedness* conditions on probability functions: anything that can happen (in the appropriate sense of 'can') is dignified with at least some

recognition from the corresponding probability function's assignment. They are precisifications of a commonsensical idea, the folk notion that "if it can happen, then it has some chance of happening".

There are many possible senses of 'possible' in the antecedent—notably logical, metaphysical, nomic, analytic, epistemic, doxastic, and so on. There are also many senses of 'probability' in the consequent—notably classical, logical, frequentist (actual and hypothetical), propensity (single case and long run), 'best systems' analyses, and subjective. Pair them up, and we get many, many regularity conditions, one for each pairing. Some are interesting, and some are not; some are plausible, and some are not.

I want to focus on pairings that are definitely interesting, and somewhat plausible, at least initially; we will get there in stages. In the consequent, let us restrict our attention to *subjective* probabilities. Versions of regularity as a rationality constraint on subjective probabilities have been proposed by Kemeny (1955), Shimony (1955, 1970), Jeffreys (1961), Edwards et al. (1963), Carnap (1963), Stalnaker (1970), Lewis (1980), Skyrms (1980), Appiah (1985), Jackson (1987), Jeffrey (1992), and others, although some of them are not entirely clear on which sense of 'possible' is at issue.

This still leaves open various regularity conditions, corresponding to the various binary modalities in the antecedent, and some are less plausible than others. Here is one that I find less plausible, although it has been suggested by Shimony (1955) and Skyrms (1995):

Logical regularity: If X is logically possible, then C(X) > 0.

Contra logical regularity, any given agent's credence function may, and perhaps even should, give 0 to some things that are logically possible and 1 to some things that are not logically necessary. There are all sorts of propositions that are knowable a priori,

but that are not logically necessary: mathematical truths, metalogical truths, mereological truths, analytic truths, and so on. It is surely no requirement of rationality to be probabilistically uncertain of such truths. Thus, the regularity condition that pairs up logical possibility with subjective probability is unpromising—at least if "logic" is interpreted in a narrow sense.

However, we might recognize logic in broader senses—the latter encompassing mathematical truths, and so on—thus reducing the space of what counts as "logically possible". So understood, logical regularity is a more plausible constraint—but not plausible enough. 'Socrates is a 3-place relation', and 'Julius Caesar is the number 17' may still be logically possible in these broader senses, but arguably they may be entirely ruled out a priori, and as such may rationally be assigned credence 0.[1]

Weaker, and more plausible still, is:

Metaphysical regularity: If X is metaphysically possible, then $C(X) > 0$.

It follows that if $C(X) = 0$, then X is metaphysically impossible. This brings us to Lewis's (1986) characterization of "regularity": "$C(X)$ is zero … only if X is the empty proposition, true at no worlds" (88, changing his "B" to "X"). (According to Lewis, X is metaphysically possible iff it is true at some world.) Lewis regards regularity in this sense as a constraint on "initial" (prior) credence functions, those that could codify the credences of agents as they begin their Bayesian odysseys—Bayesian Superbabies.[2]

---

[1] Thanks to Daniel Nolan for the examples.

[2] The term 'Superbaby' was coined by David Lewis; I heard it first in an undergraduate lecture of his on Time Travel at Princeton.

[3] Easwaran (MS) offers a similar formulation of regularity in terms of doxastic possibility, but it replaces the "is" of my formulation with "should be": his notion of 'regularity' is normative. I will find it convenient instead to regard 'regularity' as simply a property that certain probability functions—or better still, certain probability spaces, as we will see—have

[2] The term 'Superbaby' was coined by David Lewis; I heard it first in an undergraduate lecture of his on Time Travel at Princeton.

More plausible still, but still not plausible. It is metaphysically possible for no thinking thing to exist. In order to be regular in Lewis's sense, then, one must assign positive probability to no thinking thing existing. But far from being rationally required, this appears to be *irrational* (even for a Superbaby). Among other things, it renders an agent susceptible to a novel and curious kind of Dutch Book: she will be open to buying for a positive amount a bet that pays $1 if no thinking thing exists. Pays *whom*? Not her! It is a bet that she must lose. But this is overkill; it suffices to scotch Lewis's regularity constraint for it to be *rationally permissible* to assign credence 0 to no thinking thing existing (and surely it is even for a Superbaby). And an infallible, omniscient God who knows which world is actual, and so concentrates all credence on that world, is metaphysically irregular, but should hardly be convicted thereby of being irrational!

Then there are agents who fall short of Godliness, but who are nevertheless infallible over certain circumscribed domains. Perhaps, for example, they are infallible about their own mental states. Metaphysical regularity prohibits *luminosity* of one's credences:

$$\text{If } C(X) = x, \text{ then, } C[\,C(X) = x\,] = 1.$$

After all, it is metaphysically contingent that one has the particular credences that one does. But surely it is not automatically irrational to be certain of what they are. Descartes may have been wrong to think that we are perfect introspectors of our mental states, but surely his view did not entail that we are ipso facto irrational! And if there are such perfect introspectors out there, more power to them; they need not be falsely modest about their gifts. Still less should one be faulted for correct *singular instances* of luminosity. For example, suppose you assign credence 1 to the proposition that something is self-identical, and *that* fact is luminous to you:

$C[\ C(\text{something is self-identical}) = 1\ ] = 1.$

You are still metaphysically irregular, for again that fact is metaphysically contingent, but you should not automatically be charged with irrationality! And arguably it is at least rationally permissible for an agent to assign credence 1 to some propositions concerning the phenomenal experiences she is having; but again these propositions are metaphysically contingent, so again she violates metaphysical regularity, as Tang (MS) observes. Now, perhaps these introspective cases are not problems for Lewis, since arguably an agent's credence is no longer an *initial* credence by the time she first introspects (on her own credences, or on her phenomenal experiences)—she has gained some empirical information, albeit about herself. Still, these striking consequences of metaphysical regularity are worth noting.

Lewis (1986) insists that many *X*'s are metaphysically possible, but ineligible to be the contents of thought, namely, highly gerrymandered or unnatural propositions. Such *X*'s are not fit to receive credences at all, still less positive credences; and still less is it rationally required that they receive positive credences. So by Lewis' lights, metaphysical regularity cannot be a rationality constraint. Interestingly, his formulation of regularity is weaker in just such a way as to evade *this* problem, at least. Suppose that you assign no credence at all to such a gerrymandered proposition; then you violate metaphysical regularity (in failing to assign it *positive* credence), but not regularity in Lewis' sense. This alerts us to the distinction, which I will make much of, between *zero* probability assignments, and *no* probability assignments (probability gaps). In any case, Lewis's sense still falls prey to the earlier counterexamples.

So we have yet to see a viable regularity constraint. I suggest that we look instead to traditional epistemology—in particular, to its central notion of *belief*—for a more promising formulation of regularity, and related formulations.

