Philosophers' Imprint

VOLUME 19, NO. 23
JUNE 2019

IMPRECISE CHANCE AND THE BEST SYSTEM ANALYSIS

Luke Fenton-Glynn

University College London

1. Introduction

Much recent philosophical attention has been devoted to the prospects of the Best System Analysis (BSA) for yielding *high-level* chances, including statistical mechanical and special science chances. But a foundational worry about the BSA lurks: There don't appear to be uniquely correct measures of the degree to which a system exhibits theoretical virtues, such as simplicity, strength, and fit. Nor does there appear to be a uniquely correct exchange rate at which the theoretical virtues trade off against one another in the determination of a best system. Further, there may be no *robustly* best system; no system that comes out best under *any* reasonable measures of the theoretical virtues and exchange rate between them.

In the following, I argue that there plausibly *is* a set of tied-for-best systems for our world (specifically, a set of very good systems, but no robustly *best* system) and that some of these systems entail different statistical mechanical probabilities. I argue that the advocate of the BSA should conclude that (some of) the Best System chances for our world are *imprecise* (*set valued*). Since I don't mount a general defense of the BSA, my thesis is conditional: If one adopts the BSA, then one should believe in the existence of imprecise chances.[1]

In Section 2, I outline the BSA. In Section 3, I describe a recent argument for the conclusion that the best system for our world is one that entails the fundamental dynamical laws together with the probabilities of statistical mechanics (SM) and, derivatively, the probabilities of the special sciences. Specifically, I focus on the argument as it has been made in connection with so-called 'Globalist' approaches to axiomatizing SM: approaches that seek to derive SM from axioms concerning the initial state of the universe as a whole. In Section 4, I argue that plausibly there's no robustly best 'Globalist' system for our world, but

[1] Even for someone who rejects the BSA, the discussion of the relation between imprecise chance and credence in Section 6 below may be of interest because, as noted in that section, there also appear to be reasons independent of the BSA for believing in imprecise chances.

rather a set of tied-for-best systems, some of which entail diverging probabilities for SM. In Section 5, I turn my attention to so-called 'Localist' approaches to SM — approaches that seek to derive SM from axioms that concern the initial states of various *subsystems* of the universe — and argue that a similar conclusion follows in that context. In Section 6, I seek to articulate the role that chance, when imprecise, plays in constraining rational credence. In Section 7, I argue that the sets of probabilities entailed by the tied-for-best systems play that role (or at least that they do so as well as unique probabilities associated with a robustly best system would play the role defined by precise chance-credence coordination principles) and so should be regarded by the advocate of the BSA as constituting imprecise chances.

2. The BSA

According to the BSA, which received its most significant development by Lewis (1983, 366–8; 1986, xi, xiv–v, 121–31; 1994, esp. 478–82), the laws are those regularities that are entailed by the set of axioms that best systematizes the entire history of the world. The objective chances are probabilities entailed by this best system. Lewis (1994, esp. 473–5) combines the BSA with the thesis of *Humean Supervenience* (HS): the thesis that the laws and chances supervene upon the *Humean mosaic*, that is, (roughly speaking) upon the distribution of categorical (i.e. non-modal), locally-instantiated properties throughout all of space-time. When the BSA is combined with HS, the idea is that what gets systematized by the various competing systems is (parts of) the Humean mosaic. I shan't challenge HS in what follows: Most discussion of the BSA has been conducted against the background of this assumption.

The goodness with which a set of axioms systematizes the Humean mosaic is a function of the degree to which it exhibits the theoretical virtues of simplicity, strength, and fit. A system is *strong* to the extent that it says "what will happen or what the chances will be when situations of a certain kind arise" (Lewis 1994, 480). A system is *simple* to the extent that it comprises fewer axioms, or those axioms have simpler forms (e.g. linear equations are simpler than polynomials of degree greater than 1). Often greater strength can be achieved at a cost in terms of simplicity by adding or complicating axioms.

If the world is a certain way (if it contains lots of stochastic-looking events), then a candidate system may achieve a good deal of strength with little cost in simplicity by including probabilistic axioms, which define a probability function (cf. Loewer 2004, 1119). It's natural to take this to be a *conditional* function $ch(X|Y)$ that maps proposition pairs $\langle X, Y\rangle$ onto the reals in the $[0,1]$ interval, perhaps in accordance with the Rényi-Popper axioms (Rényi 1970; Popper 1972, New Appendicies *ii-*v).

The reason for thinking that the probabilistic axioms of a candidate system define, in the first place, *conditional* probabilities is that, as Lewis (1994, 480) indicates, probabilistic axioms "say... what the chances will be when situations of a certain kind arise". Where X is an outcome-specifying proposition, and Y is a proposition specifying that a situation of the relevant kind arises, the probabilistic axioms thus yield conditional probabilities, $ch(X|Y)$.[2]

Where a system entails a probability function, it may exhibit the theoretical desideratum of *fit* to a greater or lesser degree. As Lewis (1994, 480) explains it, a system *fits* the actual course of history well to the extent that the associated probability function assigns that history a higher probability (perhaps *conditional upon* initial conditions).[3]

I think that an adjustment to Lewis's definition of fit should be considered. Plausibly, the probability functions entailed by many reasonable systems *aren't defined* on propositions concerning entire possible courses of history of the world. For instance, it seems that this is true of orthodox Quantum Mechanics (QM) because it doesn't assign probabilities to measurement events (Ismael 2008, 301; Hájek 2003b, 305–8).

[2] Hájek (2003a; 2003b; 2007) provides more detailed arguments that conditional chance is more basic than unconditional chance. The arguments to follow could all be rearticulated, without loss of strength, on the alternative assumption that unconditional chance is basic.

[3] Lewis's notion of fit applies only where systems don't incorporate chance distributions over infinite sets (Elga 2004, 68; Frigg and Hoefer 2015, 554). See Elga (2004, esp. 71–2) for an extension to infinite cases.

The great success of QM is rather that the probabilities it assigns to localized outcomes — such as a silver atom's being deflected in a certain direction — given localized experimental setups — such as that atom's being fired through a Stern-Gerlach device — closely match the actual relative frequencies that obtain in such cases (cf. Ismael 2008, 301, 301n).

In light of this, it seems plausible that, on the appropriate notion of fit, a system should count as well-fitting to the extent that the conditional probabilities that it does entail are close to the actual relative frequencies, even if these conditional probabilities are not probabilities for complete future histories of the world conditional upon the (complete) initial conditions of the universe (cf. Schwarz 2014, 94). But, however exactly these details are worked out, the claim made by the Best System analyst is that the system that strikes the best balance between the theoretical virtues of simplicity, strength, and fit is the Best System, that the laws of nature are regularities entailed by the Best System, and that the probability function entailed by the Best System is the *chance function* for the world.

3. The BSA and SM

Lewis (1986, 117–21; 1994) appears to have thought that the probability function associated with the best system for our world would simply be the fundamental physical probability function: the function that yields all and only the probabilities entailed by QM, or whatever fundamental physical theory replaces it. Yet Loewer (2001; 2007; 2008; 2012a; 2012b) has influentially argued that the probabilities of SM are also entailed by the best system for our world, and therefore are genuine objective chances. Loewer appeals to the axiomatization of SM described by Albert (2000, Chs. 3–4), who suggests that SM can be derived from the following:

(FD) the fundamental dynamical laws;

(PH) a proposition characterizing the initial conditions of the universe as constituting a special low-entropy state; and

(SP) a uniform probability distribution (on the standard Lebesgue measure) over the region of microphysical phase space associated with that low-entropy state.[4]

Loewer (2012a, 16; 2012b, 124), following a suggestion of Albert's (see Loewer 2012a, 16n), dubs the conjunction of FD, PH, and SP 'the Mentaculus'.

The argument that the SM probabilities are derivable from the Mentaculus goes roughly as follows. Consider the region of the universe's phase space associated with its low-entropy initial state described by PH. Relative to the total volume of that region, the volume taken up by micro-states that lead (by FD) to fairly sustained entropy increase until thermodynamic equilibrium is reached, and to the universe staying at or close to equilibrium thereafter, is (on the Lebesgue measure) extremely high. Consequently, the uniform probability distribution (given by SP) over the entire region yields an extremely high probability of the universe following such a path.

When it comes to (approximately) isolated subsystems of the universe, the idea is that since a system's becoming approximately isolated isn't itself correlated with its initial micro-state being one that leads (via FD) to entropy decrease, it's extremely likely that any such subsystem that's in initial disequilibrium (and which is thus such that most of the subregion of *its* phase space compatible with *its* initial disequilibrium macrostate is taken up by microstates that, according to FD, are on entropy-increasing trajectories through that phase space) will increase in entropy over time (Loewer 2007, 302; 2012a, 124–5; 2012b, 17; Albert 2000, 81–5). It's thus claimed that the Mentaculus entails a probabilistic, SM approximation to the Second Law of Thermodynam-

[4] Where FD are quantum rather than classical, the uniform probability distribution is not over classical phase space, but rather over the set of quantum states compatible with PH (Albert 2000, 131–3).

ics (SLT).

Albert (2000, e.g. 22, 28–9; 2012, 28–33) and Loewer (2007, 306; 2008, 159–62; 2012a, 18) have argued that the Mentaculus entails probabilities for the special sciences. One reason for thinking this is that many special sciences are concerned with entropy-increasing processes: These include geological processes of erosion (Elga 2001, 322); meteorological processes such as the evolution of pressure systems (Loewer 2008, 159); biological processes of ageing (Albert 2000, 22), inheritance, and the workings of neurons; chemical processes involving reactions; and so on. So it seems that the Mentaculus, if it entails probabilities for thermodynamic processes in general, entails probabilities that pertain to (at least many of) the sorts of processes of concern to the special sciences. Loewer (2001, 618; 2007, 305; 2008, 159; 2012b, 129) thus claims that the Mentaculus is much stronger than a system comprising FD alone. Since it's apparently not much more complicated (it only requires the addition of PH and SP), Loewer (2001, 618; 2007, 305; 2008, 159; 2012b, 129) claims that it's plausibly the *best* system for our world.[5,6]

The claim that the Mentaculus is stronger than a system comprising FD alone is plausible, especially according to Lewis's construal of strength, on which a system is strong to the extent that it says "what will happen or what the chances will be when situations of a certain kind arise" (Lewis 1994, 480). If Albert and Loewer are correct, the Mentaculus tells us what the chances will be when certain kinds of situation arise concerning which FD is silent.

Consider a thermodynamically isolated system, *S*. The Mentaculus entails the chances for *S*'s future thermodynamic evolution given that a situation of the kind *S is in such-and-such a thermodynamic state* arises. FD (even together with bridge principles) doesn't tell us what will happen or what the chances will be when situations of such a kind arise. FD tells us what will happen when situations of the more specific kind, *S is located at such-and-such a point in its microphysical state space*, arise. Since thermodynamic states of systems are multiply realizable by points in microphysical state space, the mere fact that *S* is in the relevant thermodynamic state is not enough for FD to provide us with any predictions about *S*'s future evolution. It is, however, enough for SM to do so. Hence, by entailing SM, the Mentaculus has greater strength, in the relevant sense, than a system comprising FD alone.

Still, some modification to the BSA, as articulated by Lewis, is required if the Mentaculus is to be a candidate Best System. Observing that the (syntactic) simplicity of a system is relative to the vocabulary in which it's expressed, Lewis (1983, 367–8) takes only those systems formulated in perfectly natural kind terms to be candidate Best Systems. The trouble is that, as Schaffer (2007, 130) notes (cf. Cohen and Callender 2009, 14), the Mentaculus, as formulated by Albert, employs predicates like 'low entropy' that correspond to properties that aren't perfectly natural. Moreover, if translated into a language with only perfectly natural kind terms, the Mentaculus is likely to be syntactically very complex and thus (if we take Lewis's approach) unlikely to be a leading contender for Best Systemhood.

Yet, as Lewis (1983, 368) recognizes, naturalness admits of degrees. I think that it's plausible to take naturalness of the predicates that a system employs to be a theoretical virtue, to be weighed alongside its simplicity, strength, and fit. If an axiom system is able to achieve great simplicity and strength and fit by employing a not-too-unnatural predicate like 'low entropy' — as the Mentaculus does — then it's a plau-

5 By including PH, the Mentaculus includes information about the initial conditions of the universe. But Lewis (1983, 367) is sympathetic to the idea that the best system might include such information.

6 One common criticism of Albert and Loewer is that they don't give enough details of precisely how special science probabilities derive from the Mentaculus (Callender and Cohen 2010, 437–9; Callender 2011, esp. 99–103). Cohen and Callender (2009; 2010) propose what they call a 'Better Best System Analysis' (BBSA) of laws and chance, according to which the chances for a special science are given by the best system that's framed in that science's proprietary vocabulary (cf. Schrenk 2008). Yet since, on the BBSA, the relative goodness of such special science systems depends on their balance of simplicity, strength, and fit, the arguments below that BSA chances are imprecise are liable to apply *mutatis mutandis* to BBSA chances.

sible Best System.[7] Of course it may seem a difficult task to specify a uniquely appropriate 'naturalness-of-vocabulary' metric and exchange rate at which naturalness trades off against the other theoretical virtues in the determination of an overall best system. But, as we shall see in the next section, the BSA faces problems similar to this anyway.

4. Ties Between Systems

Lewis acknowledged that the BSA isn't unproblematic. As he says (in a passage worth quoting at length):

> The worst problem about the best-system analysis is that when we ask where the standards of simplicity and strength and balance come from, the answer may seem to be that they come from us. Now, some ratbag idealist might say that if we don't like the misfortunes that the laws of nature visit upon us, we can change the laws — in fact, we can make them always have been different — just by changing the way we think! (Talk about the power of positive thinking.) It would be very bad if my analysis endorsed such lunacy. . . . [Yet] if nature is kind to us, the problem needn't arise. I suppose our standards of simplicity and strength and balance are only partly a matter of psychology. It's not because of how we happen to think that a linear function is simpler than a quartic or a step function . . . . Maybe some of the exchange rates between aspects of simplicity, etc., are a psychological matter, but not just anything goes. If nature is kind, the best system will be *robustly* best — so far ahead of its rivals that it will come out first under any standards of simplicity and strength and balance. We have no guarantee that nature is kind in this way, but no evidence that it isn't. It's a reasonable hope. (Lewis 1994, 479)

[7] For alternative ways of relaxing the requirement that candidate systems be framed in only perfectly natural kind terms, see Cohen and Callender (2009; 2010), Dunn (2011, 87–92), Schrenk (2008), Frisch (2014, Section 5), and Frigg and Hoefer (2015, 560–2).

Lewis's view seems to be that, while there are some objective constraints on what counts as a genuine measure of simplicity ('It's not because of how we happen to think that a linear function is simpler than a quartic or step function'), strength, and perhaps even what counts as an appropriate balance between them, these constraints aren't sufficient to determine unique measures of simplicity and strength, and a unique exchange rate between the theoretical virtues as the objectively correct ones. The worry is that, in order to arrive at unique measures of the theoretical virtues and a unique exchange rate between them, whatever objective constraints there are will need to be supplemented by constraints that are 'psychological' and relative to the ways of thinking of creatures like us.

For instance, it's plausible that the particular, contingent cognitive capacities and limitations of *homo sapiens* mean that axioms of certain syntactic forms are easier than others for us to grasp and to work with in generating predictions and explanations. This makes us judge them simple and strong, while perhaps other creatures with different capacities and limitations would judge different sets of (true) axioms to be simpler and/or stronger. Although perhaps not anything goes — there are certain objective constraints on what counts as a genuine measure of *simplicity*, a genuine measure of *strength*, and an appropriate way of balancing the theoretical virtues against one another — provided that the objective constraints don't single out unique measures and balances, there may be no objective fact of the matter about what the 'best' system is and so what the laws and chances are.

Although Lewis appears to grant that objective constraints may not be sufficient to determine unique measures of simplicity, strength, and a unique exchange rate between the theoretical virtues, he takes it to be 'a reasonable hope' that this underdetermination won't issue in an underdetermination of what counts as the 'best' system; that nature's kind enough to furnish a system that's victorious on any measures and exchange rate compatible with these objective constraints.

Of the possibility that this hope is unfulfilled, Lewis says:

> I can admit that *if* nature were unkind, and *if* disagreeing rival systems were running neck-and-neck, [then] lawhood might be a psychological matter, and that would be very peculiar. I can even concede that in that case the theorems of the barely-best system would not very well deserve the name of laws. ...(Likewise for the threat that two very different systems are tied for best. . . . I used to say that the laws are then the theorems common to both systems, which could leave us with next to no laws. Now I'll admit that in this unfortunate case there would be no very good deservers of the name of laws. (Lewis 1994, 479; italics original)[8]

Lewis also appears to think that if there's no clear winner of the best system competition, then there may be nothing very well deserving of the name *chance*:

> How well the laws and chances deserve their names should depend on how kind nature has been in providing a decisive front runner [in the best system competition for our world]. (Lewis 1994, 481)

The trouble is that it has seemed to a number of authors that, particularly when we consider axioms systems that entail SM, nature hasn't been all that kind. Specifically, it appears that there is indeed a set of rival systems — each of which entails a slightly different probability function and, correspondingly, slightly different probabilistic versions of the principles of thermodynamics — that are such that it seems unlikely that one is picked out unanimously as *best* by all combinations of measures of simplicity and strength, and exchange rates between the theoretical virtues that are compatible with whatever objective constraints there are on such combinations.

As I shall explain in Section 7, I'm not as averse as Lewis appears to be to the notion that what counts as the best system and hence as the laws and chances is, to an extent, relativized to groups of epistemic agents and their particular capacities and limitations.[9] What I find more troubling is that, even allowing that the standards may be partly relativized in this way, there *still* seems to be underdetermination. Standards may be relativized without being arbitrary: This will be so provided that there's a fact of the matter about what standards will produce a winning system that's best *for epistemic agents of a given sort*, perhaps by being the *most useful* to agents of that sort. However, in light of the similarity of a range of very good systems entailing SM-like probabilities, it seems doubtful (as I'll argue in more detail in Section 7) that there's a uniquely most useful system *for agents like us* and a corresponding unique set of standards. My suggestion will be that, among the set of very good systems and the sets of standards yielding them, we shouldn't (arbitrarily) pick a unique winner and a uniquely correct set of standards.

Even so, the best system analyst ought neither to draw the conclusion that there is nothing very well deserving of the name of laws or chances in our world (*pace* Beisbart 2014), nor even that the laws and chances are limited to those theorems and probabilities that all rival systems that aren't determinately non-best entail. Rather, I'll argue, the best system analyst ought to regard some of the chances, and indeed some of the laws, for our world as indeterminate or *imprecise*, corresponding to the indeterminacy of the (objectively/non-arbitrarily) Best System for our world (for creatures like us)[10]. First, however, it's

[8] In the parentheses, Lewis speaks of a scenario in which two 'very different' systems are tied for best. I'll make no claim that the systems for our world that I'll argue are plausibly tied for best are very different, but they are different, and (crucially) entail divergent probabilities for certain propositions.

[9] Other Humeans, such as Hoefer (2007, 572), Albert (2012, 35–7), and possibly Loewer (1996, 191–2; 2004, 1122) seem sympathetic to this idea.

[10] I'll drop this rider in what follows, though I'll return again to the idea that best systemhood displays this sort of relativity in Section 7. For now, we can pretend that it doesn't. The point of raising the issue of relativity in the main text above was just to point out that I don't think that, even if we embrace it, we have a solution to the underdetermination of best systemhood.

necessary to show why consideration of SM points to the conclusion that nature hasn't furnished us with a clear winner of the Best System competition.

To see this, let's assume that the Mentaculus does indeed entail the SM probabilities, and that Loewer is correct that it constitutes a better system for our world than one comprising FD alone. If the Mentaculus comes out *best*, then the generalizations that it entails count as laws on the BSA, and the probabilities that it entails, including the SM probabilities, count as chances. But a system comprising FD alone isn't the only rival to the Mentaculus. Another rival is a system comprising FD plus an axiom giving the *precise initial conditions* (PICs) of the universe (Schaffer 2007, 130–2; Hoefer 2007, 560; Beisbart 2014, 518–9): call this the FD + PICs system. If FD is deterministic, then FD + PICs is a very strong system. Schaffer (2007, 131–2) suggests that it's maximally strong, while Hoefer (2007, 560) questions this. If FD is probabilistic, then it may still be strong, depending on the exact nature of FD. However, a proposition specifying PICs would be complex when expressed in a vocabulary that contained only perfectly natural kind terms and would be simple only when expressed in a vocabulary that had predicates corresponding to highly unnatural kinds.

Now, it's fairly plausible that we should take FD + PICs to be determinately not the best system. That's because, if FD is deterministic, it has as theorems things that we know not to be laws (cf. Woodward 2014, 97–100),[11] such as *all planets orbit the sun in the same direction* (and — in case we think we can discount this as a law in virtue of its reference to a particular object — *all stars of mass m are such that all planets orbiting them orbit in the same direction*, where *m* is the mass of the sun given precisely enough to ensure that no other star shares that precise mass). Even if FD is probabilistic, there may still be problems, for it may well come out as a theorem that there's a high probability that this generalization holds, in which case we're in danger of classifying it as a probabilistic law despite its accidental nature.

