UNSHARP BEST SYSTEM CHANCES

LUKE FENTON-GLYNN

DEPARTMENT OF PHILOSOPHY, UNIVERSITY COLLEGE LONDON,

GOWER STREET, LONDON, WC1E 6BT, U.K.

Abstract. Much recent philosophical attention has been devoted to variants on the Best System Analysis of laws and chance. In particular, philosophers have been interested in the prospects of such Best System Analyses (BSAs) for yielding *high-level* laws and chances. Nevertheless, a foundational worry about BSAs lurks: there do not appear to be uniquely appropriate measures of the degree to which a system exhibits theoretical virtues, such as simplicity and strength. Nor does there appear to be a uniquely correct exchange rate at which the theoretical virtues of simplicity, strength, and likelihood (or *fit*) trade off against one another in the determination of a best system. Moreover, it may be that there is no *robustly* best system: no system that comes out best under *any* reasonable measures of the theoretical virtues and exchange rate between them. This worry has been noted by several philosophers, with some arguing that there is indeed plausibly a set of tied-for-best systems for our world (specifically, a set of very good systems, but no robustly *best* system). Some have even argued that this entails that there are no Best System laws or chances in our world. I argue that, while it *is* plausible that there is a set of tied-for-best systems for our world, it doesn't follow from this that there are no Best System chances. (As I will argue, the situation with regard to laws is more complex.) Rather, it follows that (some of) the Best System chances for our world are *unsharp*.

Contents

1. Introduction

Much recent philosophical attention has been devoted to variants on the Best System Analysis of laws and chance. In particular, philosophers have been interested in the prospects of such Best System Analyses (BSAs) for yielding *high-level* laws and chances. Nevertheless, a foundational worry about BSAs lurks: there do not appear to be uniquely appropriate measures of the degree to which a system exhibits theoretical virtues, such as simplicity and strength. Nor does there appear to be a uniquely correct exchange rate at which the theoretical virtues of simplicity, strength, and likelihood (or *fit*) trade off against one another in the determination of a best system. Further, there may be no *robustly* best system: that is, no system that comes out best under *any* reasonable measures of the theoretical virtues and exchange rate between them. This worry has been noted by several philosophers, with some arguing that there is indeed plausibly a set of tied-for-best systems for our world (specifically, a set of very good systems, but no robustly *best* system). Some have even argued that this entails that there are no Best System laws or chances in our world.

In what follows, I will argue that plausibly there is a set of tied-for-best systems for our world, but that it doesn't follow from this that there are no Best System chances. Rather, it follows that (some of) the Best System chances for our world are *unsharp*. When it comes to laws the situation is somewhat more subtle. But I will argue that the existence of ties doesn't imply that there are no Best System laws either.

The plan is as follows. In Section 2, I outline the Best System Analysis (BSA) of laws and chance. 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 probabilistic principles of *statistical mechanics* (SM). Specifically, in that section, I examine so-called 'Globalist' approaches to axiomatizing SM. In Section 4, I describe arguments that there is no clearly best 'Globalist' system for SM. In Section 5, I argue that, contrary to what has sometimes been thought (and even if one assumes the Globalist approach to be the correct approach to axiomatizing SM), this does not show that there are no Best System chances in SM. Rather, it shows that the Best System chances are unsharp. In Section 6, I turn my attention to so-called 'Localist' approaches to SM, and argue that a similar conclusion follows in this context. In Section 7, I seek to further justify the conclusion that the BSA yields unsharp chances for our world by examining the chance-credence connection. In Section 8, I examine the so-called 'Better Best System Analysis' (BBSA) of laws and chance, which aims to capture the role of laws and chances in the special sciences. I argue that, in the light of actual attempts to systematize certain special sciences, it is plausible that the BBSA yields non-sharp special science chances.

2. The Best System Analysis (BSA) of Laws and Chance

According to the BSA, which received its most significant development by Lewis (1983, 1994), the laws are the theorems entailed by that set of axioms which best systematizes the entire history of the world. The objective chances are those probabilities that are entailed by this best system. Lewis (1994) combines the BSA with the (logically distinct) thesis of *Humean Supervenience* (HS): the thesis that the laws and chances supervene upon the *Humean mosaic*, that is, upon the distribution of categorical (i.e. non-modal) properties throughout all of space-time. Specifically, 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. While I find the thesis of HS plausible, the BSA's plausibility is not tied to that of HS. In what follows, however, I shall assume HS since most discussion of the BSA has been conducted against the background of this assumption. Nevertheless, as far as I can see, my claim that the BSA yields unsharp chances need not turn upon this assumption.

The goodness with which a candidate system systematizes the Humean mosaic is judged against the theoretical virtues of simplicity, strength, and likelihood (or *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 (e.g. by adding axioms, or making them more complicated), while simplicity can often be gained at the expense of strength (e.g. by removing axioms or simplifying them).

The Best System mustn't be dominated by any other system: that is, it mustn't be the case that it exhibits some theoretical virtue to a lower degree than another system that exhibits all other virtues in at least as high a degree. But, among the non-dominated systems, some systems will strike a better *balance* between the virtues than others. For example, if there is a system that achieves great strength with relatively few, simple axioms, then this system is to be judged superior to one that achieves maximal simplicity at the cost of minimal strength, by saying nothing at all (even though the latter system is not dominated by the former). The Best System is therefore that non-dominated system that strikes the best balance between the theoretical virtues, where (for example) neither strength nor simplicity receives zero weight in the exchange rate that determines the best balance between the virtues.

If the world's history is a certain way (if it contains a lot of stochastic-looking events), then a candidate system may achieve a good deal of strength with little cost in simplicity by being endowed with a probability function (cp. Loewer (2004, 1119)): that is, a function $Ch_t(p)$ that maps proposition and time pairs $\langle p, t \rangle$ onto real values in the $[0, 1]$ interval, and that obeys the axioms of probability. Alternatively, and I think rather plausibly, one might take conditional probability as basic. One might, for instance, prefer the Renyí-Popper axiomatisation of probability to the Kolmogorov axiomatisation (Hájek 2003a,b, 2007). In that case, one can take candidate systems to come endowed with a Renyí-Popper function $Ch(p|q)$ that maps proposition pairs $\langle p, q\rangle$ onto the reals in the $[0, 1]$ interval. If one takes conditional chance as basic in this way, then it is unnecessary to include a 'time' index to the chance function. Rather $Ch_t(p)$ can be analyzed as the chance for p that comes by conditioning upon H_t: the history of the world up to t (see Hoefer (2007, 562–565); and Glynn (2010, 78–79)). That is, $Ch_t(p) =_{def} Ch(p|H_t)$.

If one takes conditional chances to be basic, one may allow that there are chances for p conditional upon propositions less informative than one giving the complete history of the world up to some time t.[1] For example, there may be chances for p conditional upon just the *macro*-history of the world through t, H_t^M (see Loewer (2001)). There may also be chances for p conditional upon other sorts of proposition: for example, those specifying some localized chance setup (Hoefer (2007), Frigg and Hoefer (2013)). Indeed, perhaps we can get chances for p by conditioning upon just about any arbitrary proposition (Albert 2000, 2012, Ismael 2012).

Where a system comes endowed with a probability function, it may exhibit a third theoretical desideratum (besides simplicity and strength) to a greater or lesser degree: namely *likelihood* or *fit*. A system *fits* the actual course of history well (has higher likelihood) to the extent that the associated probability function assigns a higher probability to the actual course of history: the higher the probability, the better the fit (Lewis 1994, 480).[2] Where we take conditional probability as basic, we need a notion of fit according to which a system fits better to the extent that it assigns a higher probability to the actual course of history *conditional upon* initial conditions. If we're to make sense of chances in statistical physics and the special sciences, then we should also regard relatively high probability assignments

[1]There may be well-defined chances conditional upon propositions specifying the complete state of the world at a time (or on a Cauchy surface – cp. Maudlin (2007a, 18-20)). However, assuming that the dynamical laws of the world are Markovian, this is equivalent to conditioning upon the entire history of the world up to and including that time (or up to and including that Cauchy surface).

[2]This notion of fit applies only if there are only finitely many chance events. It also doesn't work well when systems may incorporate continuous chance distributions (Frigg and Hoefer 2013): that is, chance distributions over infinite sets. See Elga (2004) for an extension of the notion of fit to infinite cases.

conditional upon *coarse grainings* of those initial conditions (that is, conditional upon the initial macro-state, and not just the initial micro-state) as contributing to a system's fit.

The Best System is that which strikes the best balance between the theoretical virtues of simplicity, strength, and fit. According to the BSA, the probability function associated with the Best System is the *chance function* for the world.

The idea, then, is that the Humean mosaic, together with the theoretical virtues, serves to fix a Best System. As Lewis (1994, 480) puts it: "The arrangement of qualities provides the candidate . . . systems, and considerations of simplicity and strength [and fit] and balance do the rest". Or, more concisely, chances are "Humean [Best System]-supervenient on [the Humean mosaic]" (Frigg and Hoefer 2013).

3. The BSA and Statistical Mechanics

Lewis himself appears to have thought that the probability function associated with the best system for our world would simply be the fundamental physical probability function: that is, the function that yields all and only the probabilities entailed by quantum mechanics, or whatever fundamental physical theory replaces it (see Lewis (1986, 118); Lewis (1994)).

Yet Loewer (2001, 2007, 2008, 2012a,b) has influentially argued that the probabilities of statistical mechanics (SM) are also entailed by the best system for our world, and therefore are genuine objective chances according to the BSA. Loewer appeals to the axiomatisation of SM described by Albert (2000, Chs. 3-4). Albert suggests that SM can be derived from the following:

(FD) the fundamental dynamical laws;

(PH) a proposition characterising 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 regions of microphysical phase space associated with that low-entropy state.

Albert (2012) and Loewer (2012a,b) dub the conjunction of FD, PH, and SP 'the Mentaculus'.[3]

The argument that the SM probabilities are derivable from the Mentaculus goes roughly as follows. Consider the region of microphysical phase space associated with the low-entropy initial state of the universe implied 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 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 is not itself correlated with its initial micro state being entropy-decreasing, it is extremely likely that any such subsystem that is in initial disequilibrium will increase in entropy over time (see Loewer (2007, 302); Loewer (2012a, 124–125); Loewer (2012b, 17); and Albert (2000, 81-85)).[4]

Albert (2000, 2012) and Loewer (2007, 2008, 2012a,b) have argued that the Mentaculus entails many of the probabilities of the special sciences.[5] Loewer thus claims that the Mentaculus is much stronger than a system consisting of the fundamental dynamical laws, FD, alone. And since it is not much more complicated (it only requires the addition of the axioms PH and SP), Loewer claims that it is a plausible *best* system for our world.[6]

In fact a minor modification to the BSA, as articulated by Lewis, is required if the Mentaculus is even to be a candidate Best System for the world. It has often been observed that

[3]Note that, where the fundamental dynamics are quantum rather than classical, the uniform probability distribution – which in the classical case is given by the Statistical Postulate, SP – is not over classical phase space, but rather over the set of quantum states compatible with the PH; or at least this is the standard construal of how we recover the statistical mechanical probabilities in a quantum mechanical system (cp. Albert (2000, 131-133)).

[4]See Winsberg (2004), Earman (2006), and Callender (2011) for criticisms of this line of argument. Such criticisms will be further discussed in sections 6 & 8 below.

[5]Though I won't go into the details (such as they are) here, the idea is roughly that this is because many of the special sciences are themselves concerned with entropy-increasing processes.