3. Doxastic regularity and epistemic regularity

Belief is the ungraded counterpart of Bayesianism's graded notion of credence. Belief, in turn, induces a corresponding modality: a *doxastic possibility* for an agent is a proposition that is logically compatible with what that agent *believes*. We should add that a doxastic possibility is also *eligible* to be believed, to avoid the Lewisian problem that I just raised. Doxastic possibility is an especially natural candidate for pairing with subjective probability, and it provides a promising version of regularity:

Doxastic regularity: If *X* is doxastically possible for a given agent, then the agent's subjective probability of *X* is positive.

Better yet, it provides *many* promising versions of regularity: one for each agent, and there are many agents.[3] And we multiply still more versions if we distinguish between *narrow* and *broad* logical possibility (and thus narrow and broad logical compatibility), and between *rational* belief and belief simpliciter (and thus *rational* doxastic possibility and doxastic possibility simpliciter). As we will see, I think it will not matter how we finesse doxastic regularity, since it will be untenable on any reasonable understanding of it. But first, let's give it its due.

---

[3] Easwaran (MS) offers a similar formulation of regularity in terms of doxastic possibility, but it replaces the "is" of my formulation with "should be": his notion of 'regularity' is normative. I will find it convenient instead to regard 'regularity' as simply a property that certain probability functions—or better still, certain probability spaces, as we will see—have and others have not. The putative norm can be imposed later: agents *should* have such functions, or spaces.

It is prima facie plausible that an agent's beliefs and degrees of belief are, or at least should be, connected closely enough to uphold doxastic regularity. Indeed, a version of the so-called 'Lockean thesis' claims that rational belief just *is* rational degree of belief above a certain threshold. On this view, if a rational agent assigns probability 0 to X, then she assigns probability 1 to $\neg X$, and thus she believes $\neg X$ (whatever the threshold). Given her rationality, she then disbelieves X, so X is not doxastically possible for her. That is, if her probability of X is 0, X is not doxastically possible for her—just as doxastic regularity requires.

But even eschewing the Lockean thesis, doxastic regularity seems compelling. If it is violated, then offhand two different attitudes are conflated: one's attitude to something one outright disbelieves, and one's less committal attitude to something that is still a live possibility given what one believes. Indeed, I think the folk are tacitly committed to doxastic regularity: if something is a live possibility given what one believes, then one gives it at least *some* credence. (Not that the folk would say it in quite those words!)

Anything that is known, or even just believed, is doxastically necessary, so assignments of 1 to mathematical truths, the existence of thinking things, and so on no longer provide counterexamples. A God who knows which world is actual faces no censure from doxastic regularity, since any other world is doxastically impossible for her. And perfect introspection of one's credences, or of one's phenomenal experiences, is not threatened by doxastic regularity as long as one *believes* oneself to have the credences, or phenomenal experiences, that one has.

Above all, doxastic regularity seems to do a fine job of honouring the norm that your credences should reflect your evidence. Plausibly, if *X* is doxastically possible, then your evidence does not rule it out.[4] Nor should your credence function.

And if this version of regularity fails, then various other interesting versions will fail too. For example, consider *epistemic regularity*, which replaces "doxastically" with "epistemically" in doxastic regularity; this is Stalnaker's (1970) formulation. Now, *epistemic* possibility is weaker than doxastic possibility: since knowledge entails belief, anything that is compatible with what one believes is compatible with what one knows, but not vice versa. So epistemic regularity is stronger than doxastic regularity; if the latter fails as a rationality constraint, so does the former. But the former might fail for other reasons. For example, perhaps rationality permits one to give credence 0 to one's being a brain in a vat, even though one only believes, but does not *know* that one is not: one's being a brain in a vat is epistemically but not doxastically possible in this case.

To be sure, doxastic regularity seems at first blush remarkably weak. If *X* is doxastically possible, then doxastic regularity together with the probability axioms places almost no constraint on the credence you assign to *X:* as long as it is positive, you have committed no sin against rationality as far as doxastic regularity is concerned. Nonetheless, doxastic regularity might be regarded as a good Bayesian constraint on rational credences to add to the probability axioms, among others.

And yet on second and further blushes, doxastic regularity can be undermined. In the next section, I will characterize regularity more generally as a certain internal harmony of a probability space $\langle \Omega, F, P \rangle$. Doxastic regularity will fall out as a

---

[4] This is surely true on the Williamsonian (2003) account of evidence as what one knows. If X is compatible with what one believes, then it is compatible with what one knows. It is also plausible if it is *rational* doxastic possibility that we have in mind.

special case, as will epistemic regularity and other related regularities. Then we will be in a good position to undermine them.

4. A more general characterization of regularity

It will be useful to give a general characterization of regularity at a certain level of abstraction. We will then recognize doxastic regularity, epistemic regularity, and other regularities as salient instances.

Philosophers tend to think that all of the action in probability theory concerns probability *functions*. But for mathematicians, the fundamental object of probability theory is a *probability space*, a *triple* of mathematical entities. First we need a set of mutually exclusive and jointly exhaustive possibilities, which we will designate 'Ω'. (I think this started out in the literature as another completely unmnemonic choice, but by sheer luck it turned out evocatively: the Greek letter 'Ω' corresponds to 'W' on the standard QWERTY keyboard, which will remind us of a set of Worlds!) We then need a field (algebra) F of subsets of Ω. Finally, we define a probability function P on F that obeys Kolmogorov's axioms. When speaking of things probabilistic, we normally go straight to the relevant probability function. But I want to put Ω and F in the spotlight also.

Regularity can now be characterized in terms of a certain kind of harmony between Ω, F, and P. Firstly, between Ω and F: F is the power set of Ω, so no subset of Ω goes missing in F; secondly, between F and P: every set in F except the empty set receives positive probability from P. This means that P recognizes (bestows a positive assignment to) every proposition that F recognizes (has as a member) except the empty set, and F recognizes every proposition that Ω recognizes (has as a subset). In this sense, P mirrors Ω: P's non-zero probability assignments track the non-empty

subsets of Ω. It is a simple mathematical fact that some probability spaces are regular, and some are not—some display this kind of harmony, and some do not (with the labour of making distinctions among possibilities more divided between Ω and P).

I have spoken of regularity in terms of bridges between modality and probability. Ω represents the binary modal aspect of the probability space: its set of possibilities. P represents the probabilistic part of the probability space (of course). A bridge in one direction is secured by the probability axioms alone: Ω receives probability 1 from P (equivalently, the empty set receives probability 0 from P). Regularity provides the bridge in the other direction: *only* Ω receives probability 1 from P (equivalently, *only* the empty set receives probability 0 from P). Thanks to regularity, there is two-way traffic between binary modality and probability.

The three entities Ω, F, and P of a probability space give rise to what I will call *three grades of probabilistic* involvement with respect to a given proposition. A set of possibilities may be recognized by the space by:

(First grade) being a non-empty subset of Ω;

(Second grade) being a non-empty element of F; and by

(Third grade) receiving positive probability from P.

These are non-decreasingly committal ways in which the space may countenance a proposition. *The space is regular if these three grades collapse into one*. That is, every non-empty subset of Ω receives positive probability from P.