The key point here is that a metaphysical theory of lawhood must respect certain independent facts that contribute to determining which generalizations count as laws of nature. These facts concern the generalizations' ability to play the *role* of laws by explaining, supporting counterfactuals, predictions, and causal relations, being confirmed by their instances, and entailing probabilities that play the chance role (of guiding rational credence and explaining outcomes and frequencies of outcomes). If we're to have any data to test our metaphysical theories of lawhood against, whether a generalization plays these aspects of the law role must at least partly be a matter that's independent of our metaphysical theories.

Such facts provide an additional source of non-arbitrary independent constraint upon the appropriate standards of simplicity, strength, and balance for determining what counts as the best system (cf. Loewer 1996, 185). Put another way: Any version of the BSA that incorporates standards which are such as to entail that clearcut cases of non-laws are in fact laws (or vice versa) isn't a good metaphysical theory of laws. As already noted, it's plausible that there are regularities that are theorems of FD + PICs that are clear cases of non-laws. Therefore, standards of simplicity, strength, and balance that don't entail that FD + PICs is determinately *not* the best system are inappropriate standards to incorporate into the BSA.

Still, there are further rivals to the Mentaculus besides FD + PICs which aren't so easily ruled out. As Beisbart (2014, 519) observes, a system that applies a uniform probability distribution to a sub-region of the region to which the Mentaculus's SP applies a uniform distribution to (namely the region compatible with the Mentaculus's PH), where this sub-region contains the universe's PICs, will plausibly be better-fitting than the Mentaculus, but will also be less simple because we need to specify more information about the initial conditions of the universe than is implied by PH. This can be done by being more precise about the initial entropy level of the universe, or by actually specifying (more precisely) the values of macro-variables concerning, for instance,

[11] Thanks to an anonymous referee for pressing me on this point and for pointing out the correct response for the best system analyst.

temperatures and pressures at some initial time, or even by specifying some (but not all — as FD + PICs does) of the microphysical details of the universe at some initial time (cf. Beisbart 2014, 519).

The worry is that, by choosing different sub-regions of the phase space region associated with PH to apply a uniform probability distribution to, we get a range of candidate systems (cf. Schaffer 2007, 131n; Beisbart 2014, 519–20). At one extreme, we have the Mentaculus (which applies a uniform distribution over the whole region of phase space compatible with PH). We then have a continuum of systems involving the application of the uniform distribution to smaller and smaller sub-regions of the phase space region compatible with PH, each containing PICs. In the limit, this yields a system equivalent to FD + PICs. As Beisbart (2014, 519) suggests, such systems would appear to be increasingly well fitting, since they concentrate a higher and higher probability in a smaller and smaller region around the PICs and thus assign an increasingly high probability to the actual macroscopic course of events. But they are also increasingly complex, since picking out the progressively smaller sub-regions requires building into the axioms increasing amounts of information about PICs.

FD + PICs itself should be taken to be beyond the pale in virtue of its complexity (cf. Beisbart 2014, 518–9) because it entails accidental generalizations. Yet one would need a very large amount of information about initial conditions for FD to entail generalizations like *all planets orbit the sun in the same direction* (or for FD, together with a probability distribution over phase space, to entail a high probability of such generalizations holding). So there seems quite a lot of scope for a system to be more specific about the phase space region the PICs occupy — and thus better fitting — than the Mentaculus without going beyond the pale by entailing accidental generalizations like the one described. Given this fact, it appears that we're confronted with a range of systems each of which may come out best according to some reasonable standards of strength, simplicity, and balance.

One can also consider rivals to the Mentaculus that incorporate different probability distributions over the phase space region compatible with PH. In some cases, systems that specify different probability distributions will be equivalent to systems that specify different sub-regions of the phase space compatible with PH to apply the uniform distribution to. For instance, Frigg and Hoefer (2015, 572) suggest that among the competitors to the Mentaculus is an axiom system that — instead of SP — contains an axiom that specifies "a peaked distribution, nearly Dirac-delta style" whose peak is at the (point-sized) PICs. The delta distribution assigns probability 1 to the point at which it is peaked, and 0 to all regions that don't include this point. A system incorporating the delta distribution, peaked at the PICs, is effectively equivalent to FD + PICs.

Beisbart (2014, 520) is correct to challenge Frigg and Hoefer's (2015, 572) contention that a system comprising FD together with (something near) the delta distribution peaked at the PICs is clearly superior to the Mentaculus. Frigg and Hoefer claim that the delta distribution is just as simple as the uniform distribution specified by SP while yielding much better fit to the actual frequencies. But as Beisbart (2014, 520) points out, "[t]he delta distribution has a simple functional form, but to pick one particular delta function, you have to specify the location of the peak . . . i.e., the whole initial condition, and this is not simple at all!" So the system that invokes the delta distribution is as complex as the FD + PICs system and, because both systems are liable to entail (high probabilities for) accidental generalizations like those described above, we should take them to be beyond the pale in terms of their complexity.

Nevertheless, there's a range of systems intermediate in complexity (and fit) between the Mentaculus and the system that incorporates the Dirac-delta distribution: specifically, systems that incorporate non-flat distributions that assign increasingly high probability to smaller and smaller regions of phase space containing the PICs.[12] Many of those,

[12] Not surprisingly, given that a probability distribution can't be specified independently of a measure, different choices of measure make a difference to SM probabilities — so there may be rivals to the Mentaculus incorporating different measures. SP appeals to a probability distribution that's uniform

especially those that lie toward the Mentaculus's end of the spectrum, will not have the untoward entailments that the system incorporating the delta distribution does, and are credible contenders for best systemhood.

It's a non-trivial question precisely how different choices of region of the universe's phase space containing PICs and different choices of probability distributions over such regions translate into SM probabilities at later times. Albert (2000, 67) suggests that, provided that the subregion of the phase space region picked out by PH we apply a probability distribution to is regularly-shaped and not too small, and provided the probability distribution applied to it is reasonably 'smooth' (i.e. the probability density varies only negligibly over small distances in the phase space) with respect to the Lebesgue measure, then the resulting probabilities of normal thermodynamic behavior (i.e. monotonic entropy increase) will diverge only a little from those entailed by the Mentaculus.

Frigg argues that the assumption that underlies Albert's claim — that the micro-states that lead to SLT violating behavior are scattered in tiny clusters all over phase space (Albert 2000, 67) — is supported by "neither a priori reasons nor plausibility arguments" (Frigg 2011, 87). But, even if the assumption is correct, different choices of regularly-shaped sub-region and smooth probability distribution are liable to yield *some* small but non-zero differences in probabilities for thermodynamic behavior. Moreover, from the perspective of providing a best axiom system for the universe, choices of irregularly shaped regions and non-smooth distributions presumably shouldn't be ruled out *a priori*. In particular, the complexity involved in specifying such irregularly shaped regions and non-smooth distributions may be offset by a compensating increase in fit.

on the Lebesgue measure. Maudlin (2007; 2011) has argued that, for systems comprising a large but finite number of molecules, normal thermodynamic behavior is 'typical' in the sense of being produced by a set of initial conditions that has Lebesgue measure $1-\epsilon$ for small but positive ϵ (Maudlin 2007, 289). Yet there are alternative measures that assign different sizes to the set of initial conditions that lead to abnormal thermodynamic behavior (Maudlin 2007, 290). Probability distributions that are uniform with respect to these alternative measures are thus liable to yield different SM probabilities from the Mentaculus.

Of course, there must be a certain amount of qualitative agreement between systems that are serious contenders for best systemhood: Each must entail that the probabilities of entropy increase in (at least most) appropriately isolated non-equilibrium systems are 'very high', since any that didn't wouldn't be well-fitting enough to be among the best. Still, the different probability distributions invoked by different systems, and the different sized regions of phase space to which the probability distributions are applied, are liable to translate into non-zero differences in the probabilities assigned to entropy-increasing behavior (cf. Callender 2011, 107). Where more complex systems are better fitting, it appears likely that different reasonable standards of simplicity and balance will yield different verdicts about which system is best. In other words, it appears likely that no system will count as 'robustly' best. As shorthand in what follows, I'll sometimes talk about such a situation as one in which we have a 'tie' between systems, though of course this shouldn't be construed as meaning that we have uniquely reasonable standards that score two or more systems identically, but rather that we don't have unique reasonable standards and that different systems come out best according to different reasonable standards.

As we saw earlier, Lewis (1994, 479, 481) claims that, if there's no robustly best system, then there's nothing very well deserving of the name *law* or *chance* (cf. Beisbart 2014, 521). But we've now seen that, very plausibly, there *is* no robustly best system for our world. Must we then draw the (surprising) conclusion that there are no laws or chances?

This seems too extreme. Lewis's earlier position that, in case of ties — here interpreted as the absence of a *robustly* best system —, those theorems entailed by *all* of the tied-for-best systems would count as laws (Lewis 1983, 367) is more plausible. So too is the claim that any probabilities that the tied-for-best systems agree upon count as chances. Thus, for example, those who have considered axiomatizations of SM that are rivals to the Mentaculus have not necessarily disputed that the *fundamental dynamical principles* are the same in those axiomatizations as in the Mentaculus. If all tied-for-best axiom systems entail the same fundamental dynamical principles, then these principles should be taken to be genuine laws. Moreover, if the tied-for-best systems entail the same fundamental dynamics, then they presumably entail the same probabilities conditional upon microphysical chance setups. For example, it's possible that all tied-for-best systems agree on the probability that a particular tritium atom will decay within the next 12.32 years. If so, such probabilities should be taken to be genuine chances by the Best System analyst.

While I agree that, in case of ties, the theorems and probabilities that all tied-for-best systems agree upon count as laws and chances, I'll argue that what chances there are *goes beyond* the probabilities that all tied-for-best systems agree upon. Specifically, I'll argue that, for those phenomena where divergent probabilities are entailed by the tied-for-best systems — as I've argued is the case for thermodynamic behavior — the chances correspond to the *sets* of probabilities entailed by these tied-for-best systems. That is, the chances are *imprecise*. Where all tied-for-best systems agree on a probability — as they might for tritium decay — then the set is a singleton and the chance precise. But not so where they disagree.

What about laws? Though I'll have less to say about laws than about chances in what follows, I think that in the case where the generalizations entailed by the tied-for-best systems differ only in the probabilities that they assign to certain behavior (and not, for instance, in what factors they take to be relevant to that behavior) — as is plausibly the case when it comes to the the probabilistic principles of thermodynamics (such as SLT, the Ideal Gas Law, etc.) that are entailed by the seemingly tied-for-best systems for our world — the best system analyst should take the laws themselves to be imprecise in the sense of entailing imprecise probabilities for the behavior they concern, where of course the imprecise probabilities correspond to the sets of probabilities entailed by the tied-for-best systems.[13]

5. Localist Approaches to SM

Callender (2011) calls attempts to axiomatize SM that appeal to axioms giving a probability distribution (such as that given by SP) over a region of the phase space of *the universe as a whole* that contains its PICs (such as that described by PH) 'Globalist' approaches. The Mentaculus is an example of a Globalist approach. As we've seen, there's a range of other Globalist approaches that vary the region of the universe's phase space that a probability distribution is applied to and/or the probability distribution applied to it. Many of these approaches are plausibly just as good systematizations as the Mentaculus.

In contrast to Globalist approaches, 'Localist' approaches (Callender 2011; Frigg and Hoefer 2015) propose that SM be axiomatized instead by means of probability distributions applied to regions of the phase spaces of the various *approximately isolated subsystems of the universe* that contain the actual initial conditions of those systems (Callender 2011, 96–7). Without going into too much detail, an important motivation for those who have advocated the Localist approach is a skepticism about the well-fittingness of Globalist axiomatizations combined

[13] A slightly different approach would be to take the view that the laws are limited to the generalizations that all of the tied-for-best systems agree on but to note that, while the tied-for-best systems do not all agree on thermodynamic generalizations entailing precise probability distributions, they do agree on generalizations that say that the probability distribution lies in a certain *set*: namely any set whose elements include the probability distributions entailed by each of the tied-for-best systems. The most informative such generalizations would be the ones that say that the probability distribution lies in the set whose *only* elements are the probability distributions entailed by each of the tied-for-best systems. On this approach, it would be natural to say that the chances are an element of the set, but that there's no fact of the matter about which element, rather than that the chances are set-valued. Ultimately, it seems quite plausible that this approach and the one described in the main text are mere notational variants on one another. What's of interest is that they're quite different from the standard view that there's a precise probability distribution that gives the chances for our world, and that there's a fact of the matter about which distribution this is.

with arguments that Localist approaches are much better fitting.[14]

In this section I'll argue that, as with candidate Globalist axiomatizations, it's plausible that there's a *set* of Localist axiomatizations none of which is robustly better than any other. I'll thus argue that the best system analyst should also believe in imprecise chances if she regards Localist axiomatizations as more promising (i.e. as plausibly providing better systems) than Globalist axiomatizations.

Frigg and Hoefer (2015) endorse a Localist approach to SM, and combine this with a Best System-style analysis of chance, which they call a 'Theory of Humean Objective Chance', or 'THOC'. They argue that the distribution, P_U^μ, that's uniform on the standard Lebesgue measure, μ, should be applied to that region of each (approximately) isolated thermodynamic system's phase space corresponding to its initial macrostate in order to generate the SM chances. They argue that the distribution P_U^μ can be justified in terms of simplicity and fit with the actual frequencies. Specifically, they take Γ_p to denote the phase space region corresponding to such a system's initial macro-state (that is, its macro-state at some initial time t_0) and they ask us to consider all of the systems in the entire Humean mosaic that share this initial macrostate, saying:

> Each of these [systems] has a precise initial condition x at t_0, which, by assumption, lies within Γ_p. Now go through the entire [Humean mosaic] and put every single initial condition x into Γ_p. The result of this is a swarm of points in Γ_p. . . . THOC is essentially a refinement of finite frequentism and chances should closely track relative frequencies wherever such frequencies are available. . . . But . . . we have to reduce the complexity of the system by giving a simple summary of the distribution of points. To this end we approximate the swarm of points with a continuous distribution (which can be done using one of the well-known fitting techniques . . .) and normalise it. The result of this is a probability density function ρ on Γ_p, which can be regarded as an expression of the 'initial conditional density' in different subsets C of Γ_p.
>
> The good-fit constraint now is that $\rho(C)$ be equal to (or in very close agreement with) $\mu(C)/\mu(\Gamma_p)$ for all subsets C of Γ_p. This is a non-trivial constraint. For it to be true it has to be the case that the initial conditions are more or less evenly distributed over Γ_p because $\mu(C)/\mu(\Gamma_p)$ is a flat distribution over Γ_p. (Frigg and Hoefer 2015, 562–3)

Frigg and Hoefer's justification for applying the uniform distribution to regions of the phase spaces of subsystems of the universe (namely those regions corresponding to their initial macrostates) is thus somewhat different from Albert and Loewer's justification for applying it to a region of the phase space of the universe as a whole (namely, that region picked out by PH). Their idea is that the uniform distribution is "an elegant summary of *actual* initial conditions as they occur in the [Humean Mosaic] of a world like ours" (Frigg and Hoefer 2015, 571). As they point out, this justification of the uniform distribution "is not open to those who take [Boltzmannian SM] to be a theory about the universe as a whole, since there is only exactly one initial condition of the universe" (Frigg and Hoefer 2015, 571).

One worry about this Localist approach is that there may simply not be enough systems that have exactly the same number of phase space dimensions (and that share the same initial macrostate) to produce a set of actual initial conditions that's large enough to single out the uniform distribution as that which supplies the robustly best balance of fit and simplicity. Frigg and Hoefer (2015, 563n) acknowledge this worry in a footnote, saying:

> [O]ur discussion idealises by pretending that the histories of all sorts of different SM systems could be treated as representable via paths in a single phase space. This is an idealisation because systems with a different particle number N have different phase

[14] For details, see Callender (2011, 100–1, 106) and Frigg and Hoefer (2015).

> spaces. We think that this is no threat to our approach. SM systems such as expanding gases and cooling solids are ubiquitous in [the Humean Mosaic] and there will be enough of them for most N to ground a [Humean Best System] supervenience claim. Those for which this is not the case (probably ones with very large N) can be treated along the lines of rare gambling devices such as dodecahedra: they will be seen as falling into the same class as the more common systems and a flat distribution over possible initial conditions will be the best distribution in much the same way in which the $1/n$ rule [where n is the number of sides] is the best for all gambling devices.

But it's not clear whether Frigg and Hoefer are justified in being confident, even for systems of relatively low phase space dimensionality, that there will be *enough* such systems to single out P_U^μ as the distribution that strikes the robustly best balance of simplicity and fit with the actual distribution of initial conditions. Still less is it clear whether there are enough systems with the same low-dimensional phase spaces to guarantee that applying P_U^μ to phase spaces of higher dimensionality (of which there may be fewer instances) will be the strategy that strikes the robustly best balance between simplicity and fit.

If the set of actual initial conditions of such systems fails to nail down P_U^μ as that distribution which strikes the robustly best balance between simplicity and fit, then plausibly we will be left with a large (perhaps continuous) range of distributions that fit actual initial condition frequencies reasonably well, with the better fitting (e.g. certain non-flat distributions that concentrate probability in regions where there are clusters of actual initial conditions for the subsystems in question) being more complex, and the worse fitting (e.g. P_U^μ) being more simple.

Analogously to the Globalist approach, we can also consider a range of Localist approaches to SM that apply one or other probability distribution to variously sized sub-regions of the phase space of the appropriate subsystems of the universe (where these subregions contain the precise initial conditions of each of the systems that has a phase space of that dimensionality).[15] Again, picking out smaller sub-regions will typically buy better fit (by increasing the probability of each subsystem's actual macro-evolution) but involves more complexity (because it involves specifying the initial macro-states of the subsystems more precisely, or adding in some information about the subsystems' initial micro-states). The upshot is liable to be a range of tied-for-best Localist axiomatizations of SM entailing different SM probabilities and consequently, even if Localist axiomatizations are superior to Globalist axiomatizations, the Best System analyst should (I'll argue) still conclude that there exist imprecise SM chances, where this time the imprecise SM chances are constituted by the sets of probabilities entailed for given outcomes by the tied-for-best *Localist* axiomatizations of SM. Insofar as the Localist also wishes to recover probabilistic approximations of thermodynamic laws, perhaps by taking them to be generalizations over appropriately isolated thermodynamic systems, then this also gives us reason to think that such probabilistic laws are imprecise.

6. Imprecise Chances & Chance-Credence Coordination

The basic reason for thinking that there are imprecise chances where there's a tie between systems entailing different probability distributions is that, in such a case, players of the chance role in guiding rational credence are not confined to those probabilities that each of the tied-for-best systems agree upon. Rather, in cases of divergence, the (non-singleton) sets of probabilities entailed by the tied-for-best systems appear to play this role.[16] In order to make the case that this is

[15] Note that there's no reason to rule out non-convex sub-regions *a priori*. Suitable choices of (possibly non-convex) regions of phase space that contain the actual initial conditions of systems that share that phase space may (at the cost of some simplicity) allow us to ensure a higher probability for the actual macro-evolution of such systems.

[16] A referee pointed out that another drawback of taking the chances to be confined to those probabilities on which the tied-for-best systems all agree is that this may yield an oddly 'gappy' chance function: For instance, if all such systems agree on the probability of the conjunction $A\&B$ but disagree on the

probabilities of A and B, then such a proposal implies that the domain of the chance function isn't a Boolean algebra. But the same referee notes that there may be a worry about *my* proposal if the tied-for-best systems aren't defined on the same domain. For instance, what do we say about the chance of A if some but not all of the tied-for-best systems assign a probability to A?

In addressing this worry, first note that there's no obvious reason for thinking that the systems that, in this essay, have been envisioned to be among the tied-for-best are defined on different domains. It has been argued that there are liable to be systems that assign different probabilities to thermodynamic behavior among the tied-for-best, but nothing has been said to suggest that the tied systems entail probability distributions defined on different *kinds* of behavior or for different *kinds* of system. (Even if there were a tie between certain Globalist and Localist systems, then because Globalist systems have to entail probabilities for *subsystems* of the universe in order to be well-fitting, and because Localist systems may well entail probabilities for the thermodynamic evolution of the universe *as a whole* — if Frigg and Hoefer are correct that such axiomatizations can entail probabilities for systems comprising very large numbers of particles — it still wouldn't be unreasonable to suppose that the tied systems are defined on the same domain.)

But, by the same token, we have given no argument that among the tied systems there *won't* be systems defined on differing algebras. (Note that Lewis doesn't take definition on exactly the same algebra to be a prerequisite for the goodness of two systems to be compared: He says that a system is strong to the extent that it says "what will happen or what the chances will be when situations of a certain kind arise" (Lewis 1994, 480), which suggests that not all systems up for comparison need entail probabilities for the same kinds of situation.) This might be so if, for instance, there are systems among the tied-for-best that disagree on the fundamental dynamics.