[6]This proposal requires that facts about initial conditions, such as PH, are potential axioms of the Best System. The BSA has not always been construed as allowing for this. However, Lewis (1983, 367) himself takes the view that facts about initial conditions may well be among the axioms of the Best System. See also Maudlin (2007b, 280-281) and Maudlin (2011, 303).

the simplicity of a system is relative to the vocabulary in which it is expressed. Lewis (1983, 367-368) took this as a reason to restrict candidates for Best Systemhood to those systems whose axioms refer only to perfectly natural properties. But, as Schaffer (2007, 130) points out (see also Cohen and Callender (2009) and Callender and Cohen (2010)), the Mentaculus contains predicates like 'low entropy' that correspond to properties that are not perfectly natural, and so doesn't even seem to be a candidate Best System. One alternative to Lewis's approach – pursued by Callender (2011), Cohen and Callender (2009), Callender and Cohen (2010), Dunn (2011), and Schrenk (2008) – which will be discussed further in Section 8, is to argue that Best Systemhood is vocabulary-relative.

Yet it is not necessary to take Best Systemhood to be vocabulary relative, nor to make an *a priori* choice of some uniquely privileged vocabulary (such as the vocabulary of the perfectly natural kinds) in order to allow that the Mentaculus is a candidate Best System.[7] An alternative is to adopt the following, slight modification of the BSA (for further alternatives, see Frisch (2013) and Frigg and Hoefer (2013)). Observe that, as Lewis (1983, 368) recognises, naturalness admits of degrees. We may thus take naturalness of the predicates that it employs to be a theoretical virtue, to be weighed alongside the simplicity, strength, and fit of a system. 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 is a plausible best system.

4. Ties Between Systems

Lewis acknowledged that the BSA is not completely unproblematic. "The worst problem about the best-system analysis" (Lewis 1994, 479) is that notions such as simplicity and balance are to some extent *imprecise*. There is, for example, no unique and maximally determinate simplicity metric that is obviously the correct one to apply to candidate systems,[8] nor is there a unique and maximally determinate correct exchange rate between the competing virtues of simplicity, strength, and fit. The worry is that, within acceptable ranges,

[7]For the problems associated with making a once-and-for-all choice of a privileged vocabulary, see Cohen and Callender (2009, 11-20).

[8]Indeed the notion of *strength* is also difficult to render precise (see e.g. Woodward (2005, 288-290)).

different precisifications of the simplicity metric and of the exchange rate between the virtues will result in different verdicts about which system counts as *best*. Lewis has little more to offer than the hope that this will turn out not to be so:

> "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. Perhaps we presuppose it in our thinking about law. I can admit that *if* nature were unkind, and *if* disagreeing rival systems were running neck-and-neck, then . . . the theorems of the barely-best system would not very well deserve the name of laws. But I'd blame the trouble on unkind nature, not on the analysis; and I suggest we not cross these bridges unless we come to them." (Lewis 1994, 479; italics original)

Lewis adds parenthetically:

> "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. But what of it? We haven't the slightest reason to think the case really arises." (Ibid.)

One *might* think that the same thing that Lewis says about laws should be said of chances: if there is no clear winner of the best system competition, then there would be nothing deserving of the name *chance*. While Lewis himself does not explicitly say this, it is a conclusion that is drawn by Beisbart (2013).

The trouble is that it has seemed to a number of authors that, particularly when we consider axioms systems that entail SM, there is *indeed* a set of rival systems – each of which is associated with a different probability function – that are running neck-and-neck (at least within the limits of precision of notions such as simplicity and balance) in the best system competition for our world. The Best System analyst might therefore be hard pressed to avoid the conclusion that there simply aren't things deserving of the names of *laws* and *chances* in our world (see Beisbart, ibid.).

Specifically, let us assume that the Mentaculus does indeed entail the SM probabilities, and let us assume that Loewer is correct that it constitutes a better system for our world than one comprising the fundamental dynamical laws alone. If the Mentaculus comes out *best*, then the SM probabilities will count as objective chances on the BSA, and any generalizations (probabilistic or otherwise) that it entails will count as laws on the BSA.

But an axiom system consisting of *only* the fundamental dynamic laws (FD) is not the only rival to the Mentaculus. Schaffer (2007, 130–132), Hoefer (2007, 560), and Beisbart (2013) consider another candidate, which consists of the fundamental dynamic laws plus an axiom giving the *precise initial conditions* (PICs) of the universe. This is a very strong system. Schaffer (2007, 131-132) suggests that it is maximally strong, while Hoefer (2007, 560) questions this. Hoefer (ibid.) points out that it's not obvious how to quantify the complexity of the two candidate systems (i.e. the Mentaculus and the FD + PICs system) in such a way as to allow comparison. It is also not obvious how to decide whether any difference in complexity is adequately compensated for by a resulting a difference in strength.

As Beisbart (2013) suggests, there are still further rivals to the Mentaculus:

> "We can improve fit when we . . . assume a flat probability distribution over a certain sub-region of the past low-entropy macro-state [as opposed to over the whole of the past low-entropy macro-state – as per (SP) of the Mentaculus]. That sub-region may be defined by the demand that a certain elementary particle has a kinetic energy larger than a particular value e_0, for instance.

> If we do so, we have to pay in simplicity though because [in addition to the assumed low-entropy initial state, we have to further specify that the initial] kinetic energy of a particular particle be e_0.
>
> So overall, would we improve the system ...? This is a difficult question, and the answer is far from clear. The only thing we can say is that fit is considerably improved, but that there is a considerable cost in simplicity too. So it's not a case in which the right sort of balance favours one system rather than the other in a clear way."

Beisbart's worry is that, by choosing different sub-regions of the phase space associated with PH to apply the uniform probability distribution to, we get a range of candidate best systems (cp. also Schaffer (2007, 131n)). At one extreme, we have the Mentaculus (where a uniform distribution is applied to the whole region of phase space compatible with PH); at the other extreme, we have a system comprising the fundamental dynamic laws together with the PICs. The latter is equivalent to what we get in the limit as we apply a uniform distribution to smaller and smaller sub-regions of the phase space associated with the PH, each of which contains the point-sized region of phase space that the universe actually initially occupied. The former is relatively simple, but gives an inferior fit; the latter gives a better fit, but is less simple. In between we have a continuum of systems involving the application of the uniform distribution to progressively smaller sub-regions of the phase-space compatible with PH (where each sub-region contains the actual point in phase space at which our universe was initially located). Such systems are increasingly better fitting, since they assign an increasingly high probability to the actual macroscopic course of events,[9] but also increasingly complex, since picking out progressively smaller sub-regions requires building into the axioms an increasing amount of information about the actual initial state of the universe.

[9]The claim that this is so will be argued for in detail at the end of this section.

To be more precise, consider again the Mentaculus, i.e. the conjunction of FD, PH, and SP. Holding fixed FD, as defined above, there are three further things that need to be specified in order to give us SM-like regularities:[10] a measure μ over 6N-dimensional microphysical phase space, a phase space region Γ within which the initial micro-state of the universe is located (the relevant region Γ may be picked out by specifying that the universe was initially in some or other macro-state, M), and a probability distribution, P, over Γ. The Mentaculus takes the relevant measure to be the Lebesgue measure, λ; it takes the relevant region of phase space to be the region Γ_0 that is associated with the special low entropy macro-state, M_0, described by PH; and it takes the relevant probability distribution, P, over Γ_0 to be the distribution P_U^λ that is uniform with respect to the Lebesgue measure λ.

The systems that appear to be tied for best with the Mentaculus (at least given the limits of precision of notions like simplicity and balance) vary *either* the measure over phase space, *or* the size of the initial low-entropy region, *or* the probability distribution. In fact, the probability distribution isn't really specifiable independently of the measure. For instance, it only makes sense to talk about a *uniform* distribution when we specify *which measure the distribution is uniform with respect to.* The probability distribution-measure pair, $\langle P^\mu, \mu\rangle$, must be specified together, with the superscript to P indicating that the distribution is to be understood with respect to the measure μ. The Mentaculus takes the relevant distribution to be that which is uniform *with respect to the Lebesgue measure.* That is, it takes the relevant probability distribution-measure pair to be $\langle P_U^\lambda, \lambda\rangle$. In what follows, I will sometimes talk simply about a probability distribution, suppressing reference to the measure. Where I do so, I should be understood as assuming the Lebesgue measure.

Beisbart, as discussed, considers a rival to the Mentaculus which involves applying P_U^λ (i.e. the distribution that is uniform on the Lebesgue measure) to a sub-region Γ_B of the phase space region Γ_0 associated with the low entropy macro-state M_0 specified by the Mentaculus. This sub-region Γ_B is defined by adding to the requirement that the region be associated with

[10] At least this is so on the standard assumption that the SM regularities cannot be derived from the fundamental dynamics alone. The possibility that they *can* be so derived is one that is investigated by Albert (2000, 150-162) (cp. also Albert (2012, 39-40)).

a low-entropy macro-state, the specification that one of the elementary particles constituting the initial condition of the universe has kinetic energy above a certain value. As discussed above, one can consider a continuum of sub-regions of Γ_0, given by specifying more and more of the microphysical details of the universe's initial state. As we move along this continuum, we get better and better fit at the cost of less and less simplicity up to the extreme case where we specify the PICs of the universe, corresponding to the point-sized region of phase space, Γ_δ.

One can also consider a variety of probability distributions, P (or, more accurately, probability distribution-measure pairs, $\langle P^\mu, \mu \rangle$) over the initial phase space region. In many cases, systems that vary the initial phase space region, Γ, will be equivalent to systems that vary the probability distribution, P. Consider for example the suggestion, made by Frigg and Hoefer (2013), that among the competitors to the Mentaculus is an axiom system that – instead of the Mentaculus's SP – contains an axiom that specifies "a peaked distribution, nearly Dirac-delta style" – call it P_δ – whose peak is at the precise (point-sized) initial condition represented by Γ_δ. The delta distribution assigns probability 1 to the point at which it is peaked (in this case, the precise actual microphysical initial condition of the universe), and 0 to all regions that don't include this point (in this case, regions of phase space that don't include the actual initial condition of the universe).[11] Such a system is effectively equivalent to what we get in the limit as we apply the distribution P_U^λ to smaller and smaller phase space regions that contain the point Γ_δ.[12]

One can thus formulate a family of Mentaculus-like axiomatizations of SM. Let us call this family $Ment(\Gamma, P^\mu, \mu)$. Members of $Ment(\Gamma, P^\mu, \mu)$ comprise FD plus the following:

[11]The implications of choosing the delta-distribution as one of our axioms will be the same whether we take the measure over phase space to be the Lebesgue measure (so that our measure-probability distribution pair is $\langle P_\delta^\lambda, \lambda \rangle$), or any other measure that is absolutely continuous with the Lebesgue measure (i.e., that assigns 0s to all of the sets that the Lebesgue measure assigns 0 to).

[12]Incidentally, I think that Beisbart (2013) is correct to challenge Frigg and Hoefer (2013)'s contention that a system comprising FD together with the delta distribution peaked at the world's actual PICs is clearly superior to the Mentaculus. Frigg and Hoefer claim that the delta distribution is just as simple as the uniform distribution (over the region of phase space corresponding to the low entropy condition specified by PH) while yielding much better fit to the actual frequencies. But, as Beisbart (2013) correctly 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!"

(PH*) a proposition characterising the initial condition of the universe as some particular element of the set $\gamma = \{\Gamma_0, \ldots, \Gamma_B, \ldots, \Gamma_\delta\}$.

(SP*) a probability distribution-measure pair $\langle P^\mu, \mu \rangle$ applied to the region of microphysical phase space specified by PH* (that is, to the relevant element of γ). Assuming the measure to be the standard, Lebesgue measure (i.e. $\mu = \lambda$), the probability distribution can be taken to be some element of the set $\rho = \{P^\lambda_U, \ldots, P^\lambda_\delta\}$ (the elements of this set besides P^λ_U and P^λ_δ are to be taken to be non-flat distributions that assign increasingly high probability to smaller and smaller regions of phase space containing the PICs).

(SP* is redundant for the case where PH* specifies the PICs, Γ_δ.) As we will see shortly, non-identical SM probabilities are entailed by $Ment(\Gamma, P^\mu, \mu)$ for different choices of $\Gamma \in \gamma$ and $P \in \rho$ (even if we hold fixed $\mu = \lambda$).