So far, this is all formalism; it cries out for an interpretation. Again, philosophers have tended to focus on the interpretation of P—the main candidates being the classical, frequentist, logical, propensity and (the one relevant here) subjective interpretations. But Ω and F deserve their day in the sun, too. Ω is a set—of *what*? I

will think of the elements of Ω as worlds, the singletons of Ω as an agent's maximally specific doxastic possibilities. *F* is a set of subsets of Ω—*which ones*? They will be the privileged sets of such doxastic possibilities that are contents of the agent's credence assignments. ("$\mathscr{C}$" would be more evocative than "*F*", but I'll stick to the standard notation.)

Or start with the doxastic state of a rational agent; it cries out for a formalism so that we can model it. Philosophers have tended to focus on her probability assignments—real values in [0, 1] that obey the usual axioms. But surprisingly little is said about the *contents* of these assignments, and they deserve their day in the sun too. They are represented by a set *F* of privileged subsets of a set Ω. Thus, three tunnels with opposite starting points and heading in opposite directions meet happily in the middle:

- The graded notion of a rational agent's credence obeys the probability calculus, and that formalism finds an interpretation in such an agent's credence.

- The contents of the agent's credences form a field of privileged sets of doxastic possibilities, and the *F* of the formalism finds an interpretation in such contents.

- The agent's maximally specific doxastic possibilities form a set, and the Ω of the formalism finds an interpretation in the set of those possibilities.

We have a formalism looking for a philosophical interpretation and a philosophical interpretation looking for a formalism happily finding each other. Moreover, where traditional and Bayesian epistemology have hitherto largely proceeded on separate tracks, this way they are agreeably linked. Whereas Bayesian epistemology often shuns or ignores traditional epistemology's staple concept of belief, here that concept

earns its keep in the characterization of Ω and F (since doxastic possibility is defined in terms of belief).

So we recognize doxastic regularity as a special case of regularity, one in which the three grades of probabilistic involvement collapse into one for a probability space interpreted as I have. There are other interesting special cases, too. For example, we might interpret Ω instead as the agent's set of maximally specific epistemic possibilities, and F as the privileged sets of these possibilities to which she bestows credences. We could then define *epistemic regularity* in terms of the three grades of probabilistic involvement collapsing into one for *this* space. Or we might interpret Ω as the set of *deep epistemic* possibilities in the sense of Chalmers (forthcoming): those possibilities that cannot be ruled out a priori, and F as the set of subsets of such possibilities that receive credences from some agent. We could then define *deep epistemic* regularity in terms of the three grades of probabilistic involvement collapsing for *this* space. Then we might impose different regularity constraints on different agents. For example, perhaps Superbabies should have deep-epistemically regular probability spaces, and then doxastically regular probability spaces once they start learning, and thereafter. Or those who are suspicious of these modalities are free to propose their own interpretation of Ω and F—perhaps because they are eliminativists about the folk-psychological notion of belief, à la Churchland and Stich; perhaps because they think that the notion of knowledge belongs to a superseded epistemology, à la Jeffrey; perhaps because they are dubious about the a priori, à la Quine.

Or we might have in mind a different kind of probability function—for instance, the *chance* function (at a time). We could then ask what its Ω and F are—perhaps the set of metaphysically possible worlds, and its power set, respectively—and whether

the resulting probability space is regular. (According to Lewis it is not, since according to him all propositions about the past have chance 0 or 1.) One of the virtues of my characterization of regularity is its flexibility to capture many different versions of it. But my interest is primarily in the version that couples doxastic possibilities for an agent with her subjective probabilities, doxastic regularity.

Doxastic regularity is a property that the spaces of certain agents have and others do not. It is controversial whether agents *should* have such spaces—whether they are governed by the norm '*Thou shalt stay regular!'*

There are two ways in which an agent's space could fail to be regular, however we interpret Ω and F (in terms of doxastic, epistemic, deep possibilities, or what have you):

1) The way that gets all of the attention in the literature is for her probability function to assign zero to some member of F (other than the empty set). Then her second and third grades of probabilistic involvement come apart for this proposition. On my preferred interpretation: some doxastic possibility does not receive positive probability.

2) The far less familiar way is for her probability function to fail to assign *anything* to some subset of Ω, because the subset is not an element of F. Then her first and second grades of probabilistic involvement come apart for this proposition. On my preferred interpretation: some doxastic possibility does not receive probability at all. This neglected route to irregularity becomes salient when we look to traditional epistemology to guide our construction of a probability space.

Those who regard regularity as a norm of rationality, for some suitable understanding of the possibilities at issue, must insist that all instances of 1) *and* all instances of 2) are violations of rationality. I will argue that there are rational

instances of both 1) and 2) for doxastic possibilities constituting Ω. These arguments will carry over to various other interesting interpretations of Ω.

5. The three grades of probabilistic involvement: an example

Anything that is not a subset of Ω is not countenanced by $<\Omega, F, P>$ at all; it does not even make the first grade of probabilistic involvement for the agent with that space. Ω is probabilistically certain—again, one of Kolmogorov's axioms requires this. C. I. Lewis famously said that "if anything is to be probable, then something must be certain", and this is enshrined in the very foundations of probability theory. It is ironic, then, that Jeffrey (1992) *objected* to Lewis's dictum, famously insisting that there could be "probabilities all the way down to the roots". This is his "radical probabilism". But this is an incoherent position. *There could not be probabilities without some underlying certainty.* Far from embodying some philosophical error, Lewis's dictum is a trivial truth! Lewis may have been wrong to regard sense data as providing the probabilistic bedrock, and Jeffrey may have been right to regard them as potentially uncertain. But that is another matter.

Here is an example that displays the three grades of probabilistic involvement. Philosophers of probability are fond of throwing darts at the [0, 1] interval of the real line, and I am no exception. Here comes my throw. As usual, my dart has an infinitely thin tip, its landing point is a real number in the interval, and my throw will be fair, not privileging any part of the interval over any other: the distribution over the possible landing points is uniform. We now construct a probability space to model the attitudes of an ideal agent to this experiment (and for now let us suppose that these are the only attitudes of interest). Let $\Omega = [0, 1]$. Any landing point outside this interval is simply not countenanced—the corresponding proposition does not make even the first

grade for our space. Nor do various other possibilities—the dart turning into a shetland sheepdog, the dart singing *Waltzing Matilda*, and many others.

F could be very rich or very sparse. For example, it could consist of just Ω and the empty set. But to model our agent adequately it may include all the singleton sets, corresponding to all the exact landing points. By the closure properties of *F*, that will yield all sorts of complicated unions—e.g. the dart's landing on a rational-valued point—and intersections of sets. We may want it even to be the sigma field generated by the sub-intervals of [0, 1]. But if we want *P* to respect the symmetries imposed by the fairness of the throw—translation invariance—then *F* cannot be the power set of Ω. Famously, certain subsets of Ω—so-called *non-measurable* sets—go missing. These get no probability assignments whatsoever; their probabilities simply don't exist. We know this by working backwards from the constraints that we have put on *P*, and assuming its countable additivity, and the axiom of choice (as is standard). *P* is the Lebesgue measure on the interval. Vitali proved that *P* cannot be defined on the power set of Ω. There must be probability *gaps* as far as *P* is concerned. They achieve the first, but not the second grade of probabilistic involvement in this space.