Still, I think that we can reasonably expect significant (if not perfect) overlap in the domains on which the tied systems are defined. It is not very plausible that a system can be among the tied-for-best if it doesn't entail decay probabilities for the various standardly-recognised radioisotopes, for instance, or indeed for the macroscopic behavior of ordinary thermodynamic systems. (If, by contrast, *very* different systems — presumably deploying sets of predicates that don't overlap much — were among the tied-for-best, then perhaps there really would be nothing well deserving of the names of 'chance' and 'law', as Lewis (1994, 479) suggests. An exception would be if the disparate systems can simply be taken as characterizing different branches of science, in which case an approach such as the BBSA might be called for. But, as noted in Footnote 6, it's reasonable to take the BBSA as also implying the existence of imprecise chances.) So I think that the following conservative proposal is perfectly defensible: The domain of the chances is the Boolean closure of the maximal set of atomic propositions upon which each of the tied systems is defined and the chance associated with each element of that domain is the set of probabilities associated with it by the tied systems. (When, in what follows, I speak (for simplicity) of the set of probability functions associated with the tied-for-best systems as constitut-

so, it will be necessary to say how imprecise chances — if they exist — would guide rational credence. I attempt to do so in this section. Then, in Section 7, I argue that the set of probability functions associated with the tied-for-best systems guide rational credence in this way. In this section, I first say something about rational credence (Subsection 6.2), before turning to imprecise chance (Subsection 6.3), and then the role of the latter in guiding the former (Subsection 6.4). Before doing this, however, I'll briefly mention some previous arguments that have been given for the existence of imprecise chances (Subsection 6.1).

6.1 Existing Arguments for Imprecise Chances

The use of imprecise probability (IP) models of chancy phenomena has been investigated by Walley and Fine (1982), Kumar and Fine (1985), Grize and Fine (1987), Fine (1988), and Fierens et al. (2009), among others.[17] These authors are interested in the use of imprecise probabilities to model physical processes that give rise to frequencies that vary over time, particularly where the frequency is always within some interval.

Consider the following heuristic example, based on one given by Walley and Fine (1982, 759) (cf. Fierens et al. 2009, 1880–3). Suppose that I purchase a brand new die and proceed to roll it a very large number of times. The corners of the die start off sharply angled, but they become more worn and rounded after a large number of rolls.

ing chance, I should therefore, strictly speaking, be interpreted as meaning that it is the set of the *restrictions* (or perhaps *truncations*) of such functions to the domain of chances — with this domain being the one specified in the previous sentence — that constitute the chance for the world.) This is weaker than (and preferable to) the restriction of the chance domain to those propositions that all the tied systems agree upon, and doesn't issue in gaps such as those described by the referee. It might be possible to argue that the domain of the chance function extends further to propositions upon which only *some* of the tied systems are defined. That suggestion is liable to re-introduce such gaps (for instance: What if some systems give probabilities for A, some for B, but none for $A\&B$?) and in general is liable to be more controversial. In any case, an exploration of this more radical approach would take us too far afield.

[17] Fierens et al. (2009, 1879) use the term 'chaotic' rather than 'imprecise' probability.

Suppose the die is more likely to land face down on a face the corners of which are relatively unrounded, and yet repeated landing with a given face down tends to wear and round the corners of that face. Then it is to be expected that the frequency with which the die lands on each of its faces will vary slightly over time within some small interval around $\frac{1}{6}$ as chancy runs of lands on one face wears its corners, slightly reducing the likelihood of lands on that face for a while until the patterns of wear are evened up.

Kumar and Fine (1985, 1–2), Fine (1988), and Grize and Fine (1987, 785) observe that 'flicker noise' — a form of non-deterministic fluctuation found, for instance, in electric signals — involves such variable frequencies, while Fierens et al. (2009, 1880) claim that frequencies in weather and financial data may behave similarly. These behaviors resist modelling with precise probabilities (Kumar and Fine 1985), but can be captured using IP models. Walley and Fine (1982, 741–3) and Fierens et al. (2009, 1879) (cf. Fine 1988, 392–3) variously describe their interpretations of these models as "*objective, frequentist*" (Walley and Fine 1982, 742; cf. Fierens et al. 2009, 1879), as "a representation of propensities" (Walley and Fine 1982, 742; cf. Fine 1988, 392–3) of outcomes, and as 'ontological' probabilities or 'chances' (Walley and Fine 1982, 743), with the imprecision representing "ontological indeterminacy" rather than mere "incomplete knowledge" or "epistemological indeterminacy" (Walley and Fine 1982, 758). They clearly distinguish this interpretation from "subjectivist and epistemic" approaches (Walley and Fine 1982, 742) or a "behavioral subjective" interpretation (Fierens et al. 2009, 1879).[18]

Likewise, some authors have appealed to imprecise probability models of quantum phenomena (Suppes and Zanotti 1991; Hartmann and Suppes 2010; Galvan 2008; Hartmann 2015; Allahverdyan 2015[19]), such as entangled systems of particles that lack a classical probability distribution. It's clear that at least some of these authors wish to interpret the IP models realistically. For instance, Allahverdyan (2015, 2) claims that "[i]mprecise joint probabilities in quantum mechanics are to be regarded as fundamental entities, not reducible to a lack of knowledge."

The arguments of the authors cited above, who advocate IP models for flicker noise or QM phenomena, do not appeal to the BSA nor to the imprecision of the associated notions of simplicity, strength, and balance. A natural moral for the defender of the BSA to draw from these arguments might be that an adequate BSA should allow that *individual candidate systems themselves* needn't be associated with a unique probability distribution (i.e. the BSA should allow for intra-system imprecision): Good candidate systems should (if these authors are correct) entail sets of probabilities for phenomena such as flicker noise and perhaps entangled states of particles.

My arguments, which focus principally on SM, aren't based on any claim that SM processes yield unstable frequencies (as is the case for flicker noise) or that they lack a classical probability distribution (as is the case for entangled systems of particles). Rather the claim is that, even if they do yield stable frequencies, systems that entail probabilities exactly equal to those frequencies might not be the (robustly) best, because such systems might be exceedingly complex. For instance, the best-fitting systems might entail different probabilities for a huge range of types of thermodynamic system reflecting the frequencies, among

[18] One might object that it's implausible to interpret such processes as subject to objective imprecise probabilities. Surely — one might think — the processes are simply subject to changing precise probabilities (for instance, I failed to characterise the die example in the main text without speaking in terms of the changing 'likelihood' of its landing on a given side), which we might struggle to empirically estimate. However, Fierens et al. (2009, 1883) argue that there may be "no empirical reality existing" that's sufficient to ground these changing precise probabilities (in their terms, there may be no empirical reality that is sufficient to ground a 'selection function' F over the set $\mathcal{M}$ of probability functions used to model the process, where F selects a unique probability function at each instant of time — perhaps depending on the precise state of the system at that time). Put in Humean terms: For at least some physical processes, the mosaic just isn't rich enough to ground ever-evolving precise probabilities, but only rich enough to determine a non-singleton set of probability functions associated with the process.

[19] Fine (1974) proposes a *comparative probability* model of quantum phenomena.

systems of those types, with which their initial conditions are located within regions of phase space that issue in given sorts of thermodynamic behavior. Such systems are much more complex than those that simply treat all types of thermodynamic system as subject to a uniform distribution over those regions of their phase space compatible with their initial macrostate. (They might also strike us as too sensitive to 'accidental' facts about the frequencies with which initial conditions of systems of certain sorts are located within regions that issue in certain sorts of thermodynamic behavior.)

The argument of this essay is that plausibly there's a *set* of systems entailing (slightly) different probabilities for SM phenomena that are tied for best and hence the Best System analyst should take the *set* of probability distributions associated with the set of tied-for-best systems to constitute the (imprecise) chance distribution for the world. That is, my claim is that there's imprecision for inter-system reasons (ties between systems) and not just for intra-system reasons (the fact that a given system might be more adequate if it itself entails sets of probabilities for certain phenomena).[20] However, it's worth noting that, to the extent that the approaches cited above suggest that the best system(s) may themselves entail imprecise probabilities, the following discussion (subsections 6.2–6.4) of the imprecise chance-credence relation should be of interest even to those who are skeptical that there's a tie for best systemhood or those who believe that a tie *per se* doesn't commit the Best System theorist to the existence of imprecise chances. Moreover, because the above-cited approaches suggest that the existence of imprecise chances for our world must be permitted by *any* satisfactory metaphysical theory of chance, the following discussion of the imprecise chance-credence relation should be of interest even to those who are skeptical of the BSA.

I'm not the first to suggest that ties between systems might lead the defender of the BSA to endorse imprecise chances. This possibility has been noted in passing by Hájek (2003b). After discussing the possibility that credences may be imprecise — or as Hájek puts it 'vague'[21] —

[20] Although I won't develop this suggestion in detail, if we have ties between systems $s_1, \ldots s_n$, each of which itself posits a *set* of probability distributions $\mathcal{P}_i(\mathbf{F})$ over elements of an algebra $\mathbf{F}$ generated on a sample space Ω, then a natural suggestion is that the overall imprecise chance distribution over $\mathbf{F}$ is given by the set $\{P(\mathbf{F}) : P(\mathbf{F}) \in \mathcal{P}_i(\mathbf{F})\}$ (where $P(\mathbf{F})$ denotes an individual probability distribution over $\mathbf{F}$). Alternatively, we might regard the chances as represented by the set of sets of probability distributions (cf. Bradley 2016, Section 3.5), $\{\mathcal{P}_1(\mathbf{F}), \ldots, \mathcal{P}_n(\mathbf{F})\}$. Interestingly, the latter is the sort of entity that Kozine and Utkin (2002), Troffaes (2006), and others take as a 'second order' sample space in developing imprecise hierarchical uncertainty models (the first order sample space is just Ω). (Strictly speaking, Kozine and Utkin (2002) take a second order sample space to be a set of coherent interval-valued previsions and Troffaes (2006) takes it to be a set of coherent lower previsions for gambles on Ω. As I indicate in the main text below, I don't assume imprecise chances to be convex/interval-valued and so determined by upper and lower values.) As they interpret them, the elements of these second order sample spaces are imprecise assignments over Ω that are made by individual experts. We are instead currently interpreting them as imprecise assignments associated with individual best systems (which can be regarded as 'experts' regarding outcomes in the Humean mosaic). In imprecise hierarchical models, a(n imprecise) distribution is then placed on this second order sample space (Kozine and Utkin 2002; Troffaes 2006). Though I won't develop this idea here, the weightings implied by such a distribution could be used to model degree-of-membership of the first-order (sets of) distributions associated with the different systems in a second-order (fuzzy) set representing chance if it's a *vague* matter which systems are among the tied-for-best (a possibility raised in Footnote 21 below).

[21] I don't think that 'vague probability' is the ideal terminology for the phenomenon under consideration. 'Vague probability' is best reserved for the case where it's a vague matter which probability distributions belong to the set that constitutes an imprecise probability distribution (cf. Joyce 2005, 167n; Sturgeon 2008, 158–9; 2010, 128–33; Bradley 2016, Section 1.2). On the present interpretation, *chances* might be vague in addition to being imprecise if there's not only a tie between systems, but it's also a vague matter which systems are tied-for-best (perhaps there are some systems that are determinately among the tied-for-best, some systems that are determinately not among the tied-for-best, and some systems that don't determinately belong to either category). This might be so if what constitute reasonable standards of simplicity, strength, fit, and/or balance is itself a vague matter. I shan't explore this possibility further here. In any case, adopting the terminology recommended here, it's possible to have an imprecise probability that isn't vague — i.e. where a probability model comprises a set of distributions, but it's a perfectly determinate matter which elements belong to that set. Hájek (personal communication) has confirmed that what he means by 'vague' probability in the passages I quote in the main text corresponds to what I'm calling 'imprecise' probability (and also that he doesn't regard 'vague' as the

Hájek briefly turns his attention to objective probability:

> More controversially, let me suggest that we remain open to the possibility of vague *objective probabilities*. For example, perhaps the laws of nature could be vague, and if these laws are indeterministic, then objective probabilities could inherit this vagueness. ... [A] chance that is vague over the set *S* corresponds to a set of sharp chances, taking on each of the values in *S*. (Hájek 2003b, 278)

In a footnote, Hájek (2003b, 278n) observes that:

> This would certainly seem to be a live possibility on a Mill-Ramsey-Lewis style account of laws as regularities that appear as theorems in a 'best' theory of the universe, as long as the criteria for what makes one theory better than another are themselves vague. (In Lewis' . . . theory, for instance, the vagueness may enter in the standards for balancing the theoretical virtues of 'simplicity' and 'strength'.) Then nature may not determine a single best theory, but rather a multiplicity of such theories. Suppose, for example, that these equal-best theories disagree on the chance that a radium atom decays in 1500 years: for each real number *r* in the interval [1/3, 2/3], there is such a theory that says that the chance is *r*. Then we might say that the chance of this event is vague over this interval.[22]

The possibility suggested in this latter quote is the one I wish to push. Though Hájek uses QM as his example, I've been arguing that competing axiomatizations that entail SM probabilities provide a strong motivation for thinking that there's indeed a tie and hence — on a plausible construal of the BSA — that there are imprecise Best System chances *in our world* and that the probabilistic laws of statistical thermodynamics can correspondingly be taken to be imprecise. Hájek's suggestion that, given the imprecision of the standards for balancing the theoretical virtues (as well, I would add, as the virtues themselves) there may be systems among the tied-for-best that incorporate differing QM probabilities, might be plausible. However, this isn't a thesis I'll argue for here.

In the remainder of this section and in the next, I'll seek to bolster my argument that the best system analyst should endorse the existence of imprecise chances given that there's plausibly a tie for best systemhood. I'll do this by describing (in the remainder of this section) how imprecise chances, if they exist, would constrain rational credence and then arguing (in Section 7) that the set of probability distributions entailed by the tied-for-best systems constrain rational credence in this way.

6.2 Imprecise Credence

The thesis that *reasonable credence* may be imprecise has gained some popularity in formal epistemology.[23] Joyce (2005, 156) claims that "[t]he idea that people have sharp degrees of belief is . . . epistemologically calamitous". Joyce (2005, 170) argues in particular that "[i]t is wrong-headed to try to capture states of ambiguous or incomplete evidence using a *single* credence function". An agent doing so "is pretending to have information he does not possess" and ignoring "a vast number of possibilities that are consistent with his evidence" (Joyce 2005, 170). Instead, Joyce (2005, 171) argues that:

> As sophisticated Bayesians . . . have long recognized . . . [i]ndefiniteness in the evidence is reflected not in the values of any single credence function, but in the spread of val-

ideal term for this phenomenon).

[22] Compare Hájek and Smithson (2012, 39).

[23] Advocates include Levi (e.g. 1974; 1980; 1985; 1999), Kaplan (1983), Fine (e.g. 1988), Walley (1991), Kyburg Jr. and Pittarelli (1996), Joyce (2005; 2010), Bradley (2009), Bradley (2012), and Konek (2019).

> ues across the *family* of all credence functions that the evidence does not exclude. This is why modern Bayesians represent credal states using sets of credence functions. ...Imprecise credences have a clear epistemological motivation: they are the proper response to unspecific evidence.

As Joyce indicates in the above quotations, a common formal representation of imprecise credence appeals to the idea that the credences of a rational agent, S, are best represented, not by a single probability function, but rather by a *set* of probability functions, $\boldsymbol{cr}$. This set is the person's 'representor' (van Fraassen 1985, 249).[24] Since I take conditional probability as basic (cf. Hájek 2003a; 2003b; 2007), I assume that each element of $\boldsymbol{cr}$ is a conditional probability function $cr_i(\cdot|\cdot)$ that maps ordered pairs of propositions $\langle X, Y\rangle$ onto a unique real number, $x(\in \mathbb{R})$, in the $[0,1]$ interval: $cr_i(X|Y) = x$ (cf. Joyce 2005, 156). (Unconditional probabilities can be defined via $cr_i(X) =_{def} cr_i(X|\top)$, where $\top$ is the tautology — see Hájek 2003b, 315; Bradley 2016, Formal Appendix.) With slight abuse of notation, we can let $\boldsymbol{cr}(\cdot|\cdot)$ represent the function that maps ordered pairs of propositions $\langle X, Y\rangle$ to the set of values $\boldsymbol{x}$ that the probability functions in $\boldsymbol{cr}$ give for X conditional upon Y (where such probabilities are defined): that is $\boldsymbol{cr}(X|Y) = \{cr_i(X|Y) : cr_i \in \boldsymbol{cr}\}$ (cf. Bradley ms; Carr 2015, 69n). We can also stipulate that $\boldsymbol{cr}(\cdot) =_{def} \boldsymbol{cr}(\cdot|\top)$.

Those who advocate representing a rational agent S's credal state in terms of a set of probability functions typically model updating upon new evidence D by supposing that *each* probability function in S's representor is updated upon D and that such updates involve conditionalization upon D (see Joyce 2005, 153, 172; 2010, 287, 292–3; Chandler 2014, 1277; Bradley 2016, Section 1.1; cf. White 2010, 173; Sturgeon 2008, 157, 157n; 2010, 131, 137). Where S's initial representor, $\boldsymbol{cr}$, is assumed to be a set of conditional probability functions, the idea is that if S learns D (and nothing else), then her new representor is $\boldsymbol{cr}^D$, where:

$$\boldsymbol{cr}^D = \{cr_i^D(\cdot|\cdot) = cr_i(\cdot|\cdot \& D) : cr_i(\cdot|\cdot) \in \boldsymbol{cr}\} \qquad (*)$$

(Any probability function $cr_i \in \boldsymbol{cr}$ such that probabilities of the form $cr_i(\cdot|\cdot \& D)$ aren't defined is 'weeded out' rather than updated in the transition to $\boldsymbol{cr}^D$.)

To illustrate, suppose that S gains imprecise evidence, D, concerning some proposition A. For example, suppose that $A =$ 'a red ball will be drawn from urn 1', and that $D =$ 'all balls in urn 1 are either red or blue and a ball will be drawn from urn 1 at random' (D is imprecise evidence because it doesn't say what the proportions of red and blue balls in the urn are). If S has no other evidence bearing upon A, then — for reasons to be discussed below — imprecise probabilists will typically argue that her post-update credence in A (conditional upon $\top$) should be spread out over the whole of the $[0,1]$ interval.

Modeling S's credences by a representor $\boldsymbol{cr}$ that's updated according to $(*)$ can accommodate this. The idea is that various elements $cr_i \in \boldsymbol{cr}$ 'interpret' the evidence D differently. That is, the value of $cr_i(A|D)$ is different for various $cr_i \in \boldsymbol{cr}$. Consequently, when S updates upon D in accordance with $(*)$, various elements $cr_i^D \in \boldsymbol{cr}^D$ yield different values $cr_i^D(A)$ $(= cr_i^D(A|\top))$. More precisely, there's a set $\boldsymbol{cr'} \subseteq \boldsymbol{cr}$ such that, for each pair of elements $cr_i, cr_j \in \boldsymbol{cr'}$ $(i \neq j)$, $cr_i(A|D) \neq cr_j(A|D)$. This means that the post-update probabilities, $cr_i^D(A)$ and $cr_j^D(A)$, diverge. The idea is that, for a rational agent, the probability functions in her initial representor are such that the various values for $cr_i(A|D)$ are spread out all over the $[0,1]$ interval, so that the various values of $cr_i^D(A)$ are, too.

No completely compelling argument has been produced to show

[24] There are many other formal models of imprecise credence, and several of the authors I discuss below make use of alternative models. The advantages of the representor model over other representations are described by Joyce (2010, 285–7, 294–6) and Bradley (2016). These advantages seem decisive to me.

that (*) is the correct way of modeling belief change for a rational agent whose credal state is represented by a set of probability functions. Where the credal state of a rational agent is modeled by a single probability function, conditionalization is normally taken to be justified by diachronic Dutch Book arguments (Teller 1973; Lewis 1999), or by considerations of 'symmetry' or 'representation independence' (Hughes and van Fraassen 1984; cf. Grove and Halpern 1998). Justification of the principle that, where an agent's belief state is represented by a *set* of probability functions, rationality requires that she update by conditionalizing *each* probability function in her representor (for which conditionalization is defined) upon the evidence is less straightforward, and attempts to justify it (see Grove and Halpern 1998) are less compelling.[25] For now, however, I'll assume that (*) *is* the correct way to model rational belief change for a (rational) agent whose credal state is modeled by a representor. I'll consider an heterodox model of belief change for such an agent in Subsection 6.4 below.

Among those who have defended the view that imprecise credence is the rational response to ignorance or unspecific evidence are Levi (1985), Fine (1988), Walley (1991), Joyce (2005), and Sturgeon (2008; 2010).[26] Sturgeon treats this as a consequence of what he calls the

[25] Walley (1991, 335–6) discusses an update model for imprecise probabilities modeled using lower previsions in which an agent's unconditional lower previsions after learning D are equal to her prior lower previsions conditional upon D. However, he notes that "there is scope for other updating strategies" (Walley 1991, 336). For instance, he suggests that it might be reasonable for your post-update lower previsions to imply a more precise attitude than your prior lower previsions conditional upon D if, for instance, your prior conditional lower previsions correspond to a highly imprecise attitude reflecting the fact that, prior to learning D, D was "unexpected and You have spent little time assessing prior previsions related to" it (Walley 1991, 336). He also expresses the concern (equally applicable to the standard model of updating precise credal states) that some evidence that one might gain can't be represented in the possibility space on which one's prior previsions were defined, so 'prior' conditional previsions can at best be assessed *after* the fact relative to the more fine-grained (post-update) possibility space (Walley 1991, 337–8, 340n).