As Beisbart (2013) and others have pointed out, the PH as stated by Albert and Loewer is ambiguous. For one thing, there is an ambiguity between (at least) the following two readings of PH. On the one hand, it could be interpreted as simply saying something like 'there exists a macro-state M (and an associated phase space region Γ) and a value s such that the universe was initially in M, M has entropy value s, and s is low'.[13] On the other hand, it could be interpreted as saying 'the initial macro-state of the universe was M_0 (corresponding to phase space region Γ_0)', where M_0 in fact has low entropy ($= s$). On the former reading, PH simply specifies that the universe was initially in some-or-other macro-state with an entropy value s; on the latter reading, PH specifies exactly which macro-state the universe was in fact initially in.[14] If the latter reading of PH is intended, then it seems clear that there are still further competitor systems to the Mentaculus, which include axioms that

[13]In which case, it is not clear exactly *how low* initial entropy is specified to be by the PH. Consequently, it's not clear how big Γ_0 itself should be taken to be.

[14]Lavis (2005, 255) uses the term 'degenerate' to describe entropy values that are associated with more than one macro-state, and takes the number of macro-states sharing that value to be a measure of the value's 'degeneracy'. If s is degenerate, then the two readings of PH suggested in the main text are not equivalent. The former version of PH would make for an axiom system that is more simple (since only an entropy value, and not a precise macro-state, is specified), the latter version would make for an axiom system that is better-fitting.

achieve greater simplicity at the cost of worse fit by being less and less precise about the initial macro-state of the universe: for example by specifying the values of fewer and fewer macro-variables (thus perhaps restricting the entropy of the initial macro-state to lie below higher and higher values of s). Such systems would involve applying a uniform distribution to regions larger than Γ_0.[15]

It thus seems that we are confronted by a plethora of very good systems, none of which stands out as robustly *best*. Indeed, plausibly we are confronted with a continuum of such systems: for example, systems that apply a uniform distribution to smaller and smaller regions of phase space, each of which contains the universe's PICs. If we had a precise simplicity measure, and a precise exchange-rate between simplicity and strength/fit, then perhaps the measure and the exchange rate might produce an exact tie between systems located on this continuum. The idea would be that, according to the exchange rate, the change in strength/fit as we move along the continuum is precisely counterbalanced by the change in simplicity. More plausibly, a precise simplicity measure and a precise exchange rate is something that we cannot reasonably hope to have. If so, then it is quite plausible that none of the systems on this continuum is robustly better than all – or indeed *any* – of the others. That is, none is superior to all – or *any* – others given the imprecision of simplicity and of the exchange rate.[16]

As we saw earlier, Lewis (1994, 479) claims that, if there is not a unique best system, then there is nothing deserving of the name *law*. But we have now seen that, very plausibly, there *is* no unique best system for our world. Must we then draw the (surprising) conclusion that there are no laws for our world?

[15]Not all ways of giving less and less precise information about the universe's initial state will increase simplicity (though they will all decrease fit). For instance, while specifying the values of fewer macro-variables may increase simplicity, it doesn't seem obviously more simple to specify the values of the same number of macro-variables, but to simply specify them less precisely. For instance, an axiom that specifies that temperature t at location l is between x degrees kelvin and y degrees kelvin doesn't seem obviously more simple than one that specifies that t at l lies within a strict sub-interval of $[x, y]$.

[16]Or, more modestly: it is plausible that there is some subset of the systems on this continuum each member of which is at least as good as any system not in this subset, but none of which is robustly better than all – or indeed *any* – of the others within this subset.

This seems to be too extreme a position to take. Lewis's earlier position that, in case of ties, those theorems entailed by *all* of the tied-for-best systems would count as laws (Lewis 1983, 367) is more plausible. Thus, for example, those who have examined axiomatizations of SM that are rivals to the Mentaculus have not disputed that the *fundamental dynamical principles* are the same across all such systems. If this is indeed the case, then it is plausible that they are genuine laws of our world.

What conclusion should we draw about chances? Is it the case that if (as is plausible in light of the foregoing) there is not a unique best system for our world, there is nothing deserving of the name *chance*? Lewis doesn't explicitly say such a thing, though it might be a natural view for him to take given his position that there would be no laws in such a case. Beisbart (2013) thinks that the Best System analyst is committed to such a view:

> "The conclusion looming large here is that Humean considerations do not provide us with any clearly optimal solution to the problem of how to define a dynamics of chances. And good just isn't good enough. We need a best system that is clearly best, and not just a good one. If there isn't a best system, then there are no chances ...."

Beisbart's idea is that the Humean mosaic, together with the theoretical virtues, fails to single out a unique best system, and therefore a corresponding probability function. Consequently he claims that the BSA implies that there is nothing that counts as the *objective chance function* for our world. This is precisely analogous to Lewis's claim that there would be nothing deserving of the name of *law* if several systems were roughly tied for *best*.

Again, this seems like an extreme position to take, and it is tempting to instead adopt a position analogous to Lewis's earlier, more moderate, position on laws in the case of ties between systems. That is we might say that, in the case of a tie, any probabilities that the tied-for-best systems *agree upon* count as objective chances. 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 is possible that all the tied-for-best systems agree on the probability that a particular tritium atom will decay within the next 12.32 years. One might plausibly say that such agreed-upon probabilities count as chances.

Do the probabilities that the tied-for-best systems agree upon extend beyond the probabilities of micro-physics? It's a highly non-trivial question how different choices of initial phase space region of the universe and different probability distribution-measure pairs (i.e. different choices from the family $Ment(\Gamma, P^{\mu}, \mu)$) translate into SM probabilities at later times. If systems that include axioms that differ concerning the initial probability distribution-measure pair and/or the region of phase space to which it is to be applied, nevertheless agree on (some of) the SM probabilities, then the latter might count as objective chances.

Albert (2000, 67) argues that, provided that the sub-region of Γ_0 we choose is regularly-shaped and not too small, and provided the probability distribution that is applied to it is reasonably 'smooth' with respect to the Lebesgue measure ('smooth' in the sense that the probability density varies only negligibly over small distances in the phase space), then the resulting probabilities of thermodynamic-like behavior (that is, monotonic entropy increase) will diverge only a little from those entailed by the Mentaculus.

Frigg argues that the assumption that underlies Albert's claim – the assumption that the micro-states that lead to un-thermodynamic behavior are scattered in tiny clusters all over phase space (Albert 2000, 67) – is supported by "neither a priori reasons nor plausibility arguments . . . and merely asserting that the condition does hold is simply begging the question" (Frigg 2011, 87). If this assumption of Albert's is incorrect, then the choice of (regularly-shaped) sub-region of Γ_0 and (smooth) probability distribution may make a significant difference to the probabilities of thermodynamic-like behavior. But, even if the assumption is correct, different choices will yield *some* small but finite differences in such probabilities. Moreover, from the perspective of providing a best axiom system for the universe, it is not clear why choices of irregularly shaped regions and non-smooth distributions should be ruled out a priori. For example, the delta-distribution is as unsmooth as possible, and yet an axiom system incorporating it seems a live candidate for best systemhood.[17]

Maudlin (2007b, 2011) has argued that one can derive normal thermodynamic behavior without committing to precise assumptions about the initial probability distribution-measure pair, $\langle P^{\mu}, \mu \rangle$, provided that different choices of probability distribution-measure pair agree on which sub-regions get zero probability. Maudlin appeals to considerations of 'typicality'. When defined strictly (a more liberal definition will be considered below), a certain sort of dynamic behavior (or a limiting relative frequency for that behavior), such as thermodynamic-like behavior, counts as *typical* relative to a measure μ when that behavior (limiting relative frequency) is produced by a set of initial conditions that μ assigns measure 1 (see Maudlin (2007b, 286) and Maudlin (2011, 310)), and which hence has probability 1 on any distribution that does not concentrate finite probability on sets of measure 0 (as, for example, the delta distribution does).[18]

Maudlin's idea is as follows. Suppose that we start with some convenient measure, such as the Lebesgue measure λ,[19] and suppose that some set of initial conditions that produces a given behavior is assigned measure 1 by λ. Then, if one switches to any other measure, μ, that is absolutely continuous with λ: that is, which assigns measure 0 to all the sets that have measure 0 according to λ, the same behavior will count as typical with respect

[17]Of course, the candidate systems will all assign a high probability to thermodynamic-like behavior, otherwise they won't be well-fitting. But there is still an important difference between (for example) the extreme probabilities assigned to thermodynamic-like behavior by the system including the delta-distribution and the non-extreme probabilities entailed by the Mentaculus, even if both entail probabilities that are 'very high'.

[18]Often advocates of the 'typicality' approach to explaining thermodynamic behavior eschew talk of probability, hoping to explain thermodynamic behavior simply in terms of the large measure of the set of points (within the region of phase space associated with the low entropy initial conditions) on entropy-increasing trajectories. Maudlin, however, has no problem with probability-talk. In fact, he talks about the relevant measures as being 'probability measures' (e.g. Maudlin (2007b, 311)), suggesting that what he has in mind are measure-probability distribution pairs. In any case, it seems implausible that thermodynamic behavior can be explained in terms of the *sizes* of the respective regions, unless one can say that the size of a region is correlated to the probability that a system is located within it. One can do the latter only by appealing to some or other probability distribution (such as the one that's uniform with respect to the measure in question). I thus follow orthodoxy (and the Albert-Loewer line) in supposing that both a measure *and* a probability distribution with respect to that measure is needed to explain thermodynamic behavior.

[19]As well as being 'convenient' – the Lebesgue measure is 'convenient' in the sense that it is invariant under the dynamics (Liouville's Theorem) – the measure must also be such that we "feel comfortable that it represents a reasonable choice of *sets of measure one and zero*" (Maudlin 2007b, 287). What constitutes a reasonable choice is left unclear by Maudlin.

to μ as did with respect to λ. And provided that we choose a probability distribution P^{μ} which doesn't assign finite probability to sets that have measure 0 according to μ, then the measure-probability distribution pair $\langle P^{\mu}, \mu \rangle$ will assign probability 1 to the behavior in question (precisely because the initial conditions that don't lead to this limiting frequency still get assigned probability 0).[20]

Maudlin (2011, 311) claims that "[i]t is also extremely plausible that the choice of the exact interval [or, in the case that concerns us, the exact region of phase space] is irrelevant: make it larger and smaller, and still the same [behavior] will be typical". Maudlin does not present an argument for this latter claim, and it is not clear how such an argument would go. In particular, there is no reason to suppose that the same measure applied to different sized regions of phase space will assign zero to all of the same sets. Perhaps the idea is that the set of 'bad' points (i.e. the ones that don't yield entropy increasing behavior) has the same measure in different 'reasonable' choices of subregion of the phase space. Such a claim would seem to depend upon the sort of 'scattering' assumption discussed above, which is made by Albert, and questioned by Frigg.