But even among those subsets of Ω that do appear in *F*, and thus achieve the second grade, not all of them achieve the third grade. For various such subsets get assigned probability 0. The empty set, of course, is one of them. But so are all the singletons; all of the finite subsets; all of the countable sets, such as the set of rational numbers in [0, 1]; and even various uncountable sets, such as Cantor's 'ternary set'. In this sense, *P* is rather demanding: to register recognition from it, it doesn't suffice for a set to be large.

Examples like these pose a threat to regularity as a norm of rationality. In particular, they pose a threat to doxastic regularity. Any landing point in [0, 1] is

doxastically possible for our ideal agent. If we further imagine that the dart experiment is the *only* matter on which she is not opinionated, Ω can be interpreted as her set of doxastic possibilities, and *F* the privileged sets of such possibilities to which she assigns credences. We thus get two routes to irregularity as before, now interpreted doxastically.

I have characterized regularity—in particular, doxastic regularity—and thus what it means for it to be a rationality requirement. But should we believe that it is?

6. Against regularity: first pass

> "I believe in an open mind, but not so open
> that your brains fall out"
> –Arthur Hays Sulzberger,
> former publisher of the *New York Times*

Much of what follows can stay neutral regarding the particular version of regularity at issue—doxastic, epistemic, deep-epistemic, or what have you—but it will be fine if you keep doxastic regularity in mind throughout.

In order for there to be the kind of harmony within <Ω, *F*, *P*> that is captured by regularity, there has to be a certain harmony between the *cardinalities* of *P*'s domain—namely *F*—and *P*'s range. If *F* is too large relative to *P*'s range, then a failure of regularity is guaranteed, and this is so without any further constraints on *P*. For example, Kolmogorov's axiomatization requires *P* to be *real*-valued. This means that any uncountable probability space is automatically irregular. Indeed, any probability function defined on an uncountable algebra assigns probability 0 to *uncountably* many propositions, and so in that sense it is *very* irregular. (See Hájek 2003 for proof.) The dart throw at the [0, 1] interval furnishes one example. Tossing a coin infinitely many times provides another one.

Pruss (MS) generalizes this observation. Assuming the axiom of choice, he shows that if the cardinality of Ω is greater than that of the range of *P*, and this range is totally ordered, then regularity fails: either some subset of Ω gets probability 0, or some subset gets no probability whatsoever. The upshot is that non-totally ordered probabilities are required to save regularity—a departure from orthodoxy so radical that I wonder whether they deserve to be called "probabilities" at all.

It is noteworthy, and perhaps even curious, that Kolmogorov's axiomatization is restrictive on the *range* of all probability functions: the real numbers in [0, 1], and not a richer set; yet it is almost completely permissive about their *domains:* Ω can be any set you like, however large, and *F* can be any sigma algebra on Ω, however large. So, short of Ω forming a proper class, it can be as large as you want, and likewise *F*. In this sense we should see irregularity coming at us like a freight train, even if we restrict our attention to the doxastic possibilities and credences of rational agents. *On the one hand, we can apparently make the set of <u>contents</u> of an agent's thoughts as big as we like; on the other hand, we restrict the <u>attitudes</u> that she can bear to those contents*—they can only achieve a certain fineness of grain. Put a rich set of contents together with a relatively impoverished probability scale, and *voilà*, you have irregularity. We will see authors such as Lewis seeking to enrich the probability scale. But two can play this game, and we can super-enrich the domain in order to thwart regularity again, as we will also see.

7. Infinitesimals to the rescue?

The friend of regularity replies: if you are going to have a rich *domain* of the probability function, you had better have a rich *range*. Recall that Lewis is a friend of regularity, for Superbabies. He writes:

You may protest that there are too many alternative possible worlds to permit regularity. But that is so only if we suppose, as I do not, that the values of the function *C* are restricted to the standard reals. Many propositions must have infinitesimal *C*-values, and *C(A|B)* often will be defined as a quotient of infinitesimals, each infinitely close but not equal to zero. (See Bernstein and Wattenberg (1969).) (88)

I *have* seen Bernstein and Wattenberg (1969). But this article does not substantiate Lewis' strong claim that there are too many worlds to permit regularity *only if* the credence function is real-valued; indeed, that claim is surely false. Bernstein and Wattenberg show that using the hyperreal numbers—in particular, infinitesimals—one can give a regular probability assignment to the landing points of a fair dart throw, modelled by a random selection from the [0, 1] interval of the reals. But this is a specific sample space, with a specific cardinality (namely, that of the reals). In fact, *by Lewis's own lights* (1973) the cardinality of the set of alternative possible worlds is greater than that—at least beth 2, the cardinality of the power set of the reals. So Lewis's citation of this article does not allay the "protest" that he imagines.

Williamson (2007) argues that regularity cannot be saved, even by invoking infinitesimals, when considering all the possible outcomes of tossing a coin infinitely many times. I will offer a rather different argument.

Contra Lewis, we can thwart regularity even for a hyperreal-valued probability function by correspondingly enriching Ω. Pruss's result guarantees that, but we don't even need to appeal to it. Simply imagine a spinner whose possible landing points are randomly selected from the [0, 1) interval of the *hyperreals*. Now each point will surely get probability 0 from a hyperreal-valued probability function that respects the symmetry of the situation. Each point x is strictly contained within nested intervals of the form $[x - \varepsilon/2, x + \varepsilon/2]$ of infinitesimal width ε, for all possible ε, whose probabilities are their lengths, ε again. So the point's probability is bounded above by

all these ε, and thus it must be smaller than all of them. (If its probability were $\varepsilon_0 > 0$, then we could consider an interval of length $\varepsilon_0/2$ that covers the point, and whose probability is $\varepsilon_0/2$; but this is impossible, since the interval's probability must be at least as great as that of a point contained inside it.) Each such landing point corresponds to a possible world, in which the spinner lands on *that* point. So there are too many alternative possible worlds to permit regularity even if we suppose that the values of the credence function are *not* restricted to the standard reals, but rather to the hyperreals.

Now, perhaps we can no longer assume that probabilities of intervals can be identified with their lengths when the lengths are infinitesimal.[5] I reply that if they cannot be so identified, then so much the worse for infinitesimal probabilities. After all, the identification of probability with length seems to be a perfectly reasonable—and perhaps even a non-negotiable—constraint on one's credences about the dart experiment. It falls out of the symmetry constraint, after all. In any case, I don't need to assume this identification in order to run my argument. It suffices for there to be a sequence of nested intervals, shrinking towards the chosen point, whose probabilities are not bounded away from 0 (and in particular, there is no infinitesimal ε_1 such that all their probabilities are greater than ε_1). The identification of probabilities with lengths guarantees this, but it is not required.