[26] Not everyone is convinced. Elga (2010) and White (2010) argue that imprecise credences are irrational. Elga (2010) argues that any plausible decision rule for imprecise credences sanctions irrational sequences of actions in certain diachronic decision problems. Yet Sahlin and Weirich (2014), Chandler (2014, 1280), and Bradley and Steele (2014, esp. 284–7) have all argued that Elga's argument turns upon spuriously supposing that a rational agent will fail to engage in valid 'backwards induction' reasoning that's clearly called for when the decision problem is presented in its extensive form (see Seidenfeld 1994, esp. 459n). (Thanks to an anonymous referee for bringing this to my attention.) Sahlin and Weirich (2014), Chandler (2014, 1282–4), and Bradley and Steele (2014, esp. 285–6) also show that an agent who adopts one plausible decision rule for imprecise probabilities, namely Γ-Maximin (a rule derived from Gärdenfors and Sahlin 1982) — which says that an act A is permissible iff the minimum expected utility associated with A by any element of the agent's representor is at least as great as the minimum expected utility of any alternative act B — and engages in the warranted backwards induction, will recognize the irrational sequence of decisions as inadmissible by the lights of her decision rule. Elga has subsequently responded to Sahlin and Weirich's (2014) article by conceding the point (see Elga 2012).

Bradley and Steele (2014, esp. 285–7) argue that another decision rule for imprecise credences — namely, what they call a 'non-dominated-set' (NDS) rule, which says that A is admissible if there's no alternative act B such that at least one element of the agent's representor assigns B a higher expected utility than A and no element assigns B a lower expected utility than A — also avoids irrational acts in Elga's scenario when the agent engages in the warranted backwards induction reasoning. Meanwhile, Joyce (2010, 315–6) argues that a subclass of what he calls 'pragmatic sharpening rules' (which involve basing your decisions on a subset of elements of your representor arrived at by 'throwing out' certain extreme probability functions) also avoid the irrational sequences of decisions in Elga's case.

White (2010) focuses upon the phenomenon of *probabilistic dilation* (cf. Seidenfeld and Wasserman 1993). An example of this phenomenon, considered by Pedersen and Wheeler (2014, 1326–7) — which is the same in all relevant respects as that which White (2010, 175–81) considers — is as follows. Suppose that an agent S, who has unspecific evidence for a proposition A, has imprecise credence in that proposition. In particular, suppose that A is the proposition that the next toss of Coin 1 will result in heads, where the bias of Coin 1 is unknown to S. Suppose also that S has highly specific evidence for proposition B, and has a precise credence in B on that basis. In particular, suppose that B is the proposition that the next toss of Coin 2 will result in heads, where Coin 2 is known by S to be fair and S has credence 0.5 in B on that basis. Then learning *either* $A \leftrightarrow B$ *or* its complement $\neg(A \leftrightarrow B)$ will force S, if she updates via (*), to have imprecise credence in B. White regards it as irrational that an agent should come to have imprecise credence in a proposition like B because she has imprecise credence in the unrelated proposition A and learns the seemingly innocuous proposition $A \leftrightarrow B$ (or $\neg(A \leftrightarrow B)$), especially as she seems required to have a precise 0.5 credence in B in accordance with Lewis's (1980) Principal Principle (PP).

Hart and Titelbaum (2015) counter — correctly — that the biconditional

'Character-Match Thesis' (CMT), which is the thesis that "the attitude taken to a claim should be fixed by the character of evidence on which it is based" (Sturgeon 2010, 133; cf. Joyce 2005) or, as he elsewhere puts it, that "epistemic perfection demands *character match* between evidence and attitude" (Sturgeon 2008, 160; italics original). The consequence is, according to Sturgeon, that:

> [E]vidence and attitude should match in character: when evidence is essentially sharp, it warrants a sharp attitude; when evidence is essentially fuzzy, it warrants a fuzzy attitude. When evidence is maximally precise it warrants ... real-valued subjective probability. When evidence is not maximally precise — as in most of the time — some other kind of attitude is called for, some kind of thick confidence. (Sturgeon 2010, 133)[27]

By 'thick confidence', Sturgeon means roughly what I mean by 'imprecise credence'.[28] ('Roughly' because he rejects existing formal mod-

is far from innocuous. Indeed it is 'inadmissible' evidence (cf. Lewis 1986, 92–6) concerning B in the sense that the PP fails to constrain rational credence for an agent in possession of it (see Joyce 2010, 299–307; cf. Hart and Titelbaum 2015, 259n). Even where an agent's probabilities for both A and B are precise, updating via orthodox Bayesian conditionalization on $A \leftrightarrow B$ (or its negation) can (depending on the exact values of her prior credences in A and B) have *prima facie* surprising effects on the agent's probabilities for A and/or B. The extent of dilation in the imprecise case is just an instance of the same phenomenon: It reflects the surprising (but normatively correct) result of conditionalizing precise probability distributions on biconditionals (after all, this is all that orthodox updating for a representor amounts to).

Joyce (2010, 299–307), Bradley (2016), and Hart and Titelbaum (2015) (cf. Seidenfeld and Wasserman 1993; Joyce 2005, 173) all thus regard dilation as a rational phenomenon. The key point (expressed in terms of our running example) is that $A \leftrightarrow B$ is strong (inadmissible) evidence for B (and $\neg(A \leftrightarrow B)$ is strong (inadmissible) evidence against B) according to some elements of S's representor: namely, those that assign high initial probability to A. On the other hand, $A \leftrightarrow B$ is strong (inadmissible) evidence against B (and $\neg(A \leftrightarrow B)$ is strong (inadmissible) evidence for B) for others: namely, those that assign low initial probability to A. (The biconditional is therefore 'epistemically relevant' to B *sensu* Pedersen and Wheeler (2014, 1325), and B is epistemically relevant to the biconditional, and hence the two are not 'completely stochastically independent' *sensu* Pedersen and Wheeler (2014, 1325) (cf. Joyce 2010, 300), even though the biconditional is epistemically irrelevant to A (cf. Couso et al. 2000, Section 3.1).) Hence the dilation that results from updating each of the elements of S's representor on $A \leftrightarrow B$ (or $\neg(A \leftrightarrow B)$), can't (non-question-beggingly) be claimed to be irrational. On the other hand, if S updates on some proposition that each element of her representor agrees is independent of B — specifically, if the two propositions are 'completely stochastically independent' *sensu* Pedersen and Wheeler (2014, 1307), as $A \leftrightarrow B$ would be if all elements of her representor agreed that Coin 1 as well as Coin 2 was fair (cf. Pedersen and Wheeler 2014, 1328) — then dilation won't occur. For more detail on this line of response, see Joyce (2010, 299–307). Pedersen and Wheeler (2014) pursue a slightly different line of response. They observe that there are representors that both encode imprecise beliefs about A and where updating upon the biconditional or its negation doesn't dilate B, and defend the rationality of an agent's having a credal state modeled by such a representor.

[27] Cf. Sturgeon (2008, 159) and Joyce (2005, 170–1). Wheeler (2014) argues that the CMT doesn't hold in full generality. However, his argument is that states of full belief/disbelief/suspension of judgment needn't match the character of one's evidence. For instance, imprecise and weak evidence can sometimes warrant full belief, while strong and precise evidence sometimes fails to warrant full belief. This is because full belief in φ "is the disposition to act as if φ were true relative to some specified range of actions" (Wheeler 2014, 191–2). This means that states of full belief are sensitive, not just to the character of evidence, but also to the range of possible actions available and the risk-to-reward ratio of acting as though φ (Wheeler 2014, 192). As Wheeler shows, because states of full belief are sensitive to these other factors, sometimes mismatches in character between states of fully belief and evidence will arise. But then again, Wheeler (2014, 191) recognizes that, for these reasons, there will sometimes be a mismatch in character between states of full belief and credal states. Though it would take us too far afield to discuss Wheeler's specific examples, I don't think any of them show (nor does he claim that they show) that there are situations in which we have imprecise evidence for a proposition but are warranted in having precise credence in it, or vice versa. So Wheeler's arguments don't tell against the view that there's a character match between evidence and rational credence, which is the aspect of character matching that interests us here. (Certainly, Wheeler (2014, 189–90) acknowledges that at least sometimes (imprecise) evidence *will* issue in reasonable imprecise credence — cf. Bradley 2016, Section 2.3.) Nor do they tell against *my* claim — to be developed in Subsection 6.4 below — that knowledge of imprecise chances of outcomes issues in reasonable imprecise credence about those outcomes.

[28] Sturgeon rejects the terminology that I adopt here because (a) he prefers to keep 'credence' as a technical term for real-valued subjective probability; and (b) he thinks that 'imprecise probability' suggests the probability is (what I've referred to as) vague or (as he calls it) 'fuzzy' (Sturgeon 2010, 130n, 132–3) when it needn't be. *I* use the term 'imprecise probability' because it's the most standard term for the phenomenon in question, and I've clearly distin-

els of the phenomenon.[29]) Thus, in the urn case described above, S has highly imprecise evidence concerning whether a red ball will be drawn from the urn. The CMT supports the claim that S's credence in this outcome should be correspondingly highly imprecise.

Imprecise chances might plausibly be taken as one sort of imprecise evidence one might have about the outcomes those chances concern. My claim — to be developed in Subsection 6.4 below — will be that the rational response to knowledge of imprecise chances of outcomes is imprecise credences in those outcomes.

White (2010, 173–4) — though himself denying the rationality of imprecise credences[30] — suggests that some of those who endorse imprecise credences may do so as a consequence of their endorsement of:

> *Chance Grounding Thesis (CGT)*: "Only on the basis of known chances can one legitimately have sharp credences. Otherwise one's spread of credence should cover the range of possible chance hypotheses left open by your evidence."

In the urn example given above, it seems that for each real number $x \in [0, 1]$, S's evidence leaves open the hypothesis that the chance that a red ball will be drawn from the urn is x. CGT thus suggests that S's credence in this outcome should cover the whole $[0, 1]$ interval.[31]

The CGT — which I'll return to in Section 6.4 — says that *only* when one knows the chances is it legitimate to have precise credences. I'll

guished this from 'vague probability' above. It's also now quite common to use 'credence' neutrally between the real-valued phenomenon and the imprecise phenomenon.

[29] His reservations are based on worries about dilation — which was addressed in Footnote 26.

[30] For reasons discussed in Footnote 26 above.

[31] A referee has questioned whether the CGT is widely endorsed among imprecise probabilists. And indeed Joyce (2010, 289) argues that it's too strong on the grounds that certain known symmetries can lead to precise credences even where chances are unknown. For instance, suppose fair Coin A will be tossed. If it lands heads, then Coin B with unknown bias β towards heads will be tossed. If it lands tails, then Coin C with unknown bias $1 - \beta$ towards heads will be tossed. You know all of this, and that each coin has zero chance of landing 'edge'. Then, before Coin A is tossed, for each probability function in your representor, the probability that the second coin to be tossed will land heads (H_2) is $P(H_2) = \frac{1}{2}\beta + \frac{1}{2}(1 - \beta) = \frac{1}{2}$.

Schoenfield (2017) describes a thesis which is somewhat weaker than CGT, and which she claims is part of "the standard imprecise view": namely, that "if the only evidence you have concerning whether P is that the objective chance function for $\{P, \sim P\}$ is in the set of probability functions S, then your evidence requires you to adopt the doxastic attitude represented by S." If Joyce's example is construed as one in which you have evidence that goes beyond what the chance for H_2 is — namely evidence that a certain symmetry holds — then it's not a counterexample to the thesis that Schoenfield describes.

However, I think that the symmetry information *is* a form of chance information. As Carr (2015, 70n) points out, in Joyce's case, "you do have chance information: in particular, information that before the coin [to be tossed second] was selected, the chances of heads [on the second toss] was .5". That seems exactly right to me. Let t_1 be a time immediately prior to the toss of Coin A, and let t_2 be a time after Coin A has been tossed, but before the second coin has been tossed. Since β is unknown, at t_2 you will be completely ignorant of the chance of heads on the second coin toss whichever coin has been selected to be the second tossed, and so your credence plausibly ought to cover the full [0,1] interval. Yet the fact that the t_2 chance (better: the chance conditional upon the history of the world up to t_2) of heads on the second toss is unknown at t_2 is compatible with the t_1 chance (better: the chance conditional upon the history of the world up to t_1) of heads on the second toss being known at t_1. And indeed this is the case: The unknown value of the parameter β, while being highly relevant to the t_2 chance, is irrelevant to the t_1 chance, which is given by $P(H_2) = \frac{1}{2}\beta + \frac{1}{2}(1 - \beta)$. This is equal to $\frac{1}{2}$, no matter what the value of β. (Learning the outcome of the first toss thus *dilates* one's credence concerning H_2.) In any case, there's no time at which (and no evidential state relative to which) rational credence concerning H_2 comes apart from the range of chance hypotheses concerning H_2 that's left open by your evidence. Joyce's example is therefore a counterexample neither to Schoenfield's principle nor to the CGT.

It should be noted that my thesis in what follows isn't that the CGT holds, but is rather the distinct claim that knowledge of imprecise chances warrants imprecise credences. Still, one of the arguments for the latter claim that I'll provide in Subsection 6.4 parallels an argument for the CGT that Carr (2015) advances. If the CGT turns out to be false, then something must be wrong with Carr's argument, and this is liable to cast doubt upon my parallel argument. I think that the CGT is true but, in any case, I provide arguments for the conclusion that knowledge of imprecise chances warrants imprecise credences that are independent of the one that parallels Carr's argument for the CGT.

argue that it's not *always* legitimate to have precise credences when one knows the chances. Specifically, it won't be legitimate if those known chances are themselves imprecise.

6.3 *Imprecise Chance*

That chance is probabilistic is something that has mostly been taken for granted. The authors cited in Subsection 6.1 are exceptions, since they take chance to be imprecise (and hence not real valued[32]). As, for instance, is Feynman (1987) who countenanced *negative* chances (and chances greater than 1).

Perhaps the most influential *argument* that chance is probabilistic is given by Lewis (1986, 98) who points out that this is a direct result of the assumption that rational credence is probabilistic — an assumption which has itself been defended by Dutch Book reasoning (Ramsey 1931, 182) and considerations of accuracy (Joyce 1998) — together with the Principal Principle (*PP*) — the chance-rational credence coordination principle that Lewis (1986, 87) himself advocates, and which is defended by Pettigrew (2012; 2013) *inter alia*.

However, as Bradley (2012) (cf. Fine 1988, 399, 401–3; Joyce 2010, 292) points out, it can't be shown that someone with imprecise degrees of belief can be subjected to a Dutch Book without making highly dubious assumptions about the decision rule that they adopt (cf. Frigg et al. 2014, esp. 55–6). Moreover, while accuracy arguments can be used to show that numerically precise credences ought to obey normalization, non-negativity, and finite additivity, it's dubious whether they can be deployed to show that degrees of belief must be numerically precise (Joyce 1998, 600–2).[33] Moreover, if chance is assumed to be imprecise for the sake of argument, then the *PP* itself stands in need of revision, as I'll argue below. If the resulting principle were substituted for the *PP* in Lewis's argument, the argument would no longer be valid. So it seems that there's no non-question-begging argument along these lines for thinking that chance must be precise.

Imprecise chance — on the assumption that it exists — can be represented as a set $\boldsymbol{ch}$ of probability functions. On the present interpretation, a probability function ch_i is an element of $\boldsymbol{ch}$ iff ch_i is entailed by one of the set of tied-for-best systems. Call the set $\boldsymbol{ch}$ the *cadentor*.[34] Each element of the cadentor is a precise probability function $ch_i(\cdot|\cdot)$ that associates ordered pairs of propositions $\langle X, Y\rangle$ with a unique real number, $x(\in \mathbb{R})$, in the $[0,1]$ interval: $ch_i(X|Y) = x$. (Again, unconditional probabilities are defined via $ch_i(X) =_{def} ch_i(X|\top)$.) Abusing notation slightly, we can let $\boldsymbol{ch}(\cdot|\cdot)$ represent the function that maps ordered pairs of propositions $\langle X, Y\rangle$ to the set of values $\boldsymbol{x}$ that the probability functions in $\boldsymbol{ch}$ give for X conditional upon Y (where such probabilities are defined): that is $\boldsymbol{ch}(X|Y) = \{ch_i(X|Y) : ch_i \in \boldsymbol{ch}\}$.

6.4 *The Imprecise Chance-Credence Connection*

How ought rational credence be constrained by evidence of imprecise chances? Walley and Fine (1982, 747–51) take physical processes modeled by upper and lower probabilities to warrant corresponding upper and lower expectations. Fierens et al. (2009) (cf. Walley and Fine 1982, 752–7) provide ways of estimating a set of probability distributions (where the set is given an objective interpretation, and taken to model

[32] These authors typically replace finite additivity with finite superadditivity of lower chances and finite subadditivity of upper chances. See, for instance, Bradley (2016, Formal Appendix) for details.

[33] Schoenfield (2017) and Konek (2019) dispute whether the view that the only epistemic good is accuracy (sometimes called *accuracy-first epistemology*) is consistent with the view that imprecise credences can be rationally permitted or required, with Schoenfield giving an argument that accuracy-firsters can't allow that they're rationally permitted, and Konek arguing that accuracy-firsters should allow that they're sometimes rationally required. Note, however, that one can take accuracy to be an epistemic good without taking it to be the sole epistemic good. And if one's a pluralist about epistemic value, then it's easy to sidestep Schoenfield's argument. Indeed, without going into details, I find it more plausible to take Schoenfield's argument — if (*contra* Konek) it's otherwise successful — as constituting a *reductio* against accuracy-first epistemology than as telling against the rationality of imprecise credences. Nevertheless, I do believe that accuracy is *an* epistemic good, but just that there may be others.

[34] After the Latin *cadentia* from which the English word 'chance' ultimately derives.

a physical process) from a data sequence by computing the relative frequencies along some of its subsequences, and they treat such estimates as a basis for prediction (Fierens et al. 2009, 1883). Fine (1988, 399) also suggests that imprecise subjective probabilities are the correct response to imprecise chances (cf. Walley 1991, 358n).[35]

This points in the direction of a natural proposal for how imprecise chances constrain rational credences which, roughy speaking, is that when an agent learns what the set-valued chance is for A, her credence in A ought to come to take the same set of values. Hájek and Smithson (2012, 38) make a suggestion along these lines — though they state it in terms of a rational agent's credences *conditional* upon a set-valued chance for A rather than her credences upon *learning* a set-valued chance for A[36] — when they say:

> [Y]our credence in a proposition, conditional on the chance of that proposition being indeterminate in a particular way, should be indeterminate in the same way. The indeterminacy in the chance ... is inherited by your conditional credence.

The following is an attempt to state this idea more precisely.

Since I'm taking conditional chance to be basic, I prefer to begin with a formulation of the chance-credence connection that — in contrast to Lewis's (1980) *PP* — takes *conditional* chances to be the quantities that (in the first place) constrain rational credence.[37] I take it to be more or less platitudinous that precise conditional chances place the following constraint on rational credence. Suppose that a rational agent S's total evidence is given by the proposition E together with the proposition that the chance of A conditional upon E is x (which, for now, is assumed to be a unique real number). Such an agent has a credence in A equal to x.

More precisely, let $ch(\cdot|\cdot)$ be the ur-chance function (Hall 2004, 95). Let $cr(\cdot|\cdot)$ be any reasonable initial credence function (the credence of a Bayesian 'superbaby' – cf. Hájek ms b). (For now, both $ch(\cdot|\cdot)$ and $cr(\cdot|\cdot)$ are assumed to be unique, real-valued conditional probability functions.) Suppose that $\{A, E\}$ is any pair of propositions such that there's a well defined chance $ch(A|E)$. Then conditional chances guide rational credence in the sense captured by (*Cond*):

$$cr^{\ulcorner ch(A|E)=x \urcorner \& E}(A) = x \qquad (Cond)$$

Here, $\ulcorner ch(A|E) = x \urcorner$ is the proposition that the chance of A conditional upon E is x (cf. Hall 2004, 99), while $cr^{\ulcorner ch(A|E)=x \urcorner \& E}(\cdot)$ is the credence distribution that results from updating $cr(\cdot|\cdot)$ upon the proposition $\ulcorner ch(A|E) = x \urcorner \& E$ and conditioning upon $\top$.[38]

The most natural generalization of *Cond* to the case of imprecise

[35] While Fierens et al. (2009, 1879) prefer a representation of imprecise probabilities as sets of measures, Fine (1988) appears to prefer to work with lower probabilities.

[36] The two ways of putting the point are equivalent on the representor approach to imprecise credences if the agent updates in accordance with (*).

[37] I'm grateful to Richard Pettigrew for reminding me that, as Hall (2004, 100–1) argues, chance-credence coordination principles that appeal to *conditional* chances needn't appeal to Lewis's (1980) notion of 'admissible evidence'. This enabled me to formulate the coordination principles that I discuss in the following in a simpler way than they were formulated in an earlier version of the manuscript.