Maudlin certainly claims that this typicality reasoning applies to thermodynamics:

> "If we have two metal rods at different temperatures and then bring them into thermal contact, typical behavior (in the 'thermodynamic limit') will be for the motions of the atoms in the rods to evolve so that the temperatures in the rods equalize. This is the way that the laws of thermodynamics, which are *deterministic*, are 'reduced' to statistical mechanics: Thermodynamic behavior is shown to be typical behavior." (Maudlin 2007b, 287)

By "in the 'thermodynamic limit'", Maudlin presumably means as the number of molecules in the system increases to infinity. Goldstein and Lebowitz (2004, 57) claim that, as the

[20]Maudlin doesn't explain why any reasonable measure must be absolutely continuous with the Lebesgue measure. And it is not at all clear that it must be. In particular, Frigg (2011, 88) points out that there is no proof that all measures that are invariant under the dynamics (for a given system) are absolutely continuous with the Lebesgue measure. But let's set this worry aside, for the sake of argument.

number of molecules goes to infinity, the (Lebesgue) measure of the initial conditions that lead to significant violations of the Second Law of Thermodynamics goes to zero.[21]

The trouble with this is that no thermodynamic system, not even the universe as a whole, comprises infinitely many molecules. Consequently, thermodynamic behavior (in our world) isn't 'typical' in the sense that the initial conditions that produce it form a set that has Lebesgue measure 1. Rather, that set has Lebesgue measure only $1 - \epsilon$ (for small but finite ϵ), as Maudlin (2007b, 289) acknowledges.[22]

Given that this is so, there are measures that are absolutely continuous with λ that assign a range of sizes to the set of initial conditions that produce thermodynamic-like behavior. Maudlin claims that, since the Lebesgue measure of this set is high, it will also be very high on "almost any reasonable" alternative measure (Maudlin (2011, 314); see also Maudlin (2007b, 289)). Nevertheless, "there will be some ... degree of sensitivity ... [to] the particular measure chosen" (Maudlin 2007b, 290). Because of this sensitivity of the size of the 'good' set to the choice of measures, there are probability distributions (including the uniform distributions) relative to such measures that assign a range of probabilities to that set (even though they don't assign finite probability to sets of zero measure), and hence entail a range of probabilities for entropy-increasing behavior.[23]

It thus seems that the differences in our tied-for-best systems translate into different probabilities for entropy-increasing behavior. Of course, there must be a certain amount of qualitative agreement between the systems: each must entail that the probabilities of entropy increase in (most isolated) non-equilibrium systems are 'very high', since any system that did

[21]Frigg (2011, 88) objects that there is no proof that the set of 'bad' initial conditions has Lebesgue measure zero even in the infinite case. If it does not, then measures that are absolutely continuous with the Lebesgue measure may assign this set a range of different sizes and hence probability distributions that (e.g.) are uniform with respect to these various measures will yield a range of different probabilities for entropy-increasing behavior.

[22]In fact Frigg (2011, 85-87) points out that there is no general proof that micro-states on trajectories that lead to thermodynamic equilibrium are typical for the systems of concern to us even in the weak sense of having Lebesgue measure at least $1 - \epsilon$.

[23]Indeed, in the context of devising candidate best systems, it is unclear that it is reasonable to adhere to Maudlin's stricture that one ought to "avoid extremism: don't pick a new measure that concentrates finite probability on a set that got zero probability originally, and don't shrink the [region that the probability distribution is applied to] down to a point" (Maudlin 2011, 311). In the context of constructing candidate best systems, we are well within our rights to consider systems that incorporate the delta distribution, or that specify the universe was initially in the point-sized region of phase space, Γ_δ.

not would not be well-fitting enough to be among the best. Still the different probability-distribution-measure pairs invoked by different systems, and the different sized regions of phase space to which the probability distributions are applied are liable to translate into small but finite differences in the probabilities assigned to entropy-increasing behavior.

That different probability distributions over the initial conditions of the universe may at least yield qualitative agreement on the SM probabilities is suggested by Callender (2011, 107), when he discusses what he calls the 'Liberal Globalist' approach to axiomatizing statistical mechanics. Globalist approaches are those that – like the Mentaculus, and all of the other systems that we have considered thus far – seek to derive SM probabilities by means of the application of a probability distribution to the region of phase space associated with some or other low entropy initial macrostate *of the universe as a whole*. (Localist approaches, as we'll see in Section 6, seek to derive the SM probabilities by applying probability distributions to the regions of phase space associated with the initial macrostates of isolated *subsystems* of the universe.) Liberal Globalism acknowledges that there is not a unique distribution that yields a high probability of entropy increase for the universe:

> "Liberal Globalism notices that many other probability distributions over initial conditions will 'work,' i.e. make probable the generalizations of thermodynamics, in addition to the standard one [i.e. the one that is uniform on the Lebesgue measure]. David Albert (private communication) then suggests the following strategy. Take the set SP_i of all such probability distributions that work. There will be uncountably many of these. Dictate that physics is committed to those propositions on which SP_i plus the dynamical laws all agree. Such a picture will . . . be agnostic about claims . . . [where] the probability distributions . . . disagree and so the theory makes no claim. The advantage of this position, if it works, is that it isn't committed to any one probability distribution doing the job . . . ." (Callender 2011, 107)

Here the point seems to be that different distributions will often yield qualitative agreement about objective chances: for example, they will agree that the probability of entropy increase in a given non-equilibrium isolated system is 'high', though they do not agree on a precise numerical value for such a chance. It seems, however, that Albert and Callender take the view that, while it will then be *true* that the chance of entropy-increasing behavior, in some particular situation, is *high* (because the tied-for-best systems agree on this proposition), there will be no fact of the matter about what the objective chance *is* (since the tied-for-best systems yield different sharp probabilities for entropy increase). By contrast, in what follows, I will argue that there is a fact of the matter about what the objective chance *is*: namely, it is the set of sharp probability values entailed by the tied-for-best systems. That is to say, it is the unsharp chance that corresponds to this set of values.[24]

In general, it seems plausible that the probabilities that the tied-for-best systems all agree on ought to be counted as objective chances. But the SM probabilities entailed by the set of apparently tied-for-best systems diverge from one another. In disanalogy to the case of laws, I will argue that the chances *aren't limited* to the probabilities that are agreed upon by the tied-for-best systems. Rather, in the case where divergent probabilities are entailed by the tied-for-best systems, the chances correspond to the *set* of probabilities entailed by these tied-for-best systems.

5. Unsharp Best System Chances

The fact that the Humean mosaic, together with the (imprecise) relation of Best System-supervenience does not uniquely fix a single axiom system should not lead the Best System analyst to conclude that there are no chances in the world. It shouldn't even lead her to conclude that the *only* chances in the world are those probabilities that the set of tied-for-best systems all agree on. In cases where the tied-for-best systems disagree about the probability

[24] Note that, if the tied-for-best systems all entail that the probability for entropy increase given a non-equilibrium macro-state for an isolated system is *high*, but do not agree on sharp valued probabilities, then one might think that the correct thing to say about *laws* in this case is that there exists a qualitative probabilistic version of the Second Law of Thermodynamics (but no precise quantitative one). I find this view plausible, but nothing that I say about chances turns upon it.

for some event, the Best System analyst shouldn't say that there is no well-defined objective chance for that event. Rather, she should simply deny that the objective chance for that event is sharp.

If there is no clear winner in the Best System competition, the natural thing for the Best System analyst to say is that the *set* of probability functions corresponding to the tied systems constitutes the set of chance functions for the world.

Where the probability functions associated with the tied systems agree on the probability for a particular event then the objective chance for that event is sharp. As suggested above, this seems quite plausible when, for example, we are considering microphysical events like the decay of a tritium atom within the next 12.32 years.[25] On the other hand, when the probability functions of the tied-for-best systems yield a set of values, as appears to be the case for thermodynamic behavior in any given situation, then the set of values constitutes an unsharp chance for the event in question. Put in terms of the family $Ment(\Gamma, P^{\mu}, \mu)$ of plausibly tied-for-best Mentaculus-like systems which we formulated in the previous section, the suggestion is that differing SM probabilities are entailed by $Ment(\Gamma, P^{\mu}, \mu)$ for various choices of initial phase space region, Γ, and for different choices of probability distribution-measure pair $\langle P^{\mu}, \mu \rangle$ and that, by considering the probabilities entailed by the various tied-for-best members of the family $Ment(\Gamma, P^{\mu}, \mu)$, we obtain SM probabilities that are set-valued, i.e. are unsharp.

The possibility of unsharp chances, especially in the context of the BSA, has been noted in passing by Hájek (2003b). After discussing the possibility that credences may be unsharp – or

[25]Although one might even wonder whether there is a unique real number that constitutes the chance of this decay event (see the discussion of Maudlin (2011, 295-296)). One might take the view that the chance for the decay event is unsharp. There are two interestingly different ways of construing a situation in which the fundamental dynamics, FD, doesn't assign a unique real number as the probability of occurrence for some microphysical event. On the one hand, one might construe this as involving a tie between systems, each of which entails Quantum Mechanics-like theorems that entail various unique real-valued probabilities for decay. If so, QM chances would be unsharp for the same reason that SM chances are. Alternatively, the situation in QM might be construed, not as one in which there is a tie amongst systems each of which gives sharp QM chances (i.e. is endowed with exactly one probability function), but as a case in which the (unique) best axioms for QM themselves entail unsharp chances (i.e. include a set of probability functions). Perhaps – though I will not explore this question here – these two ways of construing the case are equivalent.

in Hájek's terminology 'vague'[26] – 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) notes 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' 1973 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."[27]

[26] I don't think that Hájek's terminology is ideal. Specifically, I think that the term 'vague probability' is best reserved for the case when it is a vague matter which elements of the set of reals in [0, 1] belong to the set that constitutes an unsharp probability (cp. Joyce (2005, 167n), Sturgeon (2008, 158-159), and Sturgeon (2010)). Adopting this terminology, it is possible to have an unsharp chance that is not vague – i.e. when more than one element of the reals in [0, 1] is an element of the set that constitutes the unsharp chance, but it is a perfectly determinate matter which elements of the former set belong to the latter set. In any case, the sort of chances that Hájek discusses correspond to what I'm calling 'unsharp' chances.

[27] Cp. also Hájek and Smithson (2012, 39).

The possibility suggested in this latter quote is one that I wish to push. As has been seen, competing ways of axiomatizing statistical mechanics provide a strong motivation for thinking that there is indeed a tie and hence – on a plausible construal of the BSA – that there are unsharp Best System chances *in our world*. It is thus particularly plausible that the BSA yields unsharp chances in the case of SM, as opposed to the case of quantum mechanics that Hájek considers.[28]

The basic reason for holding that there are unsharp chances in this context is that it seems unreasonable to claim that, in the case of a tie between systems, there simply is *nothing* playing the chance role in guiding rational credence, and hence action. This argument will be developed in detail in Section 7 below, but the following is a sketch.

Hoefer (2007, 580-587) and Frigg and Hoefer (2010) argue that Humean chances are constraints on rational credence because of the tight connection between Humean chances and actual frequencies. Specifically they argue that this allows for a 'consequentialist' justification for calibrating one's credences with the Humean chances. The basic idea is that, given the tight connection between Humean chances and actual frequencies, agents betting according to the Humean chances will do well in the long run.

The assumption that there is a unique probability function that serves as the Humean chance function does not seem essential to this argument.[29] On the BSA, in the case of a tie between systems, it seems that one can argue that the set of probability functions entailed by the tied systems would constrain reasonable credence. Specifically, when confronted with a tie between systems, an agent who knows the *set* of probability functions corresponding to the tied-for-best systems, and who knows that a certain chance setup is instantiated, and who has no inadmissible information (no information relevant to the outcome of the chance

[28]Hájek's suggestion that ties between systems yield vague or unsharp laws of nature is intriguing. As already suggested it seems plausible that, in cases of ties, the theorems (if any) common to *all* of the tied-for-best systems constitute the (sharp) laws of nature. However, since the tied-for-best systems entail different SM probabilities, but nevertheless agree on the qualitative fact that the probability of Second Law-like behavior is 'very high', it was suggested in Footnote 24 that this might be taken to imply the existence of a *qualitative* probabilistic version of the Second Law. Perhaps a law that assigns only a qualitative probability to certain behavior counts as an 'unsharp' law (since it is a kind of summary of what is common to the various quantitatively precise generalizations entailed by the different tied-for-best systems).

[29]Frigg and Hoefer (2013) themselves find it plausible that there may not be a unique best system for our world, and that there may be "a family of closely related cousins" that come out tied-for-best.

event beyond knowledge of the initial chance setup and the set of probability distributions entailed by the tied-for-best systems) rationally ought *not* to adopt a credence value that lies outside the set of probability values entailed by the systems that are tied-for-best.

For example, all tied-for-best systems entail a very low probability for entropy decrease in (most) isolated systems. An axiom system that does not would be straightforwardly excluded from the set of tied-for-best systems: it will inevitably be highly ill-fitting, since fit implies a close correspondence of the entailed probabilities to the actual frequencies. One can thus offer a 'consequentialist' justification for not adopting a credence outside of the set of probabilities yielded by the tied systems: betting as though entropy-decrease in non-equilibrium isolated systems is not very improbable would lead one to do very badly in the long run.