I envisage a kind of arms race: we scotched regularity for real-valued probability functions by canvassing sufficiently large domains: making them uncountable. The friends of regularity fought back, enriching their ranges: making them hyperreal-valued. I counter with a still larger domain: making its values hyperreal-valued. Perhaps regularity can be preserved over *that* domain by enriching the range again, as

---

[5] Thanks here to Thomas Hofweber.

it might be, making it hyper-hyperreal-valued. I counter again with a yet larger domain: making *its* values hyper-hyperreal-valued. And so it goes. Some latter-day Bernstein and Wattenberg would need to keep coming up with constructions that would uphold regularity, however big Ω gets—presumably with ever-richer fields of numbers to provide the values of the probability functions.

Could a *single* construction handle all Ω's at once? I doubt that, at least for anything recognizably like Kolmogorov's axiomatization. For recall the somewhat peculiar feature of it that a decision about the *range* of all probability functions is made once and for all, while allowing complete freedom about how large the domains can be. However rich the range of the probability functions get, I could presumably run the trick above again: fashion a spinner whose landing points come from a *domain* that is so rich, threatening to thwart regularity once again. Again, Pruss's result clinches the case. Regularity cannot be sustained with anything resembling Kolmogorov's axiomatization, however much the cardinality of Ω escalates.

Could we tailor the range of the probability function to the domain, for each particular application? Much as the general of a defense force may wait and see how big an invading army is before deciding how many troops of his own to commit, perhaps we could wait and see how big Ω and F are before deciding to how rich a system of numbers—reals, hyperreals, hyper-hyperreals (if such there be), or some quite different system …—we should commit? The trouble, as I have already foreshadowed, is that in a Kolmogorov-style axiomatization the commitment to the range of *P* comes *first*: forever more, probability functions will be mappings from sigma algebras to the reals. Or to the hyperreals. Or to the hyper-hyperreals. Or to some quite different system… Just try providing an axiomatization along the lines of Kolmogorov's that has flexibility in the range built into it. "A probability function is a

mapping from *F* to …"—well, to *what*? It is not enough to say something unspecific, like "some non-Archimedean closed ordered field …" Among other things, we need to know what the *additivity* axiom is supposed to be. Kolmogorov gave us countable additivity, but that does not even make sense in a hyperreal framework[6]; we need to be told explicitly whatever replaces it (e.g. **finite-additivity*). If we don't know exactly what the range is, we don't know what its notion of additivity will look like. And yet there is total freedom in the choice of Ω and *F*. If there is to be that sort of freedom in the range of probability functions, the theory will have to look very different from Kolmogorov's.

8. Doxastically possible credence gaps

It is time for us to enter the shadowy realm of propositions that make it to the first but not the second grade of probabilistic involvement: propositions that are subsets of Ω, but that are not elements of *F*. They are composed entirely of the agent's own maximally specific doxastic possibilities, each of which is fit to be the contents of her thought. Offhand, then, the propositions would seem to be fit to be the contents of her thought also. In some cases they will not obviously be so—for example, if they are highly gerrymandered disjunctions of such possibilities. In other cases, they obviously will—they may be the objects of her desires, for example. It is especially odd, then, that they will seem not to be fit to be the contents of her *credences*. If this is right, then there is a surprising disconnect between beliefs and credences: something that might be true for all that one believes is nevertheless not something to which one can attach a *degree* of belief.

---

[6] A countable sum is a limit of partial sums. But the usual definition of the limit, 'for all $\varepsilon > 0$, there is an $n \in \mathbb{N}$ such that …' refers to $\mathbb{N}$, the set of natural numbers. But this is an *external set* from the point of view of a non-standard model, and as such is not recognized by the model.

Decision theory recognizes the possibility of probability gaps in its distinction between *decisions under risk*, and *decisions under uncertainty*: in the latter case, probabilities are simply not assigned to the relevant states of the world. But we can make the case for credence gaps more directly. Suppose that you are currently not prepared to enter into *any* bet regarding X, whatever the odds, whatever the stakes.

Now, it may appear that rather than treating X as a credence gap, you are maximally imprecise about it: you are not prepared to buy a dollar bet on X for more than 0 cents, and you are not prepared to sell such a bet for less than a dollar, so you behave as if you were imprecise over the entire [0, 1] interval. (Not that such imprecision would be good news for regularity either: for it would then be *indeterminate* whether you assigned positive probability to some doxastic possibility.) But there is arguably a difference between gappiness and maximal imprecision. If you were maximally imprecise regarding X, we could arguably precisify your credence for X with any value in the [0, 1] interval, claiming that this was consistent with your opinion, as far as it goes, but simply going further. But we may suppose that your opinion is determinately *not* such value: that you resolutely *refuse* to assign X a probability, or for some reason *cannot* assign X a probability. Then any precisification of the probability of X distorts your opinion, giving it a value where you choose to remain silent. In that case, your opinion would be better represented as having a gap for X. Some examples should make this idea clearer.

i. Non-measurable sets

The most difficult, but also the most technically rigorous example, is one in which X is a *non-measurable* set – a set that simply cannot be assigned a probability, consistent with certain symmetry constraints that are forced upon the agent. We have

already seen such sets arising out of the dart throw at [0, 1].[7] Accordingly, any precisification of the probability of *X* would *misrepresent* the agent's credences, which are coherent.[8] The Banach-Tarski paradox furnishes further examples of such sets.

ii. Chance gaps

I have argued elsewhere (e.g. 2003) that there may be *chance gaps*. (A quick argument that we have already loaded up: suppose that the dart throw is governed by a uniform *chance* distribution; then the non-measurable sets will be chance gaps.) Suppose that *X* is such a gap. What credence should you, an ideally rational agent, assign to *X*? When you know that a proposition's chance has a particular value, by the Principal Principle you should make that value your own credence. But this is not your situation; rather, you know that there is *no* value that is *X*'s chance—not even an interval. Still, consider an extension of sorts to the Principal Principle. It says that your credence in *X*, *conditional* on *X*'s chance-gappiness, should be gappy:

$P(X \mid \text{chance}(X)$ is undefined) is undefined.

If you regard the chance function as unable to assign anything to *X*, it may be odd, and arguably irrational, for your credence to do so—at least when you have no other source of information about *X*.

Be that as it may, this extension to the Principal Principle is overkill. To refute doxastic regularity, it suffices for it to be *permissible* to reflect a hypothesized credence gap in your own credence—sometimes pleading the Fifth when the chance

---

[7] We prove that they cannot be assigned any value by reductio. We make countably many copies of *X*, which collectively partition the interval. Any putative precisification of the probability of X will lead to contradiction—assigning probability either 0 or ∞ to the entire interval.

[8] Bernstein and Wattenberg's construction also renders every subset of the interval measurable.

function does, by your lights. I find it hard to see how your rationality must thereby be discredited. For example, you certainly would be in no danger of being Dutch Booked. When you have no credence for *X* at all, you have no corresponding dispositions to enter into bets concerning *X*, and *a fortiori* no corresponding dispositions to enter into bets concerning *X* that are guaranteed to lose.