[38] *Cond* doesn't say how the update on $\ulcorner ch(A|E) = x \urcorner \& E$ proceeds. One obvious suggestion is that it proceeds by conditionalization and that, for a rational agent, *CCP* ('Conditional Chance Principle') holds (where A, E, $\ulcorner ch(A|E) = x \urcorner$, and $cr(\cdot|\cdot)$ all have the same interpretation as above):

$$cr(A|\ulcorner ch(A|E) = x \urcorner \& E) = x \qquad (CCP)$$

CCP is an analogue of *PP* where conditional chances are taken to guide rational credence. A principle along the lines of *CCP* is described by Hall (2004, 101) (cf. Hoefer 2007, 574–5; Schwarz 2014, 88).

chances and credences is given by what I'll call the Mushy Principle (or MushyP):[39]

$$\boldsymbol{cr}^{\ulcorner \boldsymbol{ch}(A|E)=\boldsymbol{x}\urcorner \& E}(A) = \boldsymbol{x} \qquad \text{(MushyP)}$$

Here, A and E have the same interpretation as before. $\boldsymbol{cr}$ is any reasonable initial representor (that is, any set of probability functions that could model a rational agent's initial epistemic state), while $\boldsymbol{cr}^{\ulcorner \boldsymbol{ch}(A|E)=\boldsymbol{x}\urcorner \& E}$ is the representor that results from updating the credal state represented by $\boldsymbol{cr}$ upon the proposition $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$, where $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner$ is the proposition that $\boldsymbol{x}$ is the set of probabilities that the various elements of the cadentor entail for A conditional upon E. MushyP thus states that a rational agent S, whose total evidence is the conjunction of E with the proposition that the set-valued chance for A conditional upon E is $\boldsymbol{x}$, has a credence in E represented by the set of values $\boldsymbol{x}$.[40]

MushyP doesn't say anything about how the update on $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$ occurs. As we saw in Section 6.2, the orthodox approach is that updating an imprecise credal state $\boldsymbol{cr}$ upon some proposition D involves updating — by conditionalization — *each* probability function $cr_i \in \boldsymbol{cr}$ for which the update operation is defined upon D. (Where the update operation is undefined, cr_i is 'weeded out' rather than updated in the transition to $\boldsymbol{cr}^D$.) Where D is imprecise evidence — as D is imprecise evidence for A if D is the proposition $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$ — the idea is that the different elements of $\boldsymbol{cr}$ 'interpret' D in different ways, in the sense described in Section 6.2.

If — as I've claimed — MushyP correctly describes how rational imprecise credence is constrained by knowledge of imprecise chances and if the orthodox model of updating an imprecise credal state upon evidence — given by (∗) — is correct, then it follows that, for any rational initial representor $\boldsymbol{cr}$, for every possible cadentor $\boldsymbol{ch}$, and for all and only values x_i such that $ch_i(A|E) = x_i$ for an element $ch_i \in \boldsymbol{ch}$, there must be an element $cr_j \in \boldsymbol{cr}$ such that $cr_j^{\ulcorner \boldsymbol{ch}(A|E)=\boldsymbol{x}\urcorner \& E}(A) = x_i$.[41]

This is slightly unsatisfactory in that we only get the correct post-update 'spread' for a rational agent's credence as a result of the somewhat *ad hoc* stipulation that her initial representor must comprise probability functions that yield this spread when updated by conditional-

[39] The name is inspired by the fact that imprecise probabilities are sometimes described as 'mushy' (and by the memorability of its abbreviation for fans of British cuisine!).

[40] In comments on an earlier version of this paper, Richard Pettigrew and Jason Konek suggested that there's another plausible principle, which is consistent with MushyP and which describes a further way in which rational credence is constrained by imprecise chance. This is MushyP*:

$$\boldsymbol{cr}^{P\&E}(\cdot|\cdot) = \boldsymbol{ch}(\cdot| \cdot \& E) \qquad \text{(MushyP*)}$$

Here, P is the proposition that $\boldsymbol{ch}$ is the cadentor. MushyP* says that the representor of an agent whose total evidence is $P\&E$ ought to be equivalent to the set of conditional distributions arrived at from the cadentor by taking E as a fixed condition. I find this principle plausible. However, it doesn't render MushyP redundant, since MushyP tells us how rational credence responds to knowledge of the set-valued chance for an *individual* proposition A conditional upon E. It's possible, and arguably quite common, to have such knowledge without knowing exactly what the cadentor is, just as in the precise case one might know the chance of some proposition (e.g. that a given tritium atom will decay within 12.32 years) conditional upon one's evidence without knowing what the complete chance distribution is over the full algebra of events on which chance is defined. In what follows, for reasons of space, I'll confine myself to an attempt to defend MushyP without also seeking to defend MushyP* and also to attempting to show how the probabilities entailed by the set of tied-for-best systems for our world play the MushyP role in guiding rational credence without also seeking to explicitly show that the entailed probability functions play the MushyP* role. However, I hope the following will at least make clear what sort of strategy I would pursue in seeking to defend MushyP* (and the view that the probability functions entailed by the tied-for-best systems play the MushyP* role), even if the details of that defence must await another occasion.

[41] Thanks to an anonymous referee for helping me to more accurately characterize the constraints that an initial representor must meet if, when updating occurs in accordance with (∗), it is to satisfy MushyP.

ization upon any possible proposition of the form $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$. Perhaps we can get a more satisfactory model of how rational post-update credence conforms to MushyP by adopting a heterodox update principle. In fact, when imprecise evidence takes the form of an imprecise chance, a particular heterodox update model suggests itself.

On this heterodox model, S's post-update representor is arrived at by updating (perhaps by conditionalization) *each* of the probability functions $cr_i(\cdot|\cdot)$ in her initial representor upon *each* proposition of the form $\ulcorner ch_j(A|E) = x_j\urcorner \& E$, where $\ulcorner ch_j(A|E) = x_j\urcorner$ is a true proposition stating that an element, j, of the cadentor entails that the probability of A conditional upon E is x_j. That is, $\boldsymbol{cr}^{\ulcorner \boldsymbol{ch}(A|E)=\boldsymbol{x}\urcorner \& E}(\cdot) = \{cr_i^{\ulcorner ch_j(A|E)=x_j\urcorner \& E}(\cdot) : cr_i \in \boldsymbol{cr}, ch_j \in \boldsymbol{ch}\}$. If each update of a $cr_i \in \boldsymbol{cr}$ upon a proposition $\ulcorner ch_j(A|E) = x_j\urcorner \& E$ (for $ch_j \in \boldsymbol{ch}$) conforms to *Cond* — that is, if for each such update, cr_i treats x_j as the unique chance of A conditional upon E and cr_i acts as though it were a unique, rational credence function — then any initial representor $\boldsymbol{cr}$ such that, for each element $ch_j \in \boldsymbol{ch}$, there's a $cr_i \in \boldsymbol{cr}$ with respect to which the update on $\ulcorner ch_j(A|E) = x_j\urcorner \& E$ is defined, will yield a post-update representor satisfying MushyP.

This heterodox update model allows us to be more permissive than the orthodox model does with regard to the initial representor that a rational agent may come endowed with if her post-update credence is to conform to MushyP. Still, there are two worries about this heterodox update model. The first concerns motivation. It's not clear *why*, in each update of a $cr_i \in \boldsymbol{cr}$ upon a proposition $\ulcorner ch_j(A|E) = x_j\urcorner \& E$ (for $ch_j \in \boldsymbol{ch}$), cr_i should treat x_j as the unique chance for A conditional upon E. It's also not clear *why*, in each such update, cr_i should act as though it were a unique, rational credence function. Finally, it's not clear *why* every $cr_i \in \boldsymbol{cr}$ should be updated on each proposition $\ulcorner ch_j(A|E) = x_j\urcorner \& E$ (for $ch_j \in \boldsymbol{ch}$) in the transition to the new representor. Perhaps an adequate answer to these questions is simply that this is a way of generating the rationally correct post-update credences from a rational pre-update representor without imposing implausibly strong constraints upon what rationality requires of the pre-update representor. Such a response might be particularly plausible if we conceive of what we're doing in 'modeling' such updates not as aiming to describe real psychological processes that a rational agent must be subject to (let alone conscious of), but just as describing some algorithm that's capable of implementing the function from rational prior credal states (which perhaps aren't as constrained as the orthodox update model would have it) to rational posterior credal states that reflect the imprecision of the evidence acquired.

Yet there's a second, and I think more serious, worry about the heterodox model described. This concerns its generalizability to other sorts of imprecise evidence. The heterodox proposal involves the idea that there are various 'precisifications' of an imprecise chance: Specifically, each proposition of the form $\ulcorner ch_j(A|E) = x_j\urcorner \& E$ is, in a sense, a precisification of the imprecise evidence for A that comprises $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$. The idea is that each probability function in the agent's representor should be updated upon each precisification of the evidence to yield the updated representor. Yet there are examples of imprecise evidence (see e.g. Elga 2010, 1) where it's not clear what a 'precisification' of the evidence would be.

So I have no entirely satisfactory story about the update mechanism. Yet note that I'm in no worse a position here than any other imprecise probabilist who claims that imprecise evidence issues in rational credence that's imprecise in a similar way (i.e. anyone who adopts something like the CMT). If they adopt the orthodox update model, they will have to say that any rational initial representor must contain elements that interpret any possible imprecise evidence D in just the right spread of ways that updating the representor in accordance with (*) yields a posterior spread of credence that matches the imprecision of the evidence updated upon. Likewise, if they adopt the heterodox model, they will face the burden of explaining what counts as a precisification of imprecise evidence where that imprecise evidence doesn't take the form of an imprecise chance. If it turns out that there's no reasonable update model for imprecise credences, then this might furnish a reason for thinking that only precise credences are rational. Still, as we've seen, independent reasons have been given for thinking that imprecise credences are rational in some circumstances. So we must acknowledge here as elsewhere that the ultimate success of the present arguments will be partly contingent upon the outcome of more general debates — having nothing specifically to do with chance — between precise and imprecise probabilists.

Hájek and Smithson (2012, 38–9) have claimed that something along the lines of MushyP is plausible:

> If you regard the chance function as indeterminate regarding [some proposition] X, it would be odd, and arguably irrational, for your credence to be any sharper. Compare: if your doctor is your sole source of information about medical matters, and she assigns a credence of [0.4, 0.6] to your getting lung cancer, then it would be odd, and arguably irrational, for you to assign this proposition a sharper credence — say, 0.5381. How would you defend that assignment?

In what follows, I'll attempt to given an explicit defense of MushyP. My strategy will be two-pronged. First, I'll consider the most obvious and/or plausible rivals to MushyP and argue that it's irrational to calibrate one's credences as these principles suggest. Of course it might be that there are alternatives of which I haven't conceived that don't lead to irrationality. Nevertheless, this first prong of my strategy at least lays down a challenge to the opponent of MushyP to devise and defend some such rival principle.[42] The second prong of my strategy will be to show how an epistemic utility argument that Carr (2015) has advanced for CGT can be readily adapted into a direct argument for MushyP. This direct defense reduces the plausibility that any unconceived rival principle is at least as defensible as MushyP.

Recall that MushyP says that, when a rational agent updates her

[42] As indicated in Footnote 40, MushyP* isn't a rival, since it's compatible with MushyP.

initial representor upon the proposition $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x} \urcorner \& E$, where this proposition represents her total evidence, her post-update credence in A is $\boldsymbol{x}$. Perhaps the most obvious rivals to this principle are:

Rival 1: S's post-update credence is the mean (or some other weighted average) of $\boldsymbol{x}$.

Rival 2: S's post-update credence is the convex hull of $\boldsymbol{x}$.

Rival 3: S's post-update credence is some proper subset of $\boldsymbol{x}$.

Rival 4: S's post-update credence contains values that lie outside $[inf\{\boldsymbol{x}\}, sup\{\boldsymbol{x}\}]$.

Rival 3 and Rival 4 aren't unique alternatives to MushyP, but describe classes of such rivals. Indeed, in some cases, Rival 1 will be a special case of Rival 3. This will be so if, for instance, $\boldsymbol{x}$ contains its own mean.

I think that instances of Rival 3 that aren't instances of Rival 1 can be passed over fairly quickly. Rival 1 (where it's an instance of Rival 3) is a specific proposal about *which* subset of $\boldsymbol{x}$ one should adopt as one's post-update credence, and (as we'll see) it comes associated with some *prima facie* motivations for *why* this is the relevant subset. I'm not aware of any motives for adopting an alternative instance of Rival 3. In general, in adopting Rival 3, you're acting as though a subset of $\boldsymbol{ch}$ is privileged. The question — as Hájek and Smithson (2012) pose it in the passage quoted above — is what justifies you in doing this? In the absence of some such justification, it's not clear *which* subset should be taken as privileged. As noted, Rival 1 comes associated with a story about this. It is of course possible that someone might devise some story that would motivate some alternative version of Rival 3. However it's not obvious to me how such a story would go. In the absence of such a story, it seems that MushyP, which is backed by the plausibility of CMT (and also the positive argument for MushyP to be given below), should be presumed preferable.

So let's now consider Rival 1. It's worth noting, first of all, that if the elements of $\boldsymbol{x}$ are infinite and $\boldsymbol{x}$ is convex, then the natural analogue of taking the mean of x would be to take the mid-point of the interval $[inf\{x\}, sup\{x\}]$. If we prefer a version of Rival 1 that, in the finite case, recommends a credence that's some weighted average of x with non-equal weights, then the natural generalization to the infinite case might be to choose some other point in the interval $[inf\{x\}, sup\{x\}]$ as our recommended post-update credence. If the elements of x are infinite and x is non-convex, then it's not obvious that there's a natural analogue of Rival 1. Still, some of the following considerations raise general concerns about responding to imprecise chances with any particular precise credence (concerns that don't simply involve invoking the CMT).

Rival 1 certainly might be appealing to those who are skeptical of the rationality of adopting imprecise credences. Wheeler and Williamson (2011, 324) describe an analogue of this view for the case where the chance of an outcome is *unknown* (thus describing a view that appears at odds with the CGT):

> As far as you are aware, the physical probability of a is now 1 or 0 and no value in between. But this does not imply that your degree of belief in a should be 1 or 0 and no value in between — a value of $\frac{1}{2}$, for instance, is quite reasonable in this case.

One way one might seek to bolster Rival 1 is by noting that the set of probability functions in the cadentor is plausibly regarded as a set of expert functions — *analyst experts* in Hall's (2004, 101–2) sense — and that you should respond to learning the set of probabilities these functions entail for A conditional upon the remainder of your total evidence E in the same way you'd respond to learning the set of probabilities for A conditional upon E assigned by any other group of experts.[43] One proposal[44] for how one should respond to experts is that one should take a linear average of their probability assignments. Where one has no grounds for trusting one of the experts any more than any other, a natural view is that one should take a linear average with equal weights (i.e. the arithmetic mean) of their probability assignments as one's own credence.

Yet there are reasons for thinking that linear averaging isn't in general the correct response to learning the opinions of experts. One reason is that it's not, in general, consistent with Bayesian conditionalization (Dawid et al. 1995; Bradley 2018). So it's not clear that it has a general attractiveness as a principle for responding to expert opinions that makes it an appealing way to respond to imprecise chances.

There are more direct objections to Rival 1. First, what justifies the (equal) weighting assigned to the various functions in the cadentor? The only obvious proposal for justifying equal weights is to appeal to the sort of Principle of Indifference-like reasoning that advocates of imprecise credence reject as failing to accurately reflect one's (lack of) evidence (cf. Bradley 2012, 4). This might not unduly concern fans of precise credences, but nevertheless, independent arguments against the Principle of Indifference (PoI) tell against this sort of justification of equal weights. And it's still less clear what would justify non-equal weights.

Perhaps even more troubling is that linear averaging doesn't preserve independence (Laddaga 1977). To see this, let us adapt an example from Jeffrey (1987). Suppose that two coins are about to be tossed, where H_1 and $\neg H_1$ are the two possible outcomes of the first toss and H_2 and $\neg H_2$ are the two possible outcomes of the second. This yields a set of four atomic events: $\{H_1\&H_2, \neg H_1\&H_2, H_1\&\neg H_2, \neg H_1\&\neg H_2\}$, over which an algebra $\mathcal{A}$ is defined. For simplicity, suppose that there are just two probability functions in the cadentor, ch_1 and ch_2. Suppose that $ch_1(H_1) = ch_1(H_2) = \frac{1}{3}$. Suppose, moreover, that ch_1 regards the tosses as independent, so that $ch_1(H_1\&H_2) = \frac{1}{9}$, $ch_1(\neg H_1\&\neg H_2) = \frac{4}{9}$, and $ch_1(H_1\&\neg H_2) = ch_1(\neg H_1\&H_2) = \frac{2}{9}$. Suppose that $ch_2(H_1) = ch_2(H_2) = \frac{2}{3}$ and ch_2 regards the tosses as independent, so that $ch_2(H_1\&H_2) = \frac{4}{9}$, $ch_2(\neg H_1\&\neg H_2) = \frac{1}{9}$, and $ch_2(H_1\&\neg H_2) = ch_2(\neg H_1\&H_2) = \frac{2}{9}$.

[43] Thanks to an anonymous referee for pushing this way of thinking about it.
[44] Discussed by, for example, Dawid et al. (1995) and Bradley (2018).

The trouble is that, although each of ch_1 and ch_2 takes the tosses to be independent, linear averages of the probabilities assigned by each of them to the atomic events $H_1\&H_2$, $\neg H_1\&H_2$, $H_1\&\neg H_2$, $\neg H_1\&\neg H_2$ result in distributions that entail non-independence of the tosses (except if extreme, 1/0, weights are given to the chances entailed by ch_1 and ch_2). For instance, taking a 'straight average' (i.e. averaging with equal weight) of the probability assignments that ch_1 and ch_2 entail, we get the assignments $p(H_1\&H_2) = p(\neg H_1\&\neg H_2) = \frac{5}{18}$ and $p(H_1\&\neg H_2) = p(\neg H_1\&H_2) = \frac{4}{18}$, implying that $p(H_2|H_1) > p(H_2)$.

If all elements of the cadentor entail that the chances of two events are independent, it seems highly dubious that any rule can be a principle of rationality if it requires (or even permits) that credence, upon learning what the distributions in the cadentor are (and nothing else of relevance), should come to treat the two events as non-independent.[45,46] Indeed, Loewer and Laddaga (1985, 90) point out that linear averaging actually results in inconsistency. To see this, suppose that we refine the partition we've been considering by subdividing each cell into two subcells, according to whether I or $\neg I$ holds, where I is the proposition that H_1 and H_2 are independent. Since ch_1 and ch_2 both assign probability 1 to I, a distribution p that takes a straight average (or any weighted linear average) of ch_1 and ch_2's assignments also assigns 1 to I even though H_1 and H_2 are not in fact independent according to p.[47] While there are adjustments that can be made to a simple linear averaging procedure that preserve independence (see Wagner 1985; 2011), these adjustments are rather *ad hoc*.

Partly because of issues surrounding the preservation of independence, Elkin and Wheeler (2018) (cf. Seidenfeld et al. 1989, esp. 241–2) regard imprecise probability assignments as the correct solution to the problem — closely related to the problem of responding to expert opinion — of opinion aggregation for groups of peers, where "the span between lower and upper probabilities for a proposition is determined by the range of judgments expressed by a group of peers" (Elkin and Wheeler 2018, 15)[48] and where convexity isn't mandated (Elkin and Wheeler 2018, 13; cf. Seidenfeld et al. 1989, esp. 241–2). Levi (1982) and Fine (1980) also treat imprecise probability assignments as the correct solution to the problem of opinion aggregation. Kyburg Jr. and Pittarelli (1992, 151–3; 1996, 334–5) give reasons for thinking that taking the convex hull of the agents' probability distributions (as opposed to the set of distributions itself) to represent the group belief may be undesirable.

Although some of these authors frame the point in terms of 'epistemic peers', it seems that, where you are confronted with a group of experts who are themselves epistemic peers with one another (and you know this), it would be a good idea for you to strive to adopt the aggregated opinion that they would if they employed a suitable aggregation rule. Indeed, Fine (1980; 1988, 403–4) argues specifically that imprecise probability assignments are the correct solution to the problem of aggregating expert opinions and that non-experts should aim to adopt the resulting imprecise probabilities and use them in decision-making (Fine 1980, e.g. 27; 1988, 403–4).[49] Thus, far from a linear averaging proposal for responding to imprecise chances being motivated by the

45 In the context of opinion aggregation, many authors have found plausible a principle that Elkin and Wheeler (2018, 5) call the Preservation of Irrelevant Evidence Principle: namely, that if each member of the group whose opinion is to be aggregated agrees that P is independent of Q (and wouldn't change her mind after learning the other agents' credences), then the aggregated opinion should take P and Q to be independent.

46 If the cadentor is convex, then the worry is slightly more subtle: namely, that it's hard to see why one's credence should treat two or more events as determinately dependent if not all elements of the cadentor treat them as dependent. Yet adopting credences in atomic events that are the mid-points of intervals of probabilities entailed by the cadentor is liable to lead to such a result.

47 Elkin and Wheeler (2018, 5–6) exploit this fact to show that someone with a credence represented by p can be subjected to an expected sure loss by a clever bookie and, indeed, if offered a bet on H_2 in ignorance of the outcome of the first coin toss, could be induced to pay for information about the outcome of the first toss.