Indeed, rather plausibly, a reasonable agent who knows that a certain chance setup is instantiated, who knows that the *set* of probability functions corresponding to the tied-for-best systems yield a (non-singleton) set of probabilities for a given outcome conditional upon the instantiation of that chance setup, and who has no inadmissible information, would have an unsharp credence in the outcome in question: specifically, her credence would be represented by a set of values corresponding to those entailed by the probability functions of the tied systems. In such a situation the agent would have no rational basis to choose between the probability functions entailed by the tied systems but would be rationally compelled to base her behavior upon the relevant unsharp chance. Thus, in such a situation, the set of values entailed by the tied systems is playing the key chance role of guiding reasonable credence, and thereby constitutes an unsharp chance.

There is one worry here. Given a set of systems that are tied for best, some of which have greater fit, and others of which have greater simplicity, it would appear to be (instrumentally) rational to set one's credences according to the probabilities entailed by that system which has the greatest fit (i.e. accords best with the actual frequencies), since betting according to those probabilities would yield the greatest payoff in the long run. This is a genuine worry, but it is a general problem for the BSA. Even in the case of a unique best system, the Best

System chances are liable to depart somewhat from the actual overall world frequencies. This is because considerations of *simplicity* go into determining a Best System, and not just considerations of fit. But it seems difficult to argue that the Best System probabilities, rather than the actual overall world frequencies, are the best players of the chance role in guiding rational credence: if one knew both, one would do better in the long run if one bet according to the actual frequencies. The advocate of the BSA requires some general explanation of why the best-fitting probabilities (i.e. the actual overall world frequencies) aren't automatically the chances (perhaps, for example, she could appeal to other aspects of the chance role). Whatever answer is deployed in this context will also be available to us in explaining why the best fitting of the tied systems doesn't automatically deliver the chances (and why simpler theories also have strong claims to do so). This issue will be taken up in more detail in Section 7 below.

6. Localist Approaches to SM

Callender (2011) calls attempts to derive the SM probabilities from a probability distribution over a phase space region corresponding to some-or-other specification of the low entropy initial macrostate of *the universe as a whole* 'Globalist' approaches to axiomatizing SM. The Mentaculus is one example of a Globalist approach. As we have seen, this can be generalised to a family of such approaches, $Ment(\Gamma, P^{\mu}, \mu)$, many of which are plausibly just as good systematizations as the Mentaculus.

In contrast to Globalist approaches, 'Localist' approaches (see Callender (ibid.) and Frigg and Hoefer (2013)) attempt to derive the SM probabilities from probability distributions over the initial states of the various *approximately isolated subsystems of the universe*. As with Globalist approaches to axiomatizing SM, it is plausible that there is a set of Localist approaches that are tied-for-best (at least within the limits of precision of notions like simplicity and balance). Many of the preceding points about unsharp chances in the Globalist case therefore carry across to the Localist case. So one should also believe in unsharp chances if one regards Localist axiomatizations as more promising (i.e. as plausibly providing better systems) than Globalist axiomatizations.

One important motivation for those who have advocated the Localist approach is the fact that there is not much evidence – let alone proof – that a high probability of entropy increase for (approximately) isolated *subsystems* of the universe can be derived from Mentaculus-like Globalist axiomatizations of SM (cp. Callender (2011, 106); Frigg and Hoefer (2013)). Indeed, Callender (2011, 100-101) points out that it simply doesn't follow from the fact that (e.g.) a uniform distribution over the low-entropy initial condition of the universe specified by the PH makes entropy increase for the universe as a whole very likely, that entropy increase in (approximately) isolated non-equilibrium subsystems of the universe is also very likely. The reason for this is that the phase spaces corresponding to these subsystems have a lower dimensionality than does the phase space of the universe as a whole. These subsystems thus have zero volume in the phase space of the universe. So increasing entropies in individual subsystems simply don't 'add up' to entropy increase in the universe as a whole. Conversely, entropy increase in the universe as a whole does not imply that entropy increases in 'most' subsystems. Yet it seems plausible that a Best System for our universe ought to entail a high probability for entropy increase in such subsystems, otherwise it will be ill-fitting (cp. Callender (2011, 100, 106)).

The Localist proposal is that, rather than simply applying a probability distribution (as in SP*), to a region of the phase space of the universe that contains the initial state of the universe (as described by PH*), we should apply probability distributions to regions of the phase spaces of (approximately) isolated subsystems of the universe that contain the actual initial conditions of those systems (Callender 2011, 96-97). Thus Callender (2011, 96) suggests that "we impose SP at (roughly) the first moment low-entropy macroscopic systems become suitably isolated". Callender calls those subsystems of the universe such that the imposition of SP on a region of their phase space corresponding to a low-entropy initial macro-state is predictively successful 'SP-systems' (Callender 2011, 106). Callender's preferred version of Localism is one that is 'agnostic' about whether the existence of SP-systems is itself somehow to be explained (as on Globalist versions of SM) in terms of the application of SP to some sub-region of the phase space of the universe as a whole.

Frigg and Hoefer (2013) 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^λ that is uniform on the standard Lebesgue measure, λ, should be applied to the initial state of SP-systems to generate the SM chances. They argue that the distribution P_U^λ 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 an SP-system's initial macro-state (that is it's macro-state at some initial time t_0), and they argue that:

> "There is a well circumscribed class of objects to which a chance rule like [P_U^λ] applies (gases, etc.). Each of these ... 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) $[\lambda(C)/\lambda(\Gamma_p)]$ 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 $[\lambda(C)/\lambda(\Gamma_p)]$ is a flat distribution over Γ_p." (Frigg and Hoefer 2013, square brackets indicate where I have changed notation for consistency with my own)

Frigg and Hoefer's idea is that, if we consider systems whose phase spaces have (roughly) the same dimensionality, then the precise actual initial conditions of those various systems will be roughly uniformly distributed over that phase space. Thus the application of a uniform probability distribution to that phase space can be justified in terms of this frequency: the uniform probability distribution *fits well* the uniform frequency with which the initial conditions of actual systems lie in the various sub-regions of the phase space. (And, in particular, it fits better than any comparably simple distribution.) Frigg and Hoefer's justification for applying the uniform distribution to subsystems of the universe is thus somewhat different from Albert and Loewer's justification for applying it to the initial conditions of the universe as a whole. 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 2013). 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 2013).

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 to produce a set of actual initial conditions that is large enough to single out the uniform distribution as that which supplies the robustly best balance of fit and simplicity. Frigg and Hoefer (2013) 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 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 is not at all clear that 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^λ 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 that there are enough to guarantee that applying P_U^λ 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 SP systems fails to nail down P_U^λ 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 – including those with peaks at each of the actual initial conditions of the subsystems in question) being more complex, and the worse fitting (e.g. P_U^λ) being more simple.

Analogously with 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 spaces of the appropriate subsystems of the universe (where these regions contain the precise initial conditions of the systems in question). Again, picking out smaller sub-regions will typically buy better fit (by increasing the probability of the system's actual macro-evolution) but will involve more complexity (because it involves specifying the initial macro-state of the system more precisely, or adding in some information about the system's initial micro-state). If this is so, then the upshot will be a range of tied-for-best Localist axiomatizations of SM entailing different SM probabilities. Consequently, even if Localist axiomatizations are superior to Globalist axiomatizations, the Best System analyst should still conclude that there exist unsharp SM chances, where this time the unsharp SM chances are constituted by the set of probabilities entailed by the tied-for-best *Localist* axiomatizations of SM.

7. Unsharp Chances and the Chance-Credence Connection

The thesis that reasonable credences may be imprecise has gained some popularity in formal epistemology. Joyce (2005, 156) claims that "[t]he idea that people have sharp degrees of belief is both psychologically implausible and epistemologically calamitous". The psychological implausibility of precise credences is emphasised by Hájek:

> "What is your subjective probability that the Democrats win the next election? If you give a sharp answer, I would ask you to reconsider. Do you really mean to give a value that is precise to infinitely many decimal places? If you're anything like me, your probability is *vague* – perhaps over an interval, but in any case over a range of values."[30] (Hájek (2003b, 293); cp. Elga (2010, 6))

The 'epistemological calamitousness' of precise credences is emphasised by Joyce (2005). Specifically Joyce (2005, 170) argues, concerning the notorious Principle of Insufficient Reason (or Indifference Principle):

> "The real difficulty is not that the Principle of Insufficient Reason might be incoherent; it is that the Principle, even if it can be made coherent, is defective epistemology. It is wrong-headed to try to capture states of ambiguous

[30]Recall from Footnote 26 that by 'vague' probability, Hájek means what I mean by 'imprecise' probability: namely a probability that is *set-valued*.

> or incomplete evidence using a *single* credence function. Those who advocate this approach play on the intuition that someone who lacks evidence that distinguishes among possibilities should not 'play favorites,' and so should treat the possibilities equally by investing equal credence in them. The fallacious step is the last one: equal treatment does not require equal credence. When [an agent who has no evidence concerning which of a set of hypotheses is true assigns to each hypothesis] an equal probability he is pretending to have information he does not possess. His evidence is compatible with *any* distribution of objective probability over the hypotheses, so by distributing his credences uniformly over them [he] ignores a vast number of possibilities that are consistent with his evidence."

Joyce (2005, 171) continues:

> "As sophisticated Bayesians ... have long recognized, the proper response to symmetrically ambiguous or incomplete evidence is not to assign probabilities symmetrically, but to refrain from assigning precise probabilities at all. Indefiniteness in the evidence is reflected not in the values of any single credence function, but in the spread of values 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. It is not just that sharp degrees of belief are psychologically unrealistic (though they are). Imprecise credences have a clear epistemological motivation: they are the proper response to unspecific evidence."

My suggestion will be that there is an additional motivation for supposing that a rational agent will sometimes have imprecise credences. It is not just that imprecise credences are more psychologically realistic, or that they are the proper response to a lack of evidence, or to unspecific evidence. They can also be the proper response to precise evidence about imprecise chances.

White (2010, 174) suggests that many of those who endorse unsharp credences subscribe to:

> "*Chance Grounding Thesis.* 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."

I will argue that, even when one knows the chances, it needn't always be legitimate for one to have sharp credences. Specifically, it will not be legitimate if those known chances are themselves set-valued.

As Joyce indicates, in the quotation given above, the standard formal representation of imprecise credences appeals to the idea that the credences of a rational person, 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'. Since I take conditional probability to be basic (cp. Hájek (2003a,b, 2007)), I assume that each element of $\boldsymbol{cr}$ is a sharp-valued 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: $cr_i(X|Y) = x$. (Unconditional probabilities can be defined via $cr_i(X) =_{def} cr_i(X|\top)$, where $\top$ is the tautology.)

Joyce (2005, 2010) observes that determinate facts about S's degrees of belief correspond to properties that are invariant across all elements of $\boldsymbol{cr}$. For example, a person has a sharp credence in X conditional upon Y when $cr_i(X|Y) = x$ for every $cr_i(\cdot|\cdot) \in \boldsymbol{cr}$, and she is more confident in X than in Y given Z if $cr_i(X|Z) > cr_i(Y|Z)$ for every $cr_i(\cdot|\cdot) \in \boldsymbol{cr}$.

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 conditionalized upon D. Specifically, if S's initial representor is $\boldsymbol{cr}$ and she learns D (and nothing else), then her new representor is $\boldsymbol{cr^D}$, where:

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

(Conditionalization is undefined for those probability functions in S's initial representor, $\boldsymbol{cr}$, for which $cr_i(D) = 0$. So these probability functions are 'weeded out' rather than updated in the transition to $\boldsymbol{cr^D}$.)