There are arguably various cases of *indeterminism without chanciness*. (Here I am indebted to Eagle 2011.) Earman (1986) argues that despite appearances, classical mechanics is indeterministic. Think of two particles *a* and *b* at rest beside each other. A violent force is applied to *a* so that its velocity increases without bound, and in finite time it is infinitely far away. Now time-reverse its motion, consistent with Newton's laws. We see a 'space invader' that starts infinitely far away, that approaches *b* at great speed, slows down, and comes to rest beside it. This is a case of indeterminism: for it is compatible with the state consisting solely of *b* and Newton's laws both that nothing interesting will happen, and that *a* will suddenly fly in from infinity. But there are no chances in this picture.[9]

It seems, then, that classical mechanics is indeterministic but not chancy: starting from a fixed set of initial conditions, it allows multiple possible futures, but they have no associated chances. Now perhaps a rational agent may correspondingly not assign them any credences, even though they may be doxastic possibilities for her. To be sure, she *may* also assign them credences; I am just suggesting that she is not rationally *compelled* to do so. And if she is not, then she provides another counterexample to regularity.

---

[9] 'Norton's dome' (Norton 2008) is another well-known example of apparent indeterminism in classical mechanics.

Now, it might be argued that agents are unlike chances—for example, agents sometimes have to bet, while chances never do![10] Well, I suppose that if someone coerced you to bet on some chance gap—perhaps by making you an offer that you couldn't refuse, à la Godfather—then sure enough we would witness some betting behaviour. But it is doubtful that it would reveal anything about your state of mind prior to the coercion. So at best this shows that it would be rational for you to *fill* a gap, when coerced. It does not impugn your having the gap in the first place—and that's all we need for a counterexample to regularity.

iii. One's own free choices

Eagle offers another potential example of indeterminism without chances: "Libertarians believe that our free will is not determined by the past states of the universe, but also that our exercises of will are not purely by chance—rather, we make them happen (through the determinations of the will, not the determination of past history)." It is not clear that one must assign credences to one's own future free actions. Indeed, some authors go further, insisting that one *cannot*. Kyburg (1988, p. 80) contends that to the extent that I am actually making a choice, I must regard that choice as free. In doing so, I *cannot* assign probabilities to my acting in one way rather than another (even though onlookers may be able to do so). Spohn (1977), Gilboa (1994), and Levi (1997) all argue for variants of this position. It is unclear, for example, how one can make sense of non-trivial subjective probabilities of one's future actions on the betting interpretation: one's fair price for a bet on such an action ought to collapse to 0 or 1. And yet acting one way or another can surely be doxastic possibilities for the agent in question. To be sure, these cases of probability gaps are

---

[10] Thanks here to John Cusbert.

controversial, and I am not convinced by them[11]; but it is noteworthy that these authors are apparently committed to there being further counterexamples to regularity due to credence gaps—indeed, for them *irregularity* is rationally required. But we need not join them to have our case against doxastic regularity as a norm—it suffices that irregularity is *rationally permissible*. It's enough if you rationally happen not to assign a credence to some future choice of yours, for whatever reason.

9. Ramifications of irregularity for Bayesian epistemology and decision theory

I have argued for two kinds of counterexamples to regularity: rational assignments of zero credences, and rational credence gaps, for doxastic possibilities. I now want to explore some of the unwelcome consequences these failures of regularity have for traditional Bayesian epistemology and decision theory.

Much of what's philosophically interesting about probability 0 events derives from interesting facts about the arithmetic of 0. Indeed, I think it is the most interesting of all numbers; it's surely the weirdest. Let's go through some of its idiosyncrasies, each one motivating a *philosophical* problem. In some ways probability 0 propositions behave like probability *gaps*. And probability gaps will cause similar problems. The two ways to stymie regularity will thus prompt us to rethink some Bayesian foundations.

i. Problems for the ratio analysis of conditional probability

As you learned in high school, if not earlier, you can't divide by 0. I made much of this fact in my (2003) critique of the usual ratio formula for conditional probability:

$$P(A \mid B) = P(A \cap B)/P(B) \text{ provided } P(B) > 0.$$

---

[11] I can surely make *comparative* judgments of the probabilities of my choices. And as Peter Vranas suggests, I can use past evidence on my choices in similar situations.

I argued that this is not a stipulative definition of conditional probability, but rather an *analysis* of a pretheoretical concept that we have. And it is a *bad* analysis. An important source of my discontent with it was what I called the *problem of the zero denominator*—cases where $P(B)$ is 0, but the conditional probability of A, given B, is apparently well-defined. For example, let our example of a probability 0 event be Williamson's coin tossed infinitely many times landing heads forever. (I choose it because it is easily stated, but you may replace it with another example if you prefer.) Intuition tells us, indeed yells at us, that the probability that the coin lands heads forever, *given* that it lands heads forever, is 1. But rather than telling us this, the ratio analysis is silent. And so it goes for any doxastically possible proposition of probability zero. For each such proposition, we have a violation of a seeming platitude about conditional probability: that the probability of a proposition *given that very proposition,* is 1. Surely that is about as fundamental a fact about conditional probability as there could be—on a par, I would say, with the platitude about logical consequence that every proposition entails itself.

Equally platitudinous, I submit, is the claim that the conditional probability of a proposition, given something else that entails that proposition, is 1: for all distinct X and Y, if X entails Y then the probability of Y, given X, is 1. But clearly this too is violated by the ratio analysis. Suppose that X and Y are distinct, X entails Y, but both have probability 0 (e.g., X = 'the coin lands heads on every toss', and Y = 'the coin lands heads on every toss after the first'). Then $P(Y \mid X)$ is undefined; but intuition demands that the conditional probability be 1. There are also less trivial examples. For instance, the probability that the coin lands heads on every toss, given that it lands heads on every toss after the first, is $\frac{1}{2}$. Again, all we get from the ratio analysis is silence—an uncomfortable silence.

Probability gaps create similar problems: undefined quantities do not enter into arithmetic operations at all, so in particular you can't divide by them. The probability that the dart lands in a particular non-measurable set N, *given* that it lands in N, is surely 1. But this is not something that the ratio analysis can tell you—if it tried, it would yield $P(N \mid N) = \text{undefined}/\text{undefined}$, which is undefined. The probability that the dart lands in [0, 1], given that it lands in N, is surely 1, but again the ratio analysis cannot say that. If any of the further examples of probability gaps that I canvassed go through, they too will provide similar counterexamples to the ratio analysis.

Regarding the ratio formula as merely providing a *constraint* rather than an *analysis* of conditional probability only makes me restate the objections. In that case, our theory of conditional probability is *incomplete*—remaining silent, when we want it to speak. Better incompleteness than falsehood, I suppose; but *that's* hardly a ringing endorsement of the traditional understanding of conditional probability.

ii. Problems for conditionalization

This problem for the conditional probability formula quickly becomes a problem for the updating rule of *conditionalization*, which is defined in terms of it.