48 Cf. Fine (1980, 27–9; 1988, 403–4).

49 Fine (1980), however, represents imprecise probabilities using interval-valued probabilities, rather than sets of probability distributions.

notion that this is in general the correct response to learning a set of expert opinions, it seems that linear averaging is the incorrect response and imprecise credence may well be the correct response to *both* the problem of responding to a set of expert opinions *and* to the problem of responding to imprecise chances.

The present considerations also tell against Rival 2 — which is that, upon learning the set of values for A that are entailed by elements of the cadentor conditional upon the remainder of one's total evidence, E, one's credence in A should correspond to the convex hull of that set of values (cf. Levi 1980, Ch. 9). Of course, if the probabilities entailed by elements of the cadentor always form an interval, then this proposal is equivalent to MushyP and unproblematic. However, if they don't, such an update rule is liable to result in trouble. For instance, consider the coin case described above and suppose that there are only two elements of the cadentor, yielding the distributions ch_1 and ch_2 described above. If, upon learning the probabilities that these distributions entail for each of the atomic events in that case, one responded with credences in each of these atomic events that comprise all convex combinations of the probabilities entailed by ch_1 and ch_2, then the result would be that the overwhelming majority (measure ≈ 1) of the distributions in one's representor regard the tosses as dependent, despite all elements of the cadentor regarding them as independent.[50] In such a scenario, one would appear to have good evidence that the sequence of tosses is i.i.d. and yet one's credal state seemingly wouldn't reflect this fact (cf. Joyce 2005, 171n; 2010, 296n).

Kyburg Jr. and Pittarelli (1992, 150–1) (cf. Kyburg Jr. and Pittarelli 1996, 333–4) show that, if an agent's representor is the convex hull of a set of probability distributions (chance hypotheses) that is itself non-convex, then the popular E-admissibility decision rule for imprecise credences — which says that an act is permissible iff it maximizes expected utility relative to at least one probability function in the agent's representor — will permit her to accept series of gambles that appear irrational, precisely because the convex hull of these probability distributions contains elements that treat as dependent events that each of the original probability distributions treats as independent. As they put it (labeling the set of chance hypotheses S): "posting odds in accordance with a probability function not in S but in the convex hull of S may result in negative long-run expectation relative to the knowledge embodied in S" (Kyburg Jr. and Pittarelli 1996, 338).

It's not clear that Γ-Maximin would also permit irrational decisions if representors were required to be convex — which seems more pertinent given that Γ-Maximin avoids Elga's (2010) objections to decision rules for imprecise credences.[51] But nevertheless, there does appear to be something epistemically defective about a situation where one has strong evidence that the tosses of a coin flip are independent, and yet one's credal state doesn't clearly reflect that evidence, even if this doesn't induce irrational betting behavior. Indeed, if convexity were required, then it becomes very difficult to see how imprecise credal states could encode knowledge or strong beliefs about independence of the coin flips in the example given above. Their being independent according to all (or even a significant proportion) of elements of the agent's representor is incompatible with convexity. A strong belief about independence could therefore only be represented if the partition were refined so that explicit probabilities concerning independence could be represented. We would then encounter the trouble that all elements of the agent's representor would assign probability 1 (or perhaps high probability) to independence even while not treating the tosses as independent.[52]

[50] Thus, in Pedersen and Wheeler's (2014, 1325) terminology, H_1 and H_2 would fail to be 'completely stochastically independent' according to the agent's representor (cf. Cozman 2012, 581) (indeed each of H_1 and H_2 would fail to be 'epistemically irrelevant' — cf. Couso et al. 2000 — to the other), even though they are completely stochastically independent according to the cadentor.

[51] See Footnote 26.

[52] See Bradley (2012, 4) for a different objection to the requirement that representors always entail convex sets of probabilities.

The very notion that a rational agent should update her credal state by taking convex combinations of possibly non-convex sets of probabilities assigned by various chance hypotheses seems wrong-headed. After all, rational agents often arrive at joint distributions by deriving them from marginals together with independence assumptions justified by their evidence (cf. Couso et al. 2000). In a case where the agent is certain about independence, but where the marginals are imprecise, it seems very doubtful that her credal state should be the set of convex combinations of the joint chance distributions, given that many such convex combinations will treat the variables as dependent.[53]

Finally, consider Rival 4. This seems to me to have fairly little *prima facie* plausibility. It's true that some — such as Easwaran et al. (2016) and Bradley (2009, 249) — have advocated principles for opinion aggregation, and even for updating on the credences of experts (Easwaran et al. 2016, 28) according to which, where $\boldsymbol{P}(A)$ is a set of (precise) probabilistic opinions regarding A to be updated on/aggregated, the post-update/aggregation probability for A needn't lie in $[inf\{\boldsymbol{P}(A)\}, sup\{\boldsymbol{P}(A)\}]$. However, Easwaran et al. (2016, 32) are clear that their rule doesn't apply where the initial opinions of the experts result from estimates of the chance of A, which is precisely the case where the elements of the cadentor are taken as the expert functions. Meanwhile, Bradley's (2009, 249) suggestion is motivated by the thought that there might be synergies between the various experts' evidence for a given outcome. But we're here concerned with how an agent should respond when she learns that different elements of the cadentor entail different probabilities *conditional on precisely the same evidence*. So neither of these approaches to opinion pooling can be used to provide a *prima facie* motivation for Rival 4. And, in general, it's difficult to see how Rival 4 — which, as Rival 1 and Rival 2 will sometimes do, recommends that upon learning a set-valued chance for A, elements of one's representor should entail probability values for A that are determinately *not* the chance of A — could be motivated. Indeed, unsurprisingly, if one responds to learning a set-valued chance in the way described by Rival 4, and one adopts just about any plausible decision rule that has been discussed for agents with imprecise credences (see Troffaes 2007 and Bradley ms for surveys), then one will be opened up to either betting according to, or at least having one's betting behavior influenced by, probabilities that are determinately not the chances, even when one knows the chances.[54]

Let us, finally, consider an explicit epistemic utility argument for MushyP. Following Hájek (ms a), Pettigrew (2012, 261), and Carr (2015, 71), I assume rational credence to be chance-directed. And, in particular, I assume that a rational agent with total evidence $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$ aims to calibrate her credence to $\boldsymbol{ch}(A|E) = \boldsymbol{x}$. Thus, what follows is not an argument for that assumption, but rather for the view that MushyP (rather than rivals 1–4, or some other rival that someone might devise) is the correct calibration principle.

The standard blueprint for giving an epistemic utility justification of some epistemic norm involves, firstly, designating some probability or probability function (or set of probabilities or set of probability functions) as 'vindicated': roughly speaking, as aim-worthy for reasonable credence. The next step is to propose some measure of the distance between probabilities or probability functions (or sets of probabilities or sets of probability functions) and some 'epistemic decision rule' — roughly speaking, some rule concerning how rational credence ought to respond to facts about distance from vindication — such that an agent adopting the epistemic norm will arrive at credences that harmonize with those of an agent following the epistemic decision rule. To the extent that the epistemic decision rule (and choice of vindicated probability (probabilities)/probability function(s) and distance

[53] Mayo-Wilson and Wheeler (2016, 61) analogously argue that, where one doesn't know which of a non-convex set of chance hypotheses is true, it would be 'strange' to require that one's credences be convex.

[54] Though I won't seek to show it here, it should be fairly obvious to anyone familiar with them that this will be true of the E-admissibility, Γ-Maximin, Γ-Maximax, Maximality, and Interval Dominance rules.

measure) is justified, the epistemic norm is supported. Or, as Konek (2019) suggests, the coherence of a plausible epistemic norm — as I claim MushyP to be — with plausible principles concerning the pursuit of epistemic value — such as those invoked in the epistemic utility argument — may provide "*symbiotic support*" (italics original) for both, providing "good reason to think that each component of your epistemology was on the right track".

An epistemic utility argument for MushyP can be arrived at by a simple adaptation of an argument due to Carr (2015) in favor of CGT, which Carr glosses as follows:

> [F]or each proposition A, evidence determines some set of possible chances of A at an appropriate time t. A rational agent's upper credence in A will be equal to the upper evidentially possible chance of A at t, and her lower credence in A will be equal to the lower evidentially possible chance of A at t. (Carr 2015, 70)

In contrast to Carr, I don't take chances to be inherently time-relative. Rather, like Hall (2004), I just take them to be relative to some evidential proposition F. I shall therefore not follow Carr in invoking the time-relativity of chances in what follows.

To start with, note that a very natural measure of the distance, δ, between two probability functions, $P(\cdot)$ and $P'(\cdot)$, is the average of the squared Euclidian distance between the probabilities assigned by the two functions to each proposition on which they're defined: that is, the Brier score.[55] The distance between the probabilities that the two functions assign to a single proposition A — i.e., $\delta(P(A), P'(A))$ — might therefore be taken to be given by $(P(A) - P'(A))^2$ (though, for present purposes, it would be harmless to alternatively take it to be the simple Euclidean distance between these probabilities). Like us, Carr wishes to allow that a rational agent's credence in A may be imprecise. Adapting definitions due to Carr (2015, 72), where (given an agent's total evidence F) there's a single vindicated probability value v for A, we can define lower and upper distances for imprecise credence as follows:

$$\delta^-(\boldsymbol{cr^F}(A), v) = \min_{cr_i^F(A) \in \boldsymbol{cr^F}(A)} \delta(cr_i^F(A), v) \qquad \text{(Lower Distance)}$$

$$\delta^+(\boldsymbol{cr^F}(A), v) = \max_{cr_i^F(A) \in \boldsymbol{cr^F}(A)} \delta(cr_i^F(A), v) \qquad \text{(Upper Distance)}$$

Carr's argument for CGT involves, rather than taking there to be a single 'vindicated' probability for A, taking there to be a *set* $\boldsymbol{v}$ of 'vindicated' probabilities for A. We can thus adapt Carr's (2015, 72–3) definitions of two further notions — lower-dominance and upper-dominance:[56]

[55] The Brier score is an instance of a quadratic loss scoring rule. Joyce (1998) and Pettigrew (2012) show that such rules satisfy a number of desiderata for distance measures in the case of precise credences (and a single vindicated function).

[56] These are modified versions of Carr's definitions. That's because she defines lower- and upper-dominance for sets of probability *distributions* and not just for sets of probability *assignments* to a single proposition, A. Her definitions thus invoke the 'global' lower- and upper-distance of a set of probability distributions from a vindicated distribution (Carr 2015, 72), rather than just the 'local' lower- and upper-distance of a set of probability assignments to a single proposition from a vindicated assignment. The definitions of the latter distance notions are the ones given in the main text above. Since MushyP is a principle concerning how knowledge of set-valued chances for individual propositions constrains credences in those propositions, we need only the notions of lower- and upper-dominance for sets of probabilities for single propositions. Although I won't attempt to show it here, by invoking Carr's 'global' lower- and upper-distance notions and her corresponding notions of lower- and upper-dominance for sets of distributions, we can provide an accuracy-based argument for MushyP* (see Footnote 40 above).

Lower-Dominance: $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A)$ *lower-dominates* $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A)$ iff, for each $v \in \boldsymbol{v}$, $\delta^-(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A), v) \leq \delta^-(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A), v)$ and for some $w \in \boldsymbol{v}$, $\delta^-(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A), w) < \delta^-(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A), w)$.

Upper-Dominance: $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A)$ *upper-dominates* $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A)$ iff, for each $v \in \boldsymbol{v}$, $\delta^+(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A), v) \leq \delta^+(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A), v)$ and for some $w \in \boldsymbol{v}$, $\delta^+(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A), w) < \delta^+(\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A), w)$.

Finally, Carr (2015, 72–3) proposes the following epistemic decision rule:

Lower-Then-Upper Dominance Avoidance: It's irrational to adopt $\boldsymbol{cr}^{\boldsymbol{F}}(A)$ as one's credence for A (given one's total evidence is F) if either (i) $\boldsymbol{cr}^{\boldsymbol{F}}(A)$ is lower-dominated; or (ii) $\boldsymbol{cr}^{\boldsymbol{F}}(A)$ is upper-dominated by any credence for A that's not lower-dominated.

Carr's strategy is to take the vindicated probabilities at t for an agent S to be the set $\boldsymbol{C}(A)$ of probabilities for A that are evidentially possible time t chances of A for S. I'm concerned with the case where S *knows* what the chance of A is, but the chance is set-valued. So the 'evidentially possible' chance of A for S is the set of probabilities for A entailed by the cadentor. Another slight difference between myself and Carr is that I take chances to be evidence- rather than time-relative. So rather than taking the vindicated probabilities at t to be the 'evidentially possible' time t chances of A, I instead take the vindicated probabilities *relative to S's total evidence* (which in the case we are interested in is given by $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$) to be given by the chance of A *conditional upon that evidence*.[57]

Carr's epistemic decision rule, when combined with the squared Euclidean distance measure, implies that the set of probabilities $\boldsymbol{cr}^{\boldsymbol{F}}(A)$ entailed for A by the elements of the representor of any rational agent with evidence F (on our interpretation $F \equiv \ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$) includes *all* elements of the set of vindicated probabilities (on our interpretation, the set $\boldsymbol{ch}(A|E)$) and that the upper- and lower-probabilities entailed for A by elements in her representor are no higher and lower (respectively) than the highest and lowest vindicated probabilities (Carr 2015, 73–4). To see this, note first that to avoid lower domination, a credence for A must include all elements $v \in \boldsymbol{v}$: Any credence in A, $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A)$ that doesn't contain some element $v \in \boldsymbol{v}$ is lower-dominated by a credence $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A)$ that contains all elements of $\boldsymbol{v}$. That's because, by containing all elements of $\boldsymbol{v}$, $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{2}}(A)$ has a lower distance of zero with respect to each of them, while, because it doesn't contain v, $\boldsymbol{cr}^{\boldsymbol{F}}_{\boldsymbol{1}}(A)$ has positive lower distance from v.

So all non-lower-dominated credences include all elements of $\boldsymbol{v}$. Among such credences, a credence will avoid upper-domination only if it contains no elements that lie outside the interval defined by the maximum and minimum elements in $\boldsymbol{v}$. Any credence that's non-lower-dominated includes the maximum and minimum elements in $\boldsymbol{v}$. Among such credences that don't include elements that lie outside of the interval defined by these maximum and minimum elements, their upper distance from each element $v \in \boldsymbol{v}$ is the same (thus no such credence is upper-dominated by any other): It's given by the squared Euclidean distance of v from the furthest end-point of the interval. Moreover, each such credence upper-dominates any that, in addition to containing all elements $v \in \boldsymbol{v}$ (thus avoiding lower-dominance), contains elements lying outside this interval. That's because any element lying outside the interval is more distant from the furthest end-point of the interval than the nearest end-point of the interval is to the furthest end-point of the interval.

Thus, an agent with evidence $\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E$ who conforms to Carr's proposed epistemic decision rule (with $\boldsymbol{ch}(A|E)$ taken to be the set of vindicated probabilities for an agent with this evidence) will be such that her credence $\boldsymbol{cr}^{\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E}(A)$ contains each element of $\boldsymbol{ch}(A|E)$ and is such that the maximum and minimum elements

[57] Strictly speaking, for simplicity, I've taken them to be given by $\boldsymbol{ch}(A|E)$ rather than $\boldsymbol{ch}(A|\ulcorner \boldsymbol{ch}(A|E) = \boldsymbol{x}\urcorner \& E)$. See Hall (2004) for a discussion of the distinction. Nothing I have to say here turns upon it.

of $\boldsymbol{cr}^{\ulcorner \boldsymbol{ch}(A|E)=\boldsymbol{x}\urcorner \& E}(A)$ are identical to the maximum and minimum elements of $\boldsymbol{ch}(A|E)$. For such an agent, calibrating her credences to imprecise chances in the way described by MushyP is thus consistent with following Carr's epistemic decision rule, whereas calibrating her credences to imprecise chances in the ways described by Rival 1, Rival 3, and Rival 4 isn't. To the extent that Carr's proposed epistemic decision rule is a good one, we thus have an additional reason to prefer MushyP over these rivals.

Carr's decision rule doesn't, however, allow us to discriminate between MushyP and Rival 2 (the proposal that a rational agent calibrates her credences to a set-valued chance by adopting a credence that's the convex hull of that chance). Presumably the reason that Carr herself doesn't seek such a rule is because she's neutral on whether imprecise credences are "determined by upper and lower credences in each proposition or whether they're more structured" (Carr 2015, 73).[58] I've already given independent reasons for preferring MushyP to Rival 2. Nevertheless, the following, fairly natural, modification of Carr's decision rule would also give us the result we want:

Lower-Then-Upper-Dominance-Avoidance-Then-Total-Distance-Minimization: It's irrational to adopt $\boldsymbol{cr^F}(A)$ as one's credence for A (given one's total evidence is F) if either

(i) $\boldsymbol{cr^F}(A)$ is lower-dominated; or

(ii) $\boldsymbol{cr^F}(A)$ is upper-dominated by any credence for A that's not lower-dominated; or

(iii) there's some $\boldsymbol{cr^{F\prime}}(A)$ that's not lower-then-upper dominated (i.e. that satisfies (i) and (ii)) such that

$$\sum_{cr_i^F(A)\in \boldsymbol{cr^F}(A), v\in \boldsymbol{v}} \delta(cr_i^F(A), v) > \sum_{cr_j^F(A)\in \boldsymbol{cr^{F\prime}}(A), v\in \boldsymbol{v}} \delta(cr_j^F(A), v)$$

The only difference from Carr's decision rule is that condition (iii) is added to penalize (among those credences that avoid lower-then-upper dominance) those such that the total distance of their elements from elements of the vindicated set of probabilities is greater than necessary.[59] This rules out any credence that contains elements that *aren't* elements of the vindicated set. Under our interpretation, it rules it irrational to have a credence in A that contains elements that aren't elements of the set-valued chance for A (conditional upon one's total evidence), as the convex hull of that set-valued chance will if the set-valued chance isn't itself convex. Obviously, this supports MushyP over Rival 2 only to the extent that this modification of Carr's decision rule is plausible. I feel some intuitive pull to the idea that one shouldn't have elements in one's set-valued credence that simply add to the total distance between elements of one's credence and elements of the set-valued chance (total distance seems like a reasonable thing for an epistemic decision rule for imprecise probabilities to be sensitive to). But this is perhaps bound up with the intuitive pull MushyP has over Rival 2 for me in the first place, as well perhaps as the independent reasons given above for preferring MushyP to Rival 2. So I won't set too much store by it, and I'm

[58] I suspect that the epistemological arguments for imprecise credences described in Section 6.2 point to rational credences being more structured. The fact that non-convex imprecise chances appear possible seems to me further reason for thinking this.

[59] Although condition (iii) gives the intuitive flavor of the requirement that we're trying to capture, since the elements of $\boldsymbol{cr^{F\prime}}(A)$ are infinite in number and don't form a convergent series, we should, strictly speaking, instead impose a requirement (iii′): *There's no* $cr_i^F(A) \in \boldsymbol{cr^F}(A)$ *such that there's no* $v \in \boldsymbol{v}$ *from which* $cr_i^F(A)$ *has zero distance*. As we've seen, condition (i) entails that a reasonable representor (for an agent with evidence F) contains all elements of $\boldsymbol{v}$. Condition (iii′) then adds that there not be any extraneous elements of the agent's representor (elements that are some positive distance from every element in $\boldsymbol{v}$ and that therefore, intuitively, just 'add' inaccuracy to the agent's representor). Note that the addition of condition (iii′) (just like the addition of condition (iii) in the finite case) renders condition (ii) redundant. So in fact our epistemic decision rule simplifies.

happy to let those reasons speak for themselves.[60]

7. Tied Systems and the Chance-Credence Role

MushyP seems plausible as an explication of the (imprecise) chance-credence connection, and I've outlined an epistemic utility justification for it, as well as arguing for it indirectly by arguing that its most obvious/*prima facie* plausible rivals are problematic. Still, there's the further question of whether the sets of probabilities entailed by the tied-for-best systems play the chance role in guiding rational credence in accordance with MushyP.

Perhaps the most promising attempts to argue that, in the case of a unique Best System, the (precise) Best System probabilities play something like the *PP* or *Cond* role are Loewer's (2004) 'reverse-engineering' argument, Schwarz's (2014) 'symmetry'-based argument, and Hicks' (2017) 'epistemic utility' argument. In this section, I'll argue that, in the case of a tie, these arguments can be adapted to support the claim that the sets of probabilities entailed by the tied systems play the MushyP role.

I don't claim that any of the above-mentioned arguments that, in the case of a unique winner, the (precise) best system chances play the *PP* or *Cond* role is entirely watertight. Rather, my more modest aim is to show that it's *as* plausible, in the case of a tie, that the probabilities entailed by the tied-for-best system play the MushyP role as it is that, in the case of a unique winner, the probabilities entailed by that system play the *PP* or *Cond* role.