Note, however, that no completely compelling argument has been produced to show that $(*)$ is the correct way of modelling rational 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) (cp. also 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 upon the evidence is less straightforward, and attempts to justify it (see Grove and Halpern (1998)) are less compelling.[31]

For now, however, let us assume that $(*)$ *is* the correct way in which to model rational belief change for a rational agent whose credal state is modeled by a representor. Such an agent S can come to have imprecise credences in response to imprecise evidence as follows. 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 next', and that D = 'all balls in urn 1 are either red or blue' (D is imprecise because it doesn't say what the proportions of red

[31]Walley (1991) presents an argument for a 'generalized Bayes rule' in the context of an account that represents imprecise probabilities in terms of upper and lower probabilities, rather than *sets* of probability functions.

and blue balls are). If S has no other evidence bearing upon A, then it is plausible that her post-update credence in A should be spread out over the whole of the $[0, 1]$ interval.

Modeling the agent S's credences by a representor $\boldsymbol{cr}$ that is updated according to $(*)$ can accommodate this. The idea is that various elements $cr_i \in \boldsymbol{cr}$ 'interpret' the evidence D differently. That is to say, 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 $cr_i^D(A)$. More precisely, the idea is that there is 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 credences, $cr_i^D(A)$ and $cr_j^D(A)$ will diverge. The idea is that, for a rational agent, the probability functions in her initial representor must be 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.

It is plausible that an unsharp chance for some proposition X is just one among many types of imprecise evidence that one might have about X. Suppose that $\boldsymbol{ch}$ represents the set of probability functions associated with the tied-for-best systems. For want of a better term, call the set **ch** the *cadentor*[32]. Each element of the cadentor is a sharp 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 $ch_i(X) =_{def} ch_i(X|\top)$, where $\top$ is the tautology.) The elements $ch_i \in \boldsymbol{ch}$ are the probability functions entailed by members of the set of tied-for-best systems. Abusing notation slightly, we can let $\boldsymbol{ch}(X|Y)$ represent the function that maps ordered pairs of propositions $\langle X, Y\rangle$ to the set of values that the probability functions in $\boldsymbol{ch}$ give for X conditional upon Y: that is $\boldsymbol{ch}(X|Y) = \{ch_i(X|Y) : ch_i \in \boldsymbol{ch}\}$.

We might suppose that updating upon an unsharp chance is aptly modeled by $(*)$. For example, suppose that T_1 is a proposition specifying that the thermodynamic state of some isolated system I at time t_1 is m_1, while T_2 is a proposition specifying that the thermodynamic state of I at time t_2 is m_2. Suppose that $\boldsymbol{ch}(T_2|T_1) = \{x : x \in \mathbb{R}, 0.2 \leq x \leq 0.3\}$: that

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

is, the probability functions in the cadentor – that is, the $ch_i \in \boldsymbol{ch}$ – are such that, for all and only values x such that x is a real number in the [0.2, 0.3] interval, there is a $ch_i \in \boldsymbol{ch}$ such that $ch_i(T_2|T_1) = x$. (Informally: the probabilities for T_2 conditional upon T_1 entailed by probability functions in the cadentor are 'spread out' all over the [0.2, 0.3] interval.) Let Ω be the proposition that $\boldsymbol{ch}(T_2|T_1) = \{x : x \in \mathbb{R}, 0.2 \leq x \leq 0.3\}$, and suppose that S learns that $\Omega \& T_1$. Then, since $\Omega \& T_1$ constitutes some imprecise evidence pertaining to T_2, the formula $(*)$ applies where we let $D = \Omega \& T_1$.

How exactly ought rational credence be constrained by evidence concerning unsharp chances? It will be difficult to answer this question definitively given that there is no consensus on the correct rational decision theory for the case where agents have unsharp credences (see, e.g., Elga (2010) and Bradley (2013)). However, one natural proposal is that, when an agent learns a set-valued chance for A, her credence ought to come to be captured by the same set of values. Let me try to state this somewhat more precisely.

Since I am taking conditional chance to be basic, I prefer to begin with a formulation of the chance-credence connection that – in contrast to Lewis (1980)'s Principal Principle (PP) – takes *conditional* chances to be the quantities that (in the first place) constrain rational credences.[33] I take it to be more or less platitudinous that conditional chances place the following constraint on rational credences. Suppose that a rational agent S knows that a certain chance setup c is instantiated and knows that the chance of A conditional upon the fact that c is instantiated is x. And suppose that the remainder of S's evidence is admissible in the sense that it doesn't contain any information about the truth-value of A that isn't simply information about whether c is instantiated or about the chance of A to which c gives rise. Such an agent has a credence in A equal to x.

[33]The 'New Principle' (NP) proposed by Hall (1994) and Lewis (1994) has conditional chances guiding rational credence. However, I have my reservations about NP, at least as Hall and Lewis formulate it, since their formulations incorporate the assumption that chances are the sorts of thing that would guide rational credences given knowledge of some complete initial microphysical history of the world and of the fundamental laws of nature. Such an assumption leads quickly to the view that the only genuine chances are those of Quantum Mechanics (see Hoefer (2007, 558-9) and Schaffer (2007, 128)). But it also effectively begs the question against those (such as myself) who think it plausible to regard the probabilities that figure in higher level theories, such as SM, as chances.

Let me try to state this principle somewhat more precisely. Let $Ch(\cdot|\cdot)$ be the chance function (which for now is assumed to be unique and to yield point values). Let $Cr(\cdot|\cdot)$ be any reasonable initial credence function (which, again, is for now assumed to be unique and point-valued). Suppose that F is some proposition specifying that a certain chance setup c is instantiated, so that $Ch(X|F)$ is well-defined for some outcome-specifying propositions X.[34] Suppose that A is just such an outcome-specifying proposition, so that there is a well defined chance $Ch(A|F)$. Finally, let E be any proposition that is admissible in the sense that it doesn't convey any information about the truth-value of A that isn't simply information about whether the chance setup c described by F is instantiated or about the chance of A conditional upon the fact that c is instantiated. Then conditional chances guide rational credence in the sense captured by COND:[35]

(COND) $$Cr(A|Ch(A|F) = x\&F\&E) = x$$

The principle as stated must be qualified: it is important that F itself not be inadmissible. That is, F must not bear upon reasonable credence about A otherwise than bearing upon reasonable credence that c is instantiated or upon reasonable credence about the chance of A to which c gives rise. Where c is a setup involving an unbiased coin about to be flipped by a fair flipper, and A is the proposition that the coin lands heads, F should *not* be the proposition that an unbiased coin is about to be flipped by a fair flipper *and the Oracle says that the coin will land heads* (at least not unless the latter is counted as a complex chance setup by our best theory of chance).

[34]Lewis took the only genuine chance setups to be complete initial microphysical histories of the world. I find it plausible that many other sorts of things (e.g. the macro-state of some thermodynamically isolated system, or an unbiased coin together with a flipper) may serve as chance setups (cp. Hoefer (2007, 564-5)). This, however, is a matter that is ultimately to be decided by our best metaphysical theory of chance.

[35]As far as I can tell COND is essentially the principle that is advanced by Hoefer (2007, esp. 574-5) as an interpretation of the PP .

We should thus require that F be a *minimal* specification of a chance setup c, in the sense of providing 'just enough' information about c so that $Ch(A|F)$ is well-defined and expresses the chance of A to which c gives rise (without conveying additional, potentially inadmissible, information). One first pass at attempting to capture this more precisely would be to require that F mustn't be such that (a) there is a strictly less informative proposition F' such that $Ch(A|F) = Ch(A|F')$,[36] unless (b) there is some proposition F'' of intermediate informativeness such that $Ch(A|F) \neq Ch(A|F'')$.

Let H be the proposition that the coin will land heads. According to the criterion just outlined, the proposition O: *an unbiased coin is about to be flipped by a fair flipper and the Oracle says that the coin will land heads* is not minimal in the requisite sense, since there is a strictly less informative proposition U: *an unbiased coin is about to be flipped by a fair flipper* such that $Ch(H|O) = Ch(H|U)$.[37] Moreover, since the second conjunct of O contains no information relevant to the chance of H, it seems that there is no proposition T that is intermediate between U and O in informativeness such that (b) is satisfied. We thus get the desired result that O is not minimal in the requisite sense.[38]

The most natural way of extending a principle like Cond to set-valued chances and set-valued credences is as follows. As before, let $\boldsymbol{ch}$ be the set of probability functions that constitutes our unsharp chance, and let $\boldsymbol{cr}$ be any reasonable representor (that is, any set of credence functions that could model a rational agent's epistemic state). As before (and with a slight abuse of notation), let $\boldsymbol{ch}(X|Y)$ 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: that is $\boldsymbol{ch}(X|Y) = \{ch_i(X|Y) : ch_i \in \boldsymbol{ch}\}$. With a similar slight

[36]If we understand propositions as sets of possible worlds, then F' is strictly less informative than F iff $F \subset F'$.

[37]Again, I'm assuming that the state of affairs described by O will not itself count as a (complex) chance setup, yielding different chances for A than the simple coin-flip setup does, according to our best theory of chance.

[38]The reason for condition (b) is to deal with the sort of case in which, coincidentally, the chance for H conditional upon U exactly matches the chance for H conditional upon a proposition Q giving the quantum mechanical state of the coin and the flipper. Plausibly U is strictly less informative than Q. So condition (a) would be satisfied where $F = Q$. Yet of course it is still the case that $Cr(H|Ch(H|Q) = x\&H\&E) = x$. This is provided for by the fact that (b) is satisfied: there are propositions of intermediate informativeness (e.g. ones that specify the state of the coin-plus-coin-flipper system at an intermediate level of detail) conditional upon which the chance of H differs or (perhaps more plausibly) goes undefined.

abuse of notation, let $\boldsymbol{cr}(X|Y)$ represent the function that maps ordered pairs of propositions $\langle X, Y\rangle$ to the set of values that the probability functions in $\boldsymbol{cr}$ give for X conditional upon Y: that is $\boldsymbol{cr}(X|Y) = \{cr_i(X|Y) : cr_i \in \boldsymbol{cr}\}$ (cp. Bradley (2013)). Finally, let F be some admissible proposition describing a chance setup c, let A be a proposition describing one of the possible outcomes of c, and let E be any admissible proposition (i.e. any proposition that doesn't convey information about the truth-value of A that isn't simply information about whether the chance setup c described by F is instantiated or about the unsharp chance of A conditional upon the fact that c is instantiated). Then the generalized chance-credence connection is captured by COND*:

(COND*) $$\boldsymbol{cr}(A|\boldsymbol{ch}(A|F) = \boldsymbol{x}\&F\&E) = \boldsymbol{x}$$

Informally, COND* says that a rational agent's credence in A, conditional upon the proposition F, the proposition that the set-valued chance for A conditional upon F is $\boldsymbol{x}$, and any admissible proposition E, is $\boldsymbol{x}$. More precisely, COND* says that any reasonable representor is such that the result of conditioning each probability function within it upon the proposition F and the proposition that the set of probabilities for A generated by conditioning each of the probability functions in the cadentor upon F is $\boldsymbol{x}$, and any other admissible information E, is the set $\boldsymbol{x}$ of probabilities for A.

Hájek and Smithson (2012, 38) seem to have a similar extension of the PP in mind when, in considering the possibility that chances might be unsharp (or, in their terminology 'indeterminate'[39]), they say:

> "[A]ssume a version of a chance-credence coordination principle, such as Lewis's well-known Principal Principle ... according to which your credence

[39]This terminology seems unhappy to me, since it has epistemic connotations. Yet when Hájek and Smithson speak of 'indeterminate chance', they mean chance that is objectively indeterminate – i.e. the same as I mean when talking about 'unsharp chance'

in a proposition, conditional on the chance of that proposition being x, should be x (see Lewis for further fine-tuning). Extend this in a natural way to allow for indeterminate chances: your 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."

Note that COND and the PP gain much of their plausibility from the assumption that updating upon evidence proceeds by conditionalization: that is, $Cr^D(\cdot|\cdot) = Cr(\cdot|\cdot\,\&D)$. This is because the plausibility of these principles depends largely upon the intuition that chances guide rational action. But it is unclear that chances would serve as a guide to action for an agent whose credences conformed to some such principle but who did not update by conditionalization.