Suppose that your degrees of belief are initially represented by a probability function $P_{\text{initial}}(-)$, and that you become certain of a piece of evidence E. What should be your new probability function P_{new}? Since you want to avoid any gratuitous changes in your degrees of belief that were not prompted by the evidence, P_{new} should be the minimal revision of P_{initial} subject to the constraint that $P_{\text{new}}(E) = 1$. The favoured updating rule among Bayesians is *conditionalization: P_{new}* is derived from P_{initial} by taking probabilities conditional on E, according to the schema:

(Conditionalization) $P_{\text{new}}(X) = P_{\text{initial}}(X \mid E)$ (provided $P_{\text{initial}}(E) > 0$)

But what are you supposed to do if you become certain of something that you initially gave probability 0? Suppose you *learn* that the coin landed heads on every toss: this is your total evidence. What should be your *new* probability that it landed heads on every toss? 1, surely. But $P_{initial}$(heads on every toss | heads on every toss) is undefined, so conditionalization cannot give you this advice.

Similarly, if you learn that the dart landed in *N*, your probability that it did so should become 1. But conditionalization cannot give you this advice. Or suppose that you learn that the coin landed heads on every toss after the first. This should *raise* your probability that it landed heads on every toss, to ½. Or suppose that you are initially ignorant *both* about whether the dart lands in *N*, *and* whether the first toss of the coin was heads. Your probability for the conjunction *dart lands in N & heads* is initially gappy. Then you learn that the dart did indeed land in *N*, but you learn nothing about the coin toss. Your probability for the conjunction should become ½. But again, conditionalization cannot say this.

Probability 0 cases are notoriously problematic for the Bayesian account of learning, but probability gaps are equally problematic. Now that regularity seems to be untenable, we have reason to embrace a more sophisticated account of conditional probability, one that allows conditional probabilities to be defined even when the conditions have probability zero, or are probability gaps. Popper (1959) has provided such an account, offering an axiomatization of primitive conditional probability functions—so-called "Popper functions".

iii. Problems for independence

0 is the multiplicative destroyer: multiply anything by it, and you get 0 back. This, in turn, spells trouble for the usual *factorization analysis* of probabilistic independence: *A* and *B* are *independent* just in case

$$P(A \cap B) = P(A)P(B).$$

Paralleling my discussion of conditional probability in the previous section: If this were simply a stipulative definition of a new technical term, then there could be no objection to it. But 'independence' is clearly a concept with which we were familiar before this analysis arrived on the scene, even more so than the concept of 'conditional probability'. The choice of word, after all, is no accident. Indeed, it has such a familiar ring to it that we are liable to think that this factorization account captures what we always meant by the word 'independence' in plain English. The concept of 'independence' of *A* and *B* is supposed to capture the idea of the *insensitivity* of (the probability of) *A*'s occurrence to *B*'s occurrence, or the information that *B*, and vice versa. It is at best an open question whether the factorization analysis has succeeded in capturing this idea.

It has not succeeded. In fact, I think that some 80 years of probability theory has simply got the notion of independence wrong. And independence is a staple of Bayesian epistemology. It has assumed the factorization analysis of independence—wrongly, I maintain. Branden Fitelson and I (MS) argue for this at greater length; some of the discussion there is repeated here.

According to the factorization analysis, anything with probability 0 has the bizarre property of being probabilistically independent of *itself*:

If $P(X) = 0$, then $P(X \cap X) = 0 = P(X)P(X)$.

Yet identity is the ultimate case of *de*pendence of the kind that we are trying to characterize. Suppose that you are interested in whether the coin landed heads on every toss. Now suppose you get as evidence: the coin landed heads on every toss. Can you tell me with a straight face that this evidence is *entirely uninformative* regarding the proposition of interest? On the contrary, nothing could be more informative! This is a refutation of that analysis.

More generally, one would expect the dependence of *X* on *X* to be *maximal*, not minimal. Much as I took it to be a platitude that

for every proposition *X*, the probability of *X* given *X* is 1,

so I take it to be a platitude that for every proposition *X*, *X* is *de*pendent on X. What better support, or evidence, for *X* could there be than *X* itself?

More generally, according to the factorization analysis, any proposition with extreme probability has the bizarre property of being probabilistically independent of *anything*. This includes its negation. It also includes anything that entails the proposition, and anything that the proposition entails. Much as I took it to be a platitude that

for all *X* and *Y*, if *X* entails *Y* then the probability of *Y*, given *X*, is 1,

so I take it to be a platitude that for every proposition *X*, *X* is *de*pendent on anything that entails *X*.

The situation is *worse* for Kolmogorov's account of independence than it was for his ratio account of conditional probability. For that was guilty of what we might call *sins of omission*—delivering no answers where there were clear answers to be delivered (in the coin case, and so on). But the factorization account of independence is guilty of *sins of commission*—delivering *incorrect* answers. It explicitly says that certain pairs of events are independent when they are obviously not. These are

outright counterexamples to it. I think this is a scandal of orthodox probability theory, and of philosophical positions that employ it uncritically, such as Bayesianism.

Probability gaps are similarly problematic for the factorization analysis. Suppose that you are interested in whether the dart landed in *N*. Now suppose you get as evidence: the dart landed in *N*. I say that this is maximally informative regarding the proposition of interest. But the factorization analysis does not say anything.

I think that a theory of independence should be based on Popper's account of conditional probability. Branden Fitelson and I (MS) are developing such a theory.

iv. Problems for expected utility theory

Curiously, while 0 is the most potent of all numbers when it comes to multiplication, it's the most impotent when it comes to addition and subtraction. After all, it's the additive identity: adding it to any number makes no difference. This, in turn, creates difficulties for *decision theory*. Arguably the two most important foundations of decision theory are the notion of *expected utility*, and *dominance reasoning*. The former is a measure of choiceworthiness of an option: the weighted average of the utilities associated with that option in each possible state of the world, the weights given by corresponding probabilities that those states are realized. The latter codifies the platitude that if one option is at least as good as another in every possible state, and strictly better in at least one possible state, then it is preferable. (We should add that the options have no causal influence on the states.) These are compelling ideas that are largely taken for granted.

And yet consideration of probability 0 propositions shows that they can give conflicting verdicts. Suppose that two options yield the same utility except on a proposition of probability 0; but if that proposition occurs, option 1 is far superior to

option 2. For example, suppose that we throw the dart at the [0, 1] interval. You can choose between these two options:

Option 1: If the dart hits a rational number, you get a million dollars; otherwise, you get nothing.

Option 2: You get nothing.

Expected utility theory says that these options are equally good: they both have an expected utility of 0. But dominance reasoning says that option 1 is strictly better than option 2. Which is it to be? Intuition shouts that option 1 is better. This is a counterexample to expected utility theory.

Again, this case is *worse* than that of the conditional probability ratio formula. It isn't that we get an uncomfortable silence. Rather, two of our decision rules speak, and they contradict each other. Moreover, we know that expected utility theory is speaking falsely.[12]

Similar problems are caused by probability gaps. Suppose you can choose between these two options:

Option 1: If the dart hits N, you get a million dollars; otherwise, you get nothing.