First, let us consider Loewer's (2004) 'reverse-engineering' argument. Loewer's (2004, 1122) key idea is that a theory counts as *best* just by *being* a theory whose probabilities a rational agent would calibrate her credences to. In other words, the standards for 'bestness' are to be thought of as 'reverse-engineered' from the chance-credence connection. For a system to count as *best* just is for it to be victorious under those standards of strength, simplicity, and fit, and that exchange rate between them, such that a system that's victorious under just those standards and that exchange rate is one that entails rational-credence guiding probabilities. Presumably, if the best-fitting system isn't automatically best, then there must be some story about why probabilities that diverge from the actual frequencies because of the relative simplicity of the system that entails them are targets for rational credence. One such reason will be considered below.

In any case, it seems that we can adapt Loewer's argument to the case where there are ties. If we adopt Loewer's interpretation then, where there are ties, there are multiple systems entailing probabilities that have equal claim on rational credence. On the face of it, it seems quite plausible that this is so[61] given that, as was argued in Section 4, there appears to be a multitude of very similar systems entailing similar but not identical probabilities for SM. The systems that have equal claim on rational credence are victorious under different (reverse-engineered) standards of simplicity, strength, fit, and balance. The fact that the probabilities entailed by such systems have strong and equal claims upon rational credence is analogous to the situation in which one is confronted with the probabilistic opinions of a group of experts and, as noted in Section 6.4, there are arguments in the literature (e.g.

[60] One might be skeptical of the foregoing epistemic utility argument for MushyP in light of Schoenfield's (2017) (cf. Seidenfeld et al. 2012; Mayo-Wilson and Wheeler 2016) proof that — assuming certain plausible desiderata for distance/inaccuracy measures — for every imprecise credal state ***cr***, there's a precise credence *cr* such that, relative to every possible vindicated function (Schoenfield doesn't take (non-extremal) chance hypotheses to be candidate vindicated functions — but her proof doesn't turn on this), the inaccuracy of *cr* is no greater than the inaccuracy of ***cr***. Though I won't seek to respond to this in detail here, it's worth pointing out that Carr (2015, 77) suggests that her argument — which I've adapted for my purposes — evades objections based on Schoenfield's proof because it doesn't invoke a single inaccuracy measure, but rather invokes both lower and upper distances. But even if it didn't, Konek (2019) points out that it's not obvious that we should reject accuracy-based arguments for imprecision rather than the putative desiderata that Schoenfield invokes. In any case, as a last resort (and given my disposition to accept a form of epistemic value pluralism — see Footnote 33), I could simply fall back on my more negative arguments for MushyP that involved disparaging the alternatives.

[61] Further reason to think this will be given below.

Elkin and Wheeler 2018; cf. Seidenfeld et al. 1989, esp. 241–2) that can be adapted in support of the view that one should respond to the probabilistic opinions of a group of experts in the way that MushyP implies that one should respond to learning a set-valued chance. And, as we've seen, alternative proposals like averaging have undesirable consequences (such as a failure to preserve independence).

Next, consider Schwarz's (2014) 'symmetry'-based argument. This requires a little more unpacking. The initial step that Schwarz takes is to provide an argument for why known *frequencies* constrain rational credence in something like the way *Cond* has chance constrain rational credence. He then argues that, where a frequency is unknown but the best system probability (which he supposes to be unique) is known, the best system probability constrains rational credence in that way because of the relationship between best system probabilities and frequencies.[62]

When it comes to frequencies, following de Finetti (1937), Schwarz (2014, 84–5) observes that if one's initial credal state is represented by a probability function that treats a sequence of outcomes as *exchangeable* — meaning, roughly, that for a set $\boldsymbol{T}$ of mutually exclusive and jointly exhaustive event types to which outcomes in that sequence belong (e.g. $\{heads, tails, edge\}$), all permutations of that sequence involving each of those event types occurring in the same proportions are equiprobable[63] — then it's a consequence of the probability calculus that one's credence that the i^{th} outcome in this sequence is an event of some type $T \in \boldsymbol{T}$ (e.g. *heads*), conditional upon (just) the fact that the frequency with which outcomes of type T occur in the sequence is x, is x. Of course, it's not *a priori* that some sequence is *exchangeable*. The most obvious way of justifying such an assumption is to appeal to some facts about the chances to which the sequence is subject (for instance, that the sequence is i.i.d.). So this doesn't give us a non-circular argument that frequencies play the chance role in guiding rational credence (cf. Hall 2004, 108).

Consequently, Schwarz (2014, 86–7, 92–4) suggests that, rather than appealing to exchangeability, we appeal to a certain sort of symmetry to rationalize conforming one's credences to the actual frequencies. Roughly speaking, the idea is that, if one is concerned with the i^{th} event in a sequence of outcomes and one knows that the frequency of outcomes of type T in that sequence is x, but one doesn't know in what position in the sequence i lies,[64] then it follows from what looks like a reasonable application of the PoI[65] that one's credence that i is of type T should be x.

When it comes to Best System probabilities, Schwarz's (2014, 96) idea is, roughly speaking, as follows. For reasons already described, it's reasonable to calibrate one's credences to actual frequencies, where these are known. But suppose that one is interested in whether some outcome o in a sequence is of type T, and one doesn't know the actual frequency f of outcomes of type T in that sequence, but one does know that x is the best system probability of a given outcome in that sequence being T. Then one can reasonably take x to be close f because 'mostly' or 'typically' a system's probabilities must be close to the actual frequencies if it's to be reasonably well-fitting (and thus a candidate for Best systemhood). Of course, x may deviate somewhat from f (after all, Best System probabilities are simplicity-constrained

[62] Of course, this suggests that, where the agent knows both the frequency and the best system probability and these diverge, she should calibrate her credence to the frequency. This is a point that I return to in the main text below.

[63] Strictly speaking, then, a sequence is not exchangeable or not *simpliciter*, but only exchangeable *relative to some such set* $\boldsymbol{T}$.

[64] This is commonly our situation with respect to chance events. For example, we might be interested in whether the decay of *this* Po-215 atom will be an instance of alpha- or beta-decay, but have very little idea in what position in (say) the time-ordered sequence of all Po-215 decay instances the decay of *this* one lies.

[65] Schwarz (2014, 93), acknowledging the concerns that some philosophers have about the PoI, describes this as a "very restricted" application and argues that it is less controversial than other applications. Of course, not everyone will accept even this limited application of the PoI (some imprecise probabilists appear to reject all applications). So this is a respect in which Schwarz's argument is not watertight.

as well as fit-constrained). But, for any (moderate) value of α, one has no more reason to suspect that $f = x - \alpha$ than to believe that $f = x + \alpha$. One's expectation of the frequency should therefore be equal to x and, if one's ignorant of where o lies in the relevant sequence, then one's credence that o is T should also be x.

Suppose we relax the assumption of a unique Best System. Then we might reason as follows. Each of the tied-for-best systems is constrained (via the desideratum of fit) to entail probabilities that, globally, don't deviate far from the actual frequencies. In ignorance of the actual frequencies, if one only knew the probabilities entailed by one of the tied-for-best systems, then Schwarz's (2014, 96) arguments could be deployed for taking the probabilities entailed by that system as one's expected frequencies and thus for calibrating one's credences to the probabilities entailed by *that* system. In other words, one is in the position of treating that system as though it were an expert concerning the frequencies. On the other hand, where one knows the *set* of probabilities entailed by the set of tied-for-best systems, then this is tantamount to being confronted with a *group* of experts' estimates of the actual frequencies and one can do no better than to adopt this set of estimates as one's credence. By the foregoing reasoning, then, we seem to arrive at a justification for thinking that sets of probabilities entailed by the tied-for-best systems play something like the MushyP role.

But one might object to this. For instance, one might think that the probabilities entailed by the tied-for-best systems are likely to be symmetrically distributed around the frequency. If so, it seems that one would do well to take the mean of the set of probabilities entailed by the systems as one's credence rather than the set of probabilities itself.

Yet this is implausible (and in particular, doesn't follow from Schwarz's reasoning that, for any given system, one has no more reason to suspect that the frequency is $x - \alpha$ than one does to believe that it's $x + \alpha$), as can be seen from the following example. Suppose that the frequency, throughout all of spacetime, of coin flips (of a certain physical character) that result in *heads* is 0.5. Suppose that it just so happens that the frequency is slightly elevated (0.515) among coin flips that occur on Mondays, and slightly elevated again among coin flips that occur on Monday afternoons (0.52). And suppose that one's interested in whether *this* coin flip, which happens to be taking place on a Monday afternoon, will land *heads*. If one knew the overall world frequency for *heads* among coin flips on Monday afternoons (and not just for coin flips on Mondays or coin flips in general) and one didn't have any more specific information, I take it that one should conform one's credence to *that* frequency.[66]

Now suppose that there are just three tied-for-best systems (perhaps the world of the example is a very simple one): One (which, *ceteris paribus*, is relatively simple but relatively non-well-fitting) entails a uniform 0.5 probability for coin flips resulting in *heads*; another (which, *ceteris paribus*, is middlingly complex and middlingly well-fitting) discriminates between days of the week but not times of day and assigns a uniform 0.515 probability for coin flips on Mondays resulting in *heads*; and the third (which, *ceteris paribus*, is relatively complex but relatively well-fitting) discriminates between days of the week *and* times of day and assigns a probability 0.52 for coin flips on Monday afternoons resulting in *heads*. Then taking the average (≈ 0.512) of the probabilities of *heads* for a Monday afternoon coin flip entailed by these three systems won't result in your having a credence that's closer to the actual frequency of *heads* on Monday afternoons than the probabilities entailed by all (or even most) of the tied-for-best systems.

Another problem with the averaging strategy is the following. As was indicated above, where one knows the *set* of probabilities entailed by the set of tied-for-best systems, this is tantamount to being confronted with a group of experts' estimates of the actual frequencies. Now it might be that, on each of these estimates, certain outcomes are independent under the actual frequencies (for instance, perhaps each of the systems described above takes coin flips on Monday afternoons to be probabilistically independent of one another). Yet, as

[66] See Hoefer (2007, 580–7), especially his discussion of 'Next-*n* Frequentism', for arguments supportive of this claim.

the discussion of Section 6.4 makes clear, an agent who simply averages the probabilities that each distribution entails is liable to end up with credences that take these outcomes as dependent, and could be induced to pay money for information about one such outcome when considering whether to accept a bet on another.

One might, however, think that there's an alternative strategy for an agent who knows the probabilities entailed by the tied-for-best systems that will ensure that her credences are as close as possible to the actual frequencies (relative to the smallest possible reference classes): namely, to adopt as her credence the probabilities entailed by the *most complex* system.[67] She can reason as follows: Only by being relatively better-fitting than the other systems can a relatively more complex system earn its membership of the set of tied-for-best systems. Even though the *overall* best fitting system may not best fit *every* frequency, unless she already knows the frequencies, such an agent might reason that the probabilities that the overall best fitting system entails will more typically be closer to the frequencies than the probabilities entailed by simpler systems. The worry is thus that a rational agent will respond to knowledge of the set of probabilities entailed by the tied-for-best systems by calibrating her credences to those entailed by the most complex system. This response is at odds with the way that rational credence responds to (set-valued) chance according to MushyP.

Before responding in earnest to this concern, note that, as we've seen, ties might not only arise because of the indeterminacy of the correct exchange rate between strength, simplicity, and fit, but also because simplicity (as well as strength and fit) is itself indeterminate. So, for at least some subset of the tied-for-best systems, there may not be any clear answer to the question of which is the simplest, and so the foregoing reasoning might not allow the agent to identify a unique probability to calibrate her credence with. Nevertheless, provided that ties result, in part, from indeterminacy concerning the correct exchange rate, the indeterminacy of simplicity needn't ensure that the agent's credences calibrate to the full range of probabilities entailed by the tied-for-best systems (because, by the foregoing reasoning, she'll prefer to go with those entailed by the determinately more complex systems).

It's worth noting that the present worry has an analogue which concerns the ability of best system probabilities to play the *Cond* role in the case where there's a unique winning system. The worry there is that a rational agent who has knowledge both of the (precise) best system probabilities and also of some better-fitting probabilities, such as the actual frequencies, will calibrate her credences to those better-fitting probabilities (see Hoefer 2007, 583). (The best system probabilities are liable to diverge from the actual frequencies because considerations of simplicity, and not just of fit, go into determining which system is *best*.) Indeed, as we've seen, Schwarz argues directly that it's rational to calibrate one's credences to the actual frequencies, and argues only that it's rational to calibrate one's credences to the best system probabilities *when one's ignorant of the actual frequencies* (but knowledgeable of the best system probabilities) and then *only because they equal one's expectations of the frequencies*. Schwarz's reasoning suggests that, given knowledge of both the actual frequencies and the best system probabilities, one ought to calibrate one's credences to the actual frequencies.

In general, the probabilities entailed by some probability function won't guide the credences of a rational agent who's aware of alternative probabilities that she knows to derive from a better-fitting probability function (unless, for some reason, she has reason to believe that the latter function is locally worse-fitting for the particular type of sequence she's concerned with). In the case of a unique winner of the best system competition, the probabilities entailed by the winner won't guide the credences of an agent who knows the actual frequencies. In the case of a set of tied-for-best systems, the probabilities entailed by the (determinately) simpler among these systems won't guide the credences of someone who knows the probabilities that are entailed by the (determinately) more complex of these systems *and* who knows both the relative complexity of the systems and that they're tied-for-

[67] Thanks to an anonymous referee for pressing a concern along these lines.

best. After all, such an agent can conclude that the latter systems are better-fitting by the reasoning given above.

One could doggedly maintain that the best system probabilities nevertheless perfectly play the chance role in guiding rational credence (as I've suggested is defined by *Cond* in the precise case and MushyP in the imprecise case), by arguing that rational agents only fail to calibrate their credences to the best system probabilities in these instances because they're in possession of *inadmissible* information (cf. Hoefer 2007, 583). However, this seems like a cheat for two reasons. First, one of the key advantages of a formulation of the chance-credence connection that appeals to conditional (rather than unconditional) chance — like *Cond* and MushyP — is that one shouldn't need a restriction to 'admissible' evidence, for reasons discussed by Hall (2004, esp. 101). Reintroducing admissibility to deal with the present objection seems *ad hoc*. Second, not only does it seem *ad hoc*, but it seems question-begging: Claiming that knowledge of the actual frequencies is inadmissible (in the sense that chance doesn't guide rational credence in the face of such knowledge) straightforwardly begs the question against actual frequentist accounts of chance.

I think the correct response is to admit that the Best System probabilities (whether precise or imprecise) are imperfect players of the chance role in guiding rational credence, but observe (*contra* Lewis 1980, 266) that the chance-credence connection doesn't exhaust the chance role (cf. Loewer 2001; Arntzenius and Hall 2003; Schaffer 2003; 2007) and to argue that the Best System probabilities earn the name 'chance' by playing other aspects of the chance role better than rival candidates. (Indeed, I take it that it's partly because they're persuaded of the latter point that most Humean philosophers nowadays think that the BSA provides a better interpretation of chance than actual frequentism.) For instance, it's plausibly one aspect of the chance role to be a lawfully-entailed magnitude (Schaffer 2007, 126). If one thinks on independent grounds that the BSA is the correct account of laws (as many Humeans do), then one's liable to think that the Best System probabilities are better players of *this* aspect of the chance role than the actual frequencies.[68] Indeed, if, as suggested above, in case of ties, the best system analyst ought to say that the laws are imprecise in the sense of entailing the sets of probabilities for outcomes yielded by the tied systems, then it's these set-valued chances that play this aspect of the chance role.

It's also important not to overstate the degree of imperfection with which the BSA probabilities (precise or imprecise) play that aspect of the chance role that involves guiding reasonable credence. We don't know exactly what the actual frequencies are relative to many reference classes, and never will, though we may be able to confine them (with higher probability!) within narrower intervals the more evidence we accumulate. Likewise, among the range of similar systems that entail SM-like probabilities, we don't know what the best-fitting are and never will, though we may be able to rule out more and more candidates (with higher probability!) the more evidence we accumulate. We have a sense of what the simpler systems are (systems that, for instance, draw upon uniform distributions over simply-characterizable regions of the universe's phase space) but, among the more complex ones, we often don't know which fit better and which fit worse than the simpler ones. For instance, because we don't know exactly where in its phase space the PICs of the universe are located, we don't know for sure which of the more complex of the Mentaculus-like systems concentrate probability in subregions in which they *are* located, as opposed to subregions in which they aren't.

If agents more rapidly accumulate *evidence* for the reasonably-well-fittingness of simpler systems than they do for the even-better-fittingness of those more complex systems that *are* better fitting, then this might explain why, relative to all but the extremely large amounts

[68] And if one thinks that there's some sort of frequency-tolerance platitude concerning chance (Armstrong 1983, 32; Loewer 2001, 613; Frigg and Hoefer 2015, 553–4) — namely that, although chances explain actual frequencies and actual frequencies are evidence for chances, actual frequencies may nevertheless diverge from the chances — one might also point out that, unlike the Best System probabilities (which exhibit moderate frequency-tolerance), the actual frequencies exhibit zero frequency-tolerance!

of evidence needed to establish the better-fittingness of more complex systems, simpler systems have an equal or greater claim on rational credence. Hoefer (2007, 583–7) and Hicks (2017) have suggested that something like this phenomenon may explain why the best system probabilities (in the case of a unique winner) are better guides to rational credence than the actual frequencies for an agent that lacks an extremely large amount of evidence (sufficient to determine what the actual frequencies are). Below, I'll return to the idea that, in the case of ties, this phenomenon might explain why the set of probabilities entailed by the tied-for-best systems constrain rational credence in something like the manner described by MushyP for all but agents with an extremely large amount of evidence. Before doing so, it will be helpful to discuss Hicks' (2017) argument that (in the precise case) the BSA probabilities play the chance role in guiding rational credence because it appeals to something like this phenomenon.

Hicks (2017) gives an epistemic utility argument that the best system probabilities play the chance role in guiding rational credence. Hicks first proposes a slightly revisionary interpretation of the theoretical desideratum of 'fit': namely, that the 'fit' of a system should be understood in terms of the distance of the probability function that it entails from the 'truth' function that assigns 1 to all truths and 0 to all falsehoods. (The argument would go through just as well if we took the 'vindicated' probabilities to be actual frequencies, thus recovering something closer to the orthodox notion of 'fit'.) Hicks then argues that — in the case of a unique Best System — because the Best System probabilities are constrained by the requirement of fit to be close to the truth function, an agent who knows the probabilities entailed by the Best System has an epistemic justification for minimizing the distance of her credences from these probabilities. In the case of ties between systems, 'minimizing the distance' of an agent's credence from a set of probabilities entailed by the tied-for-best systems might just mean respecting the principle 'Lower-Then-Upper-Dominance-Avoidance-Then-Total-Distance-Minimization' (see Section 6.4 above) with this set of probabilities taken as the vindicated ones.

The reason, according to Hicks (2017), that it's not the *best fitting* theory that constrains our credence is that we're typically in the position of having better evidence for the well-fittingness of simpler theories. For a simple example, suppose that one has observed an initial sequence of three flips of a coin, with the outcomes being $H_1T_2H_3$ (with the subscripts denoting the order in the sequence in which the outcome occurs). And suppose that one's interested in the outcomes of the next three flips. Let's suppose that the outcomes of these next three flips will in fact be $H_4T_5H_6$. So, as things pan out, (where the vindicated function is taken to be the truth function) the most accurate credence (conditional upon one's evidence) is one such that $cr(H_4|H_1T_2H_3) = cr(T_5|H_1T_2H_3) = cr(H_6|H_1T_2H_3) = 1$.

But now consider, on the one hand, a simple chance hypothesis that says that the coin has a uniform 0.6 chance of landing heads and, on the other, a more complex chance hypothesis that says that $ch(H_4|H_1T_2H_3) = ch(T_5|H_1T_2H_3) = ch(H_6|H_1T_2H_3) = 1$. The latter hypothesis is more complex because it assigns different chances to outcomes of the type *heads* (and *tails*) depending on their position in the sequence. (In this sense, a hypothesis that treats a sequence as i.i.d. — as the first hypothesis does — is, *ceteris paribus*, simpler than one that doesn't). But, as it turns out, an agent who conforms her credences to the latter hypothesis has the most accurate credences possible, where accuracy is construed as distance from the 'truth' function.[69]

[69] It's plausible — especially if one thinks that the aim-worthiness of frequencies is inherited from the aim-worthiness of truth — that such credences should also be taken as more accurate than credences conforming to the 0.6 hypothesis even if 'accuracy' is construed as distance from 'the' frequencies. After all, frequencies relative to narrower reference classes are often (at least) more aim-worthy than those relative to broader reference classes. For instance, take the overall frequency, among all atoms of all isotopes ever to exist, of decay within one year of coming into existence. This frequency seems less aim-worthy for rational credence than the specific frequencies for the various different isotopes. The frequencies we're presently considering the agent calibrating her credences to are the ones relative to reference classes comprising single elements, which might seem the most aim-worthy of all.

However the agent in question — before having witnessed the outcomes of tosses 4, 5, and 6 — doesn't have very good evidence for the well-fittingness of the more complex chance hypothesis and consequently would be irrational to calibrate her credences to it. Indeed, for an agent who updates via the Rule of Succession, this more complex chance hypothesis has an expected Brier Score (over the next three outcomes) of approximately 0.468, whereas the simpler chance hypothesis has an expected Brier Score of approximately 0.241 and so has lower expected inaccuracy according to a popular measure. As Hicks (2017, 939) puts the point:

> [W]e gain information about the chances from the frequencies. In order for the chances to be epistemically accessible to us, we need to be able to infer them from observation. And in order to observe them, we need a broad class of events whose outcomes are assigned the same chance. So simpler chance theories are more *epistemically accessible*.