Yet, where an agent's epistemic state is represented by a *set* of probability functions, rather than a single probability function, it is not so clear how to model the process of updating upon evidence. It is still less clear how updating should be modeled when the evidence may itself be imprecise.

So far we have supposed, as is orthodox in the literature on imprecise credences, that the correct way of modeling an update upon some proposition D is to suppose that *each* probability function $cr_i(\cdot|\cdot)$ in S's initial representor is updated by conditionalization upon D to yield a function $cr_i^D(\cdot|\cdot) = cr_i(\cdot|\cdot\,\&D)$ (this is the update principle $(*)$ that we considered above). Where D is imprecise evidence concerning A, the idea is that S's resulting credences in A are unsharp because various functions in her representor 'interpret' D differently, in the sense that, for different values of i, the values of $cr_i(A|D)$ $(= cr_i(A|\top\&D))$ and so of $cr_i^D(A)$ $(= cr_i^D(A|\top))$ vary.

If this is correct then, where S's evidence D comprises F together with the proposition that the (possibly unsharp) *chance* for A conditional upon F is given by the (possibly non-singleton) set of values $\boldsymbol{x}$, as well as an admissible proposition E, her post-update credence is given by UPD:

(UPD) $$\boldsymbol{cr}^{\boldsymbol{ch}(A|F)=\boldsymbol{x}\&F\&E}(A) = \boldsymbol{cr}(A|\boldsymbol{ch}(A|F) = \boldsymbol{x}\&F\&E)$$

It follows from COND* and UPD that S's post-update credence in A is given by what I will call the 'Mushy Principle' (or MUSHYP):[40]

(MUSHYP) $$\boldsymbol{cr}^{\boldsymbol{ch}(A|F)=\boldsymbol{x}\&F\&E}(A) = \boldsymbol{x}$$

MUSHYP states that S's post-update credence in A (upon learning F, the set-valued chance for A, conditional upon F, and the admissible proposition E) will be represented by precisely the same set of values as is the chance for A conditional upon F. So, according to MUSHYP, credence is constrained by unsharp chance in precisely the way that it was argued that it ought to be in Section 5.

While I find MUSHYP to be extremely plausible, I'm less convinced by UPD. Recall that UPD incorporates the orthodox assumption that the correct way of modeling an update upon some proposition D is to suppose that *each* probability function cr_i in S's initial representor is updated by conditionalization upon D. The idea is that, when D is imprecise evidence pertaining to some proposition A, the different probability functions in S's initial representor 'interpret' D differently in the sense that the values of $cr_i(A|D)$ are different for different

[40]This name is inspired by the fact that unsharp probabilities are sometimes described as 'mushy'.

$cr_i \in \boldsymbol{cr}$. On this orthodox view, things should presumably work no differently when S's unsharp evidence D for A is a proposition – such as $\boldsymbol{ch}(A|F) = \boldsymbol{x}\&F\&E$ – implying an unsharp *chance* for A.

The reason that I find this orthodox approach slightly unsatisfactory is that we only get the correct post-update 'spread' for S's credences by building it in by hand. We must simply stipulate that any rational agent S must be endowed with a representor that contains just the right probability functions in the first place: that is, a set of probability functions that will 'interpret' any piece of imprecise evidence concerning a proposition A in just the right range of ways to leave S with a post-update representor that yields just the right unsharp credence in A. Where the imprecise evidence comes in the form of an imprecise chance, we must stipulate that S is endowed a set of probability functions that will 'interpret' this imprecise chance in just the right range of ways to leave S with a post-update representor that satisfies MushyP.

Admittedly, this orthodox approach is just a model: it is perhaps not intended to be altogether psychologically realistic (even to the extent that we approximate ideally rational agents). Still the example of imprecise chances suggests an alternative model which is more elegant in that it doesn't involve a rather ad hoc reverse-engineering of the desired post-update representor to determine the requisite pre-update representor, and then the stipulation that any rational agent must come equipped with a pre-update representor of just the right sort. The alternative model that I wish to suggest is rather permissive concerning the initial representor that a rational agent may come endowed with. It then suggests an update mechanism by which a wide range of plausible initial representors may come post-update to have the desired properties, such as satisfaction of MushyP.

Let's start with the case where S's imprecise evidence D concerning A comprises an imprecise chance for A. More specifically, suppose that it comprises (i) an admissible proposition F; (ii) the proposition that the (possibly unsharp) *chance* for A conditional upon F is given by the (possibly non-singleton) set of values $\boldsymbol{x}$; and (iii) any further admissible proposition E. In this case, there seems to be a very natural way of modeling S's update on her evidence. Roughly speaking, the idea is that a rational agent S's post-update representor comprises the set of probability functions that result from updating *each* of the probability functions $cr_i(\cdot|\cdot)$ in her initial representor upon *each* proposition $ch_j(A|F) = x_j\&F\&E$ such that $ch_j(\cdot|\cdot)$ is a probability function in the cadentor. In the case where S has a finite number N of probability functions in her initial representor and there is a finite number, M, of probability functions in the cadentor, S will thus end up with $N \times M$ (minus the number of instances in which a conditional probability of the form $cr_i(\cdot| \cdot \&ch_j(A|F) = x_j\&F\&E)$ is undefined) probability functions in her updated representor.

If each $cr_i \in \boldsymbol{cr}$ responds in an COND-like way to learning $ch_j(A|F) = x_j\&F\&E$,[41] then for each $cr_i \in \boldsymbol{cr}$ and for all $ch_j \in \boldsymbol{ch}$ we have:

$$(\dagger) \qquad cr_i(A|ch_j(A|F) = x_j\&F\&E) = x_j$$

And assuming that updating proceeds by conditionalization then, for each $cr_i \in \boldsymbol{cr}$ and for all $ch_j \in \boldsymbol{ch}$, we have:

$$(\ddagger) \qquad cr_i^{ch_j(A|F)=x_j\&F\&E}(A) = cr_i(A|ch_j(A|F) = x_j\&F\&E) = x_j$$

If S updates her initial representor $\boldsymbol{cr}$ in the light of the evidence $\boldsymbol{ch}(A|F) = \boldsymbol{x}\&F\&E$ by updating *each* of the probability functions $cr_i \in \boldsymbol{cr}$ upon *each* proposition of the form $ch_j(A|F) = x_j\&F\&E$ such that $ch_j(\cdot|\cdot) \in \boldsymbol{ch}$, and if each of the latter updates conforms to

[41] I admit that this assumption is somewhat ad hoc, and motivated only by the fact that it helps to yield the desired results about the nature of S's post-update representor. So both the orthodox model of updating one's representor in the light of evidence *and* the model that I'm currently articulating involve some ad hocness: I'll leave it to the reader to judge whether the ad hocness is worse in one case than the other.

(‡), then we get just the right mix of sharpening and dilation to ensure that S's post-update representor satisfies MushyP.

To see this, suppose that, for each probability function $ch_j \in \boldsymbol{ch}$, $ch_j(A|F) = x_j = x$. In other words, the chance of A conditional upon F is sharp: $\boldsymbol{ch}(A|F) = \{x\}$. Then since (‡) holds for all $cr_i \in \boldsymbol{cr}$ and for all $ch_j \in \boldsymbol{ch}$, it entails that, if S's initial credence is unsharp, but S then learns $\boldsymbol{ch}(A|F) = \{x\}\&F\&E$, then S's post-update credence in A will be sharp and equal to x. That's because, by (‡), for each $ch_j \in \boldsymbol{ch}$, all functions in S's representor, once updated on $ch_j(A|F) = x_j\&F\&E$, will take the value x post-update. This seems exactly as it should be (it is also what MushyP entails). For example, suppose you start off knowing just that all the balls in urn 1 are red or blue and that a draw will be made from urn 1, but you then learn that the draw will be made randomly and that there is a sharp chance 0.5 of a red ball being drawn conditional upon a ball being drawn randomly from urn 1. Then your credence that a red ball will be drawn should also be 0.5.

But where S learns $\boldsymbol{ch}(A|F) = \boldsymbol{x}\&F\&E$, where $\boldsymbol{x}$ is not a singleton, updating each probability function in her representor upon $ch_j(A|F) = x_j\&F\&E$ for each $ch_j \in \boldsymbol{ch}$ has a tendency to cause dilation. Specifically, for each credence function $cr_i(\cdot|\cdot)$ in S's initial representor, she ends up with a set of post-update credence functions of the form $cr_i^{ch_j(A|F)=x_j\&F\&E}(\cdot|\cdot)$.

Note that, where S has a finite number N of probability functions in her initial representor and there is a finite number, M, of probability functions in the cadentor, this need not entail that S's post-update credence in A is represented by a set of $N \times M$ (minus the number of instances in which a conditional probability of the form $cr_i(\cdot|ch_j(A|F) = x_j\&F\&E)$ is undefined) different *values*. This is because, for all N probability functions cr_i in S's representor, for each of the M probability functions ch_j in the cadentor, $cr_i^{ch_j(A|F)=x_j\&F\&E}(\cdot|\cdot) = x_j$. So S ends up with at most a set of M values representing her post-update credence in A.

Indeed, this value may even be considerably lower, since (in addition to the fact that conditionalization may not be well-defined in all cases), it may be that some or all of the probability functions in $\boldsymbol{ch}$ entail the same value x for the probability of A conditional upon

F. In the extreme case where, for all $ch_j \in \boldsymbol{ch}$, $ch_j(A|F) = x_j = x$, S's post-update credence will have the sharp value x. An example would be if all of the probability functions in the cadentor entailed the same probability of a given tritium atom decaying within the next 12.32 years (conditional upon some arbitrary proposition).

One worry about the present proposal – namely that updating on the imprecise evidence for A that comprises the proposition $\boldsymbol{ch}(A|F) = \boldsymbol{x}\&F\&E$ involves updating each probability function $cr_i \in \boldsymbol{cr}$ upon the proposition that $ch_j(A|F) = x_j\&F\&E$ for every $ch_j \in \boldsymbol{ch}$ – is that it appears to appears to make updating upon an unsharp chance a rather different operation to updating upon any other sort of evidence. After all, it was noted earlier that the orthodox way of modeling updating upon evidence D where S's epistemic state is represented by a *set* of probability functions, is to suppose that S's post-update representor is arrived at by updating *each* of the probability functions in S's representor by conditionalizing upon D. The idea was that, when D is imprecise evidence concerning A, various functions in S's initial representor 'interpret' D differently, resulting in S's post-update credence in A being unsharp.

I admit that it would be unfortunate to model updates upon unsharp chances of A differently from the way in which one models updates upon other sorts of imprecise evidence concerning A. But perhaps the present proposal for updating upon imprecise chances could be generalized to other sorts of imprecise evidence (i.e. the orthodox view of how to model the update of an agent's representor in response to imprecise evidence could be rejected entirely). The proposal that I have outlined concerning updates on imprecise chances basically involves the idea that there are various precisifications of an imprecise chance: specifically, each proposition of the form $ch_j(A|F) = x_j\&F\&E$ is a precisification of the imprecise evidence that $\boldsymbol{ch}(A|F) = \boldsymbol{x}\&F\&E$. The idea is that, where the imprecise evidence for A comes in the form of an imprecise chance for A, S's update should involve conditionalizing each probability function in her initial representor upon each precisification of the imprecise evidence: that is, upon each proposition of the form $ch_j(A|F) = x_j\&F\&E$ such that ch_j is in the cadentor.

Yet plausibly an imprecise chance of A is not the only sort of imprecise evidence concerning A that can be precisified in various ways. For example, suppose that A is the proposition that *there is a red ball in urn 1*, and that D is the proposition that *there are some balls in urn 1 and either all balls in urn 1 are red or all balls in urn 1 are blue.* What should your credence in A be upon learning that D? Plausibly it should be the unsharp credence $\{0,1\}$ (i.e. it should be represented by the binary set whose elements are 0 and 1). The 'precisification' account that I am proposing delivers this result. Specifically, there are two 'precisifications' of D, namely:

D_1 : There are some balls in urn 1 and all balls in urn 1 are red.