Option 2: You get nothing.

Expected utility theory cannot compare these options: option 1's expected utility is undefined. But dominance reasoning says that option 1 is better than option 2. Which is it to be? Intuition shouts that option 1 is better. This reveals expected utility theory to be incomplete.

---

[12] Getting fancy with causal as opposed to evidential decision theory won't help. This is as much a problem for causal decision theory as it is for evidential decision theory. Take an outcome that has 0 probabilistic weight (for a given act) according to whichever causal decision theory you prefer; now consider another act that improves just that outcome. You ought to prefer this sweetened act, but its causal expected utility is the same as the original act, since the sweetening makes no contribution to that expected utility.

Regularity would allow us to have a complete, unified decision theory. Dominance reasoning would never conflict with expected utility theory, since every state would get *positive* probability. *A fortiori*, a state in which one option is strictly better than another would get positive probability, and thereby make a difference to the expected utility calculations. Moreover, there would be no threat of expected utilities going undefined, since there would be no threat of probability gaps. Indeed, dominance reasoning would be rendered redundant: any verdict it could deliver would automatically be delivered by expected utilities. The demise of regularity spells the demise of this hope.[13]

10. Conclusion

Irregularity makes things go bad for the orthodox Bayesian; that is a prima facie reason to insist on regularity. The trouble is that regularity appears to be untenable. I think, then, that irregularity is a reason for the orthodox Bayesian to become *unorthodox*. I have already advocated replacing the orthodox theory of conditional probability, conditionalization, and independence with alternatives based on Popper functions. Expected utility theory appears to be similarly in need of revision.

One might object that a theory based on Popper functions is more complicated than the elegant, simple theory that we previously had. But elegance and simplicity are no excuses for the theory's inadequacies. Moreover, the theory had to be complicated in any case by the introduction of infinitesimals; and it seems that even they are not enough to overcome the inadequacies.

---

[13] One could contend that expected utility falls silent about what to choose when multiple options have equal utility, and one can fill the gaps with, e.g., a dominance principle. One could contend something similar about undefined utilities: expected utility reasoning just does not tell us what to do in these cases, and we have to fall back on something else in the decision theorist's toolkit. The cost, of course, is that this doesn't give us a unified decision theory, just a jumble of tools. (Thanks here to Rachael Briggs.)

Or perhaps some genius will come along one day with an elegant theory that preserves regularity after all? Fingers crossed![14]

<div align="right">

Philosophy Program
Research School of Social Sciences
Australian National University

</div>

REFERENCES

Appiah, Anthony (1985): *Assertion and Conditionals*, Cambridge University Press, Cambridge.

Bernstein, A. R., and F. Wattenberg (1969): "Non-standard measure theory", in W. A. J. Luxemburg, ed., *Applications of Model Theory of Algebra, Analysis, and Probability*. New York: Holt, Rinehard and Winston.

Carnap, Rudolf (1963)" "Carnap's Intellectual Autobiography", in *The Philosophy of Rudolf Carnap*, The Library of Living Philosophers, Vol. XI, ed. Paul Arthur Schilpp.

Chalmers, David (forthcoming): "The Nature of Epistemic Space", in Andy Egan and Brian Weatherson, eds, *Epistemic Modality*. Oxford University Press, forthcoming.

de Finetti, Bruno (1931): "Probabilismo", *Logos*, 163-219.

Eagle, Antony (2011): "Chance versus Randomness", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/chance-randomness/>.

Earman, John (1986): *A Primer on Determinism*, Dordrecht: D. Reidel.

Easwaran, Kenny (MS): "Regularity and Infinitesimal Credences".

Edwards, W., Lindman, H., and Savage, L. J. (1963): "Bayesian Statistical Inference for Psychological Research", *Psychological Review* 70, 193-242.

Fitelson, Branden and Alan Hájek (MS): "Declarations of Independence".

Gilboa, Itzhak (1994): 'Can Free Choice Be Known?', in C. Bicchieri, R. Jeffrey and B. Skyrms (eds.), *The Logic of Strategy*, Oxford University Press.

Hájek, Alan (2003): "What Conditional Probability Could Not Be", *Synthese*, Vol. 137, No. 3, December 2003, 273-323.

Hájek, Alan (2009): "Arguments For—Or Against—Probabilism?", *British Journal of Philosophy* 59 (December), 793-819. Reprinted in *Degrees of Belief*, eds. Franz Huber and Christoph Schmidt-Petri, Springer, 2009, 229-251.

Jackson, Frank (1987): *Conditionals*, Blackwell.

Jaynes, E. T. (2003): *Probability Theory: The Logic of Science*, Cambridge: Cambridge University Press.

Jeffrey, Richard (1992): *Probability and the Art of Judgment*, Cambridge University Press.

Jeffreys, Harold (1961): *Theory of Probability*, 3rd ed., Oxford, Clarendon Press.

Keynes, John Maynard (1921): *A Treatise on Probability*.

Kyburg, Henry (1988): 'Powers', in William L. Harper and Brian Skyrms (eds.), *Causation in Decision*, *Belief Change*, *and Statistics*, Kluwer Academic Publishers, pp. 71–82.

Levi, Isaac (1997): *The Covenant of Reason: Rationality and the Commitments of Thought*, Cambridge University Press.

Lewis, David (1973): *Counterfactuals*, Harvard University Press, Cambridge, Massachusetts.

Lewis, David (1980): "A Subjectivist's Guide to Objective Chance", in *Studies in Inductive Logic and Probability*, Vol II., ed. Richard C. Jeffrey, University of California Press.

Lewis, David (1986): *Philosophical Papers*, Volume II, Oxford University Press.

Norton, John (2008): "The Dome: An Unexpectedly Simple Failure of Determinism", *Philosophy of Science*, 75: 786–98.

Popper, Karl (1959): *The Logic of Scientific Discovery*, Basic Books; reprint edition 1992, Routledge.

Pruss, Alexander (MS): "Regularity and Cardinality".

Rényi, Alfred (1970): *Foundations of Probability*, Holden-Day.

Shimony, Abner (1955): "Coherence and the Axioms of Confirmation", *Journal of Symbolic Logic* 20, 1-28.

Shimony, Abner (1970): "Scientific Inference", in Robert Colodny (ed.), *The Nature and Function of Scientific Theories*, University of Pittsburgh Press.

Skyrms, Brian (1995): "Strict Coherence, Sigma Coherence, and the Metaphysics of Quantity", *Philosophical Studies* 77, No. 1, 39-55.

Spohn, Wolfgang (1977): "Where Luce and Krantz Do Really Generalize Savage's Decision Model", *Erkenntnis* **11**, 113–134.

Stalnaker, Robert C. (1970): "Probability and Conditionals", *Philosophy of Science* 37, 64-80.

Tang, Weng Hong (MS): "Regularity Reformulated".

Williamson, Timothy (2007): "How Probable Is an Infinite Sequence of Heads?" *Analysis* 67: 173–80.