In our example, the simple chance hypothesis treats all coin tosses in the sequence as belonging to the same class of events, hence the Rule of Succession can be applied in order to gain rational confidence that it has a high degree of accuracy. The more complex chance hypothesis treats tosses 4, 5, and 6 as *sui generis* events and so although, as it turns out, it's even more accurate than the simpler hypothesis, its accuracy is less amenable to confirmation. Before observing the tosses, we thus have a lower degree of rational confidence in its accuracy. It's not hypotheses with greater *accuracy* (as the more complex theory does in this case) that rational credence calibrates to, but rather hypotheses with greater (rationally) *expected accuracy* (as the simpler theory does in this case). Only for agents who have a large amount of evidence relative to the total possible evidence is the theory with the highest accuracy liable to coincide with the theory with the highest *expected* accuracy.

I think it's natural to integrate this thought with a Loewer-style 'reverse engineering' of the best system from the chance-credence connection. The idea would be to say that the 'best system' just is that system with the probability function that has the greatest expected accuracy. Since the expected accuracy of a system is relative to an evidential state (the complex chance hypothesis considered above has greater expected accuracy for an agent who has observed the outcomes of tosses 1–5 — and updates via the Rule of Succession — than it does for an agent who has observed tosses 1–3), the question naturally arises as to what evidential state the probability function must have greatest expected accuracy with respect to. The most natural answer is 'the sort of evidential state that we human beings, collectively, find ourselves in around about now' (in other words, the evidential state that comprises something like our current corpus of accumulated scientific knowledge). This re-introduces into the BSA some of the sort of anthropocentrism that Lewis was keen to avoid.

As I suggested in Section 4 above, I doubt that anthropocentrism is to be avoided entirely. For the Humean, what's fundamental is the mosaic. The laws and chances derive from systematizations of the mosaic, and it's doubtful that it's possible or even desirable to give a complete account of what counts as a good systematization without invoking features that are good-making *for creatures like us*. God, for instance, who knows everything that happens throughout the entire mosaic — and can easily hold all of this in His memory and recall it at will — would have no need for a simple systematization of it (cf. Albert 2012, 35–7). Moreover, the evidence-relativity of good systemhood just alluded to is unlikely to be too radical. For most reasonable confirmation functions (e.g. the Rule of Succession), where evidence is fairly significant yet still low relative to the total amount of possible evidence that could be acquired (as it is for most of our most serious candidates for lawhood and chancehood), the expected accuracy of chance hypotheses tends to vary relatively little with small increments of evidence (at least where the chance hypotheses entail non-extreme chances for that evidence).

On the present approach, given the number of similar and very good systems that entail SM probabilities, ties can reasonably be expected on this Loewerian interpretation. In light of the above reasoning, it appears that evidence of the well-fittingness of certain simpler systems accrues at a higher rate than evidence of the even-better-fittingness of more complex systems, which could well yield a tie in expected fit (relative to a reasonable confirmation function). But what's even more plausible is that there's no system that will come out as uniquely maximizing expected accuracy/fit under every reasonable confirmation function and every reasonable choice of priors. And since, on this picture, the correct standards of simplicity, strength, fit, and balance are treated as those that confer victory on the system that maximizes expected accuracy, the conclusion on this interpretation is that there are no uniquely correct standards and no unique winner.

If an agent knows the set of probabilities for some proposition A, conditional upon the remainder of her evidence E, that's entailed by a set of systems that's tied when it comes to the maximization of expected accuracy, then it's plausible that she ought to adopt that set of probabilities as her credence. For one thing, if she takes some alternative approach, like averaging, then she's liable to take some sequences as non-i.i.d. that all of the tied-for-best systems take as i.i.d.. Since treating sequences as i.i.d. is one component of simplicity in the relevant respect (and it's easier to gain evidence for the accuracy of hypotheses that — like the simpler hypothesis considered above — treat sequences as i.i.d.), averaging is liable to leave her with a credence distribution that determinately doesn't maximize expected accuracy.

8. Conclusion

It has been argued that, if we accept the BSA, then we have good reason to think that there are imprecise chances in our world. This is for the following reasons. *Firstly*, it's implausible that there's a single axiom system for our world that strikes the robustly best balance between the theoretical virtues of simplicity, strength, and fit. *Secondly*, it appears that the tied-for-best systems entail divergent (conditional) probabilities for certain outcomes: In particular, it has been argued that this is the case for thermodynamic outcomes (conditional upon earlier thermodynamic states). But, *thirdly*, the *sets of probabilities* entailed by the set of probability functions associated with the tied-for-best systems play the chance role in (for example) guiding rational credence in accordance with MushyP, and so such sets of probabilities constitute imprecise chances.

9. Acknowledgments

For helpful discussion and comments, I would like to thank Claus Beisbart, Seamus Bradley, Luke Elson, Alan Hájek, Jason Konek, Richard Pettigrew, Leszek Wroński, and three referees for this journal. I would also like to thank audiences at EPSA15, the *Formal Methods and Science in Philosophy* conference in Dubrovnik in 2017, the *Formal Epistemology Gathering* at the University of Bristol in 2017, the *Chance Encounter* workshop at the University of Groningen in 2016, the University of Leeds, the University of Cambridge, LSE, and the Institute of Philosophy in London. Special thanks to Radin Dardashti, Matthais Frisch, and Karim Thébault for many detailed discussions of the issues involved in this paper. I gratefully acknowledge the support of the Alexander von Humboldt Foundation during part of the time that I was writing this.

References

Albert, D. (2000). *Time and Chance*. Harvard University Press, Cambridge, MA.

Albert, D. (2012). Physics and chance. In Ben-Menahem, Y. and Hemmo, M., editors, *Probability in Physics*, pages 17–40. Springer, Berlin.

Allahverdyan, A. (2015). Imprecise probability for non-commuting observables. *New Journal of Physics*, 17:1–18.

Armstrong, D. (1983). *What is a Law of Nature?* Cambridge University Press, Cambridge.

Arntzenius, F. and Hall, N. (2003). On what we know about chance. *British Journal for the Philosophy of Science*, 54(2):171–179.

Beisbart, C. (2014). Good just isn't good enough — Humean chances and Boltzmannian statistical physics. In Galavotti, M., Dieks, D., Gonzalez, W., Hartmann, S., Uebel, T., and Weber, M., editors, *New Directions in the Philosophy of Science*, pages 511–529. Springer, Dordrecht.

Bradley, R. (2009). Revising incomplete attitudes. *Synthese*, 171:235–256.

Bradley, R. (2018). Learning from others: Conditioning versus averaging. *Theory and Decision*, 85(1):5–20.

Bradley, S. (2012). Dutch book arguments and imprecise probabilities. In Dieks, D., González, W., Hartmann, S., Stöltzner, M., and Weber, M., editors, *Probabilities, Laws, and Structures*, pages 3–17. Springer, New York.

Bradley, S. (2016). Imprecise probabilities. *Stanford Encyclopedia of Philosophy*, Winter 2016 Edition.

Bradley, S. (ms.). Criteria of adequacy for an imprecise decision theory. `http://www.seamusbradley.net/Papers/adequacy-decision.pdf`.

Bradley, S. and Steele, K. (2014). Should subjective probabilities be sharp? *Episteme*, 11(3):277–289.

Callender, C. (2011). The past histories of molecules. In Beisbart, C. and Hartmann, S., editors, *Probabilities in Physics*, pages 83–113. OUP, New York.

Callender, C. and Cohen, J. (2010). Special sciences, conspiracy and the better best system account of lawhood. *Erkenntnis*, 73(3):427–447.

Carr, J. (2015). Chancy accuracy and imprecise credence. *Philosophical Perspectives*, 29(1):67–81.

Chandler, J. (2014). Subjective probabilities need not be sharp. *Erkenntnis*, 79(6):1273–1286.

Cohen, J. and Callender, C. (2009). A better best system account of lawhood. *Philosophical Studies*, 145(1):1–34.

Couso, I., Moral, S., and Walley, P. (2000). A survey of concepts of independence for imprecise probabilities. *Risk, Decision, and Policy*, 5(2):165–181.

Cozman, F. (2012). Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600.

Dawid, A., DeGroot, M., and Mortera, J. (1995). Coherent combinations of experts' opinions. *Test*, 4(2):263–313.

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institute Henri Poincaré*, 7(1):1–68.

Dunn, J. (2011). Fried eggs, thermodynamics, and the special sciences. *British Journal for the Philosophy of Science*, 62(1):71–98.

Easwaran, K., Fenton-Glynn, L., Hitchcock, C., and Velasco, J. (2016). Updating on the credences of others: Disagreement, agreement, and synergy. *Philosophers' Imprint*, 16(11):1–39.

Elga, A. (2001). Statistical mechanics and the asymmetry of counterfactual dependence. *Philosophy of Science*, 68(S3):S313–S324.

Elga, A. (2004). Infinitesimal chances and the laws of nature. *Australasian Journal of Philosophy*, 82(1):67–76.

Elga, A. (2010). Subjective probabilities should be sharp. *Philosophers' Imprint*, 10(5):1–11.

Elga, A. (2012). Errata for 'subjective probabilities should be sharp'. Unpublished, `https://www.princeton.edu/~adame/papers/sharp/sharp-errata.pdf`.

Elkin, L. and Wheeler, G. (2018). Resolving peer disagreements through imprecise probabilities. *Noûs*, 52(2):260–278.

Feynman, R. (1987). Negative probability. In Hiley, B. and Peat, F., editors, *Quantum Implications: Essays in Honour of David Bohm*, pages 235–246. Routledge, London.

Fierens, P., Rêgo, L., and Fine, T. (2009). A frequentist understanding of sets of measures. *Journal of Statistical Planning and Inference*, 139(6):1879–1892.

Fine, T. (1974). Towards a revised probabilistic basis for quantum mechanics. *Synthese*, 29:187–201.

Fine, T. (1980). Remarks on the assessment, representation, aggregation, and utilization of expert opinion. Technical Report UCID - 18665.

Fine, T. (1988). Lower probability models for uncertainty and nondeterministic processes. *Journal of Statistical Planning and Inference*, 20(3):389–411.

Frigg, R. (2011). Why typicality does not explain the approach to equilibrium. In Suárez, M., editor, *Probabilities, Causes, and Propensities in Physics*, pages 77–93. Springer, Dordtrecht.

Frigg, R., Bradley, S., Du, H., and Smith, L. (2014). Laplace's demon and the adventures of his apprentices. *Philosophy of Science*, 81(1):31–59.

Frigg, R. and Hoefer, C. (2015). The best Humean system for statistical mechanics. *Erkenntnis*, 80(S3):551–574.

Frisch, M. (2014). Why physics can't explain everything. In Wilson, A., editor, *Chance and Temporal Asymmetry*. OUP, Oxford.

Galvan, B. (2008). Quantum mechanics and imprecise probability. *Journal of Statistical Physics*, 131(6):1155–1167.

Gärdenfors, P. and Sahlin, N.-E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese*, 53(3):361–386.

Grize, Y. and Fine, T. (1987). Continuous lower probability-based models for stationary processes with bounded and divergent time averages. *The Annals of Probability*, 15(2):783–803.

Grove, A. and Halpern, J. (1998). Updating sets of probabilities. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 173–182. Association for Uncertainty in Artificial Intelligence.

Hájek, A. (2003a). Conditional probability is the very guide of life. In Kyburg Jr., H. and Thalos, M., editors, *Probability is the Very Guide of Life*, pages 183–203. Open Court, Chicago.

Hájek, A. (2003b). What conditional probability could not be. *Synthese*, 137(3):273–323.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, 156(3):563–585.

Hájek, A. (ms. a). A puzzle about partial belief.

Hájek, A. (ms. b). Staying regular.

Hájek, A. and Smithson, M. (2012). Rationality and indeterminate probabilities. *Synthese*, 187:33–48.

Hall, N. (2004). Two mistakes about credence and chance. *Australasian Journal of Philosophy*, 82(1):93–111.

Hart, C. and Titelbaum, M. (2015). Intuitive dilation? *Thought*, 4(4):252–262.

Hartmann, S. (2015). Imprecise probabilities in quantum mechanics. In Crangle, C., García de la Sienra, A., and Longino, H., editors, *Foundations and Methods from Mathematics to Neuroscience: Essays Inspired by Patrick Suppes*. CSLI Publications, Stanford, CA.

Hartmann, S. and Suppes, P. (2010). Entanglement, upper probabilities and decoherence in quantum mechanics. In Suárez, M., Dorato, M., and Rédei, M., editors, *EPSA Philosophical Issues in the Sciences*, pages 93–103. Springer, Dordrecht.

Hicks, M. (2017). Making fit fit. *Philosophy of Science*, 84(5):931–943.

Hoefer, C. (2007). The third way on objective probability: A sceptic's guide to objective chance. *Mind*, 116(463):549–596.

Hughes, R. and van Fraassen, B. (1984). Symmetry arguments in probability kinematics. *Philosophy of Science*, 1984(2):851–869.

Ismael, J. (2008). Raid! Dissolving the big, bad bug. *Noûs*, 42(2):292–307.

Jeffrey, R. (1987). Indefinite probability judgment: A reply to Levi. *Philosophy of Science*, 54(4):586–591.

Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603.

Joyce, J. (2005). How probabilities reflect evidence. *Philosophical Perspectives*, 19:153–178.

Joyce, J. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, 24:281–323.

Kaplan, M. (1983). Decision theory as philosophy. *Philosophy of Science*, 50(4):549–577.

Konek, J. (2019). Epistemic conservativity and imprecise credence. Forthcoming in *Philosophy and Phenomenological Research*.

Kozine, I. and Utkin, L. (2002). Processing unreliable judgements with an imprecise hierarchical model. *Risk, Decision and Policy*, 7(3):325–339.

Kumar, A. and Fine, T. (1985). Stationary lower probabilities and unstable averages. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 69(1):1–17.

Kyburg Jr., H. and Pittarelli, M. (1992). Some problems for convex Bayesians. In *Proceedings of the Eighth International Conference on Uncertainty in Artificial Intelligence*, pages 149–154. Morgan Kaufmann, San Francisco, CA.

Kyburg Jr., H. and Pittarelli, M. (1996). Set-based Bayesianism. *IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans*, 26(3):324–339.

Laddaga, R. (1977). Lehrer and the consensus proposal. *Synthese*, 36:473–477.

Levi, I. (1974). On indeterminate probabilities. *Journal of Philosophy*, 71(13):391–418.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, MA.

Levi, I. (1982). Conflict and social agency. *Journal of Philosophy*, 79(5):231–247.

Levi, I. (1985). Imprecision and indeterminacy in probability judgment. *Philosophy of Science*, 52(3):390–409.

Levi, I. (1999). Value commitments, value conflict, and the separability of belief and value. *Philosophy of Science*, 66(4):509–533.

Lewis, D. (1980). A subjectivist's guide to objective chance. In Jeffrey, R., editor, *Studies in Inductive Logic and Probability*, volume 2, pages 267–297. University of California Press, Berkeley & Los Angeles.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4):343–377.

Lewis, D. (1986). *Philosophical Papers*, volume 2. OUP, New York.

Lewis, D. (1994). Humean supervenience debugged. *Mind*, 103(412):473–490.

Lewis, D. (1999). Why conditionalize? In Lewis, D., editor, *Papers in Metaphysics and Epistemology*, volume 2, pages 403–407. Cambridge University Press, New York.

Loewer, B. (1996). Humean supervenience. *Philosophical Topics*, 24(1):101–127.

Loewer, B. (2001). Determinism and chance. *Studies in History and Philosophy of Science Part B*, 32(4):609–620.

Loewer, B. (2004). David Lewis's Humean theory of objective chance. *Philosophy of Science*, 71(5):1115–1125.

Loewer, B. (2007). Counterfactuals and the second law. In Price, H. and Corry, R., editors, *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, pages 293–326. OUP, Oxford.

Loewer, B. (2008). Why there is anything except physics. In Hohwy, J. and Kallestrup, J., editors, *Being Reduced: New Essays on Reduction, Explanation, and Causation*, pages 149–163. OUP, Oxford.

Loewer, B. (2012a). The emergence of time's arrows and special science laws from physics. *Interface Focus*, 2(1):13–19.

Loewer, B. (2012b). Two accounts of laws and time. *Philosophical Studies*, 160(1):115–137.

Loewer, B. and Laddaga, R. (1985). Destroying the consensus. *Synthese*, 62:79–95.

Maudlin, T. (2007). What could be objective about probabilities? *Studies In History and Philosophy of Modern Physics*, 38(2):275–291.

Maudlin, T. (2011). Three roads to objective probability. In Beisbart, C. and Hartmann, S., editors, *Probabilities in Physics*, pages 293–319. OUP, New York.

Mayo-Wilson, C. and Wheeler, G. (2016). Scoring imprecise credences: A mildly immodest proposal. *Philosophy and Phenomenological Research*, 93(1):55–78.

Pedersen, A. and Wheeler, G. (2014). Demystifying dilation. *Erkenntnis*, 79(6):1305–1342.

Pettigrew, R. (2012). Accuracy, chance, and the principal principle. *Philosophical Review*, 121(2):241–275.

Pettigrew, R. (2013). A new epistemic utility argument for the principal principle. *Episteme*, 10(1):19–35.

Popper, K. (1972). *The Logic of Scientific Discovery*. Hutchinson, London.

Ramsey, F. (1931). Truth and probability. In Braithwaite, R., editor, *The Foundations of Mathematics and Other Logical Essays*, pages 156–198. Routledge & Kegan Paul, London.

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

Sahlin, N.-E. and Weirich, P. (2014). Unsharp sharpness. *Theoria*, 80(1):100–103.

Schaffer, J. (2003). Principled chances. *British Journal for the Philosophy of Science*, 54(1):27–41.

Schaffer, J. (2007). Deterministic chance? *British Journal for the Philosophy of Science*, 58(2):113–140.

Schoenfield, M. (2017). The accuracy and rationality of imprecise credences. *Noûs*, 51(4):667–685.

Schrenk, M. (2008). A Lewisian theory for special science laws. In Walter, S. and Bohse, H., editors, *Ausgewählte Beiträge zu den Sektionen der GAP*, volume 6, pages 121–131. Mentis, Paderborn.

Schwarz, W. (2014). Proving the principal principle. In Wilson, A., editor, *Chance and Temporal Asymmetry*, pages 81–99. OUP, Oxford.

Seidenfeld, T. (1994). When normal and extensive form decisions differ. In Prawitz, D., Skyrms, B., and Westerståhl, editors, *Logic, Methodology and Philosophy of Science IX*, pages 451–463. Elsevier, Amsterdam.

Seidenfeld, T., Kadane, J., and Schervish, M. (1989). On the shared preferences of two bayesian decision makers. *Journal of Philosophy*, 86(5):225–244.

Seidenfeld, T., Schervish, M., and Kadane, J. (2012). Forecasting with imprecise probabilities. *International Journal of Approximate Reasoning*, 53(8):1248–1261.

Seidenfeld, T. and Wasserman, L. (1993). Dilation for sets of probabilities. *The Annals of Statistics*, 21(3):1139–1154.

Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, 42:139–165.

Sturgeon, S. (2010). Confidence and coarse-grained attitudes. *Oxford Studies in Epistemology*, 3:126–149.

Suppes, P. and Zanotti, M. (1991). Existence of hidden variables having only upper probabilities. *Foundations of Physics*, 21(12):1479–1499.

Teller, P. (1973). Conditionalization and observation. *Synthese*, 26:218–258.

Troffaes, M. (2006). Generalizing the conjunction rule for aggregating conflicting expert opinions. *International Journal of Intelligent Systems*, 21(3):361–380.

Troffaes, M. (2007). Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29.

van Fraassen, B. (1985). Empiricism in the philosophy of science. In Churchland, P. M. and Hooker, C. A., editors, *Images of Science: Essays on Realism and Empiricism*, pages 245–308. University of Chicago Press, Chicago.

Wagner, C. (1985). On the formal properties of weighted averaging as a method of aggregation. *Synthese*, 62:97–108.

Wagner, C. (2011). Peer disagreement and independence preservation. *Erkenntnis*, 74(2):277–288.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Walley, P. and Fine, T. (1982). Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10(3):741–761.

Wheeler, G. (2014). Character matching and the Locke pocket of belief. In Lihoreau, F. and Rebushi, M., editors, *Epistemology, Context, and Formalism*, pages 187–195. Springer, Cham.

Wheeler, G. and Williamson, J. (2011). Evidential probability and objective Bayesian epistemology. In *Handbook of the Philosophy of Science. Volume 7: Philosophy of Statistics*, pages 307–331. Elsevier, Amsterdam.

White, R. (2010). Evidential symmetry and mushy credence. *Oxford Studies in Epistemology*, 3:161–186.

Woodward, J. (2014). Simplicity in the best systems account of laws of nature. *British Journal for the Philosophy of Science*, 65(1):91–123.