D_2 : There are some balls in urn 1 and all balls in urn 1 are blue.

Conditionalizing all of the probability functions in S's initial representor (for which the operation of conditionalization is well defined) upon D_1 results in the updated functions assigning probability 1 to A; conditionalizing them all upon D_2 results in their assigning probability 0 to A. Since the present, 'precisification' proposal is that S's updated representor contains the probability functions that result from both of these operations, her post-update credence in A is the unsharp one $\{1, 0\}$.

If this is the correct way of modeling updates upon imprecise evidence – namely, by supposing that S's post-update representor is arrived at by conditionalizing each probability function in her initial representor upon each precisification of her imprecise evidence – then modelling an update on an unsharp chance in terms of updating *each* of the probability functions in a person's representor upon *every* sharp-valued probability that constitutes the unsharp chance (these being the various 'precisifications' of the unsharp chance) would just be to treat the latter as a special case of updating on unsharp evidence more generally.

I'm not suggesting that there are no problems with this 'precisification' approach;[42] the orthodox approach remains a live option. But the precisification approach certainly strikes

[42]For one thing, it is not clear how to rigorously cash out the notion of a 'precisification' of imprecise evidence in the case where that imprecise evidence doesn't take the form of an unsharp chance. There are also some examples of imprecise evidence (e.g. that given by Elga (2010, 1)) where it's not clear what a 'precisification' of the evidence would amount to.

me as an elegant way of modeling updates in response to unsharp evidence that comprises an unsharp chance. This is because it represents a 'mechanism' by which agents with a wide variety of initial representors may arrive at a post-update representor that satisfies MushyP.

In any case, the key claim for present purposes is that, whatever the precise update mechanism, a rational agent's updated representor, once she updates on an imprecise chance, ought to respect MushyP. MushyP is supposed to be intuitive and, as such, not in need of explicit justification. Hájek and Smithson (2012, 38-39) claim 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?"

Admittedly, MushyP doesn't enjoy quite the same intuitive pull as the PP. But, then again, neither does the so-called 'New Principle' (NP) that was proposed by Lewis (1994) and Hall (1994) when it was realized that PP, in combination with Humean theories of chance, leads to contradiction. Rather MushyP, like NP, is an attempt to capture as best we can our pre-theoretical intuitions about the chance-credence connection, while simultaneously remaining reasonably true to other, sometimes competing, intuitions (such as that the rational response to imprecise evidence is imprecise credences) and to the implications of our best theories of chance (such as that chances derive from axioms that systematize the complete history of the world in the way that strikes as good a balance as possible between the competing theoretical virtues).[43]

[43]Cp. Lewis (1994, 489).

But even if MushyP enjoys a fair degree of intuitive support as an explication of the chance-credence connection, it remains to be shown in any great detail that the probability functions entailed by the tied-for-best systems play the chance role in guiding rational credence in accordance with MushyP. I do not have any completely satisfactory demonstration that they do so. This should come as no surprise, given that no entirely satisfactory explanation exists in the literature of why, in the case where there is a unique winner of the Best System competition (and on the assumption that credence must be sharp), the probabilities entailed by the winning system play the role that chance does according to either the PP or the NP.

But there is an additional obstacle to showing that the probability functions entailed by the tied-for-best systems play the MushyP role in guiding rational credence. The most promising attempts to argue that, in the case of a unique Best System, the (sharp) Best System probabilities play the Principal Principle role in guiding rational credence appeal to the *instrumental* rationality of conforming one's credences to the Best System probabilities (Hoefer 2007, Frigg and Hoefer 2010). Such arguments have a fairly well worked-out decision theory to draw upon in the form of Expected Utility Theory. However, there is no comparably well worked out theory of rational decision in the face of imprecise credences (see, e.g., Elga (2010) and Bradley (2013)).

In spite of these obstacles, some remarks are worth making that are of relevance to this issue. Some of these were already made briefly in Section 5 above. The argument considered there was that that systems earn membership of the set of tied-for-best systems by entailing conditional probabilities that are close to the actual relative frequencies. Systems that do not do so will be ill-fitting and not among those tied for best. Consequently, an agent who bets according to the conditional probabilities that are entailed by a member of the family of tied-for-best systems will do well in the long run. Moreover, an agent who knew the set of probability functions associated with the tied-for-best systems would lack a rational basis for choosing between them. So she appears to be rationally obliged to adopt a credence that takes the set of values corresponding to the probabilities entailed by the tied-for-best systems.

Of course, as we noted in Section 5, there will likely be some well-fitting systems that aren't among those that are tied-for-best, since they are too complex. Why should the agent not calibrate her credences to the probability functions entailed by these systems? This appears to be a genuine problem, but it is one that Best System analyst must face in any case (whether or not she takes Best System chances to be set-valued). The Best System analyst needs to have some account of why simplicity considerations ought to constrain rational credence. Otherwise she will have no answer to the question of why the probabilities of the overall Best System should play the chance role in guiding rational credence, rather than the probabilities of the best-fitting (but presumably highly complex) system.[44]

Alternatively, and to my mind more promisingly, the Best System analyst might admit that the Best System probabilities are imperfect players of the chance role in guiding rational credence, as captured by the Principal Principle (or indeed NP or MushyP),[45] but maintain (*contra* Lewis (1980)) that the chance-credence connection doesn't exhaust the chance role (cp. Loewer (2001), Arntzenius and Hall (2003), Schaffer (2003, 2007)). She might then argue that the probability function of the overall best system (for now, assume that it is unique) counts as the chance function because it plays *other* aspects of the chance role (besides guiding rational credence) better than the the probability function associated with the best *fitting* system.

For example, if one thinks that there is some sort of frequency-tolerance platitude concerning chance (cp. e.g. Armstrong (1983, 32); Loewer (2001, 613); Frigg and Hoefer (2013)), then this could motivate the view that the BSA probabilities are better players of the chance

[44] In this vein, Hoefer (2007, 583-7) mounts an argument that (much of the time) Humean chances serve as better guides to rational credence than the actual frequencies.

[45] Lewis thought that it is the Principal Principle that is "the key to our concept of objective chance" (Lewis 1994, 489). However, he admitted (ibid.) that Best System probabilities at best merely satisfy the NP, and so are imperfect players of the chance role. Lewis (ibid.) nevertheless claimed that the Best System probabilities are both the best players of the chance role, and that they are good enough players of that role to deserve the name *chance*.

role than the actual frequencies. The idea behind frequency-tolerance is that, while actual frequencies are evidence for chances (and while chances perhaps explain frequencies), nevertheless actual frequencies may diverge (arbitrarily?) from the chances. The actual frequencies themselves exhibit zero frequency tolerance, and so fail to conform to this putative platitude. Best System probabilities are constrained, via the theoretical desideratum of *fit*, not to diverge arbitrarily from all frequencies throughout the entire universe. That is, they may not exhibit arbitrarily large *global* divergence from the actual frequencies.[46] In this respect they are perhaps imperfect players of the chance-frequency aspect of the chance role. Nevertheless, the BSA allows that the Best System probabilities may exhibit at least some global divergence, and a great deal of local divergence from the actual frequencies (especially if this buys significant simplicity for the system that entails them), and are thus much better players of this aspect of the chance role than are the actual frequencies themselves.

It would seem that such arguments can be adapted to our purposes. If the defender of the BSA can admit that, on the assumption of a unique Best System, that the Best System probabilities don't play the Principal Principle (or NP) role as well as, for example, the actual overall world frequencies, but can emphasize some other aspect of the chance role (such as frequency-tolerance) that they play better then, relaxing the assumption of a unique Best System, we can admit that the set of all and only the probability functions of the tied-for-best systems doesn't play the MushyP role better than any other set of probability functions (such as certain sets containing the function $freq(X|Y)$ that, for all X and Y, entails the actual overall world frequencies for X given Y), but we can nevertheless argue that the set of (simplicity-constrained) probability functions associated with the tied-for-best systems plays other aspects of the chance role (e.g. frequency-tolerance) better than those sets of probability functions that are its rivals for the name *chance*.

A different sort of worry concerning the ability of imprecise probabilities to play the chance role in guiding reasonable credence seems to follow from arguments due to Elga (2010) and White (2010). They argue that having unsharp credences in response to unspecific evidence

[46]This is what leads to the *undermining* problem for the BSA (see Lewis (1994)).

is incompatible with perfect rationality. White (2010) focuses upon the phenomenon of *probabilistic dilation*, whereby imprecise credences in a proposition A, for which one has unspecific evidence, can force one to adopt imprecise credences in an apparently unrelated proposition B, about which one has highly specific evidence: for example, one might have knowledge of a sharp *chance* for B. The resulting imprecise credence in B appears irrational: for one thing, it appears to be in violation of the chance-credence connection, given one's knowledge of the sharp chance of B. On the other hand, Elga (2010) argues that imprecise credences permit an agent to make decisions that she knows to be dominated by (that is, to yield lower payoffs come-what-may than) alternative decisions that she knows to be available to her. Both White's and Elga's arguments appear to apply to those adopting unsharp credences in response to imprecise evidence that takes the form of an imprecise chance, as well as to those adopting unsharp credences in response to other types of imprecise evidence (it is the latter upon which they focus). If their arguments go through, then perhaps any player of the chance role in guiding rational credence must itself be sharp. For one thing, insofar as unsharp credences issue in irrational betting behavior, this threatens to undercut the 'instrumental' or 'consequentialist' argument that unsharp chances can play the MushyP role in guiding rational credence.

However, Joyce (2010)'s defenses of unsharp credences against Elga's and White's arguments are compelling. In response to White (2010), Joyce (2010, 299-307) argues that the circumstances that force dilation with respect to the proposition B are in fact circumstances in which one gains imprecise evidence that is relevant to B but not (only) because it is relevant to the originally-known sharp chance for B. In other words, the new imprecise evidence that one gains in such circumstances is evidence that is inadmissible *sensu* the Principal Principle and MushyP. *Given* this fact, it is not incompatible with perfect rationality to have a post-update credence in B that is unsharp (and to bet accordingly). In response to Elga, Joyce (2010, 315-316) argues that Elga considers only an impoverished set of betting rules that someone with imprecise credences might adopt. Joyce argues that, given a betting rule that is sensitive, not just to the *set* of probabilities that the functions representing one's credence entail, but also to other features that are common to every such function, one can avoid bets that one knows to be dominated. So imprecise credences about a proposition B (whether arrived at upon the basis of knowledge of an imprecise chance for B or some other sort of imprecise evidence concerning B) aren't irrational and (when combined with reasonable decision-making procedures) do not lead to unreasonable betting behavior.

In this section, I have proposed a principle – MushyP – concerning how unsharp chances ought to guide reasonable credence. I have suggested a mechanism for updating credences upon unsharp chances – one that involves, roughly speaking, an agent conditionalizing each probability function in her representor upon each chance in the cadentor – that leads agents starting with a wide range of initial representors to have post-update representors that respect MushyP. Finally, I have argued that the probability functions associated with the tied-for-best systems play the MushyP role in guiding rational credence reasonably well and that it is plausible that, when we take into account other aspects of the chance role, they are the *best* players of the chance role. I thus hope to have gone at least some way toward establishing that the unsharp Best System probabilities deserve to be called unsharp *chances*.

8. Special Sciences and the 'Better Best Systems' Analysis

Albert (2000, 2012) and Loewer (2008, 2012a) claim that probabilistic approximations to many of the generalizations of the special sciences are theorems of the Mentaculus. If the Mentaculus were the robustly best system for our world then it would follow, on the BSA, that the theorems in question are genuine probabilistic laws, and that the associated probabilities are genuine chances. But if the Mentaculus is tied for best with a number of other systems then, since the probabilistic special science generalizations that derive from these systems may diverge somewhat, we will be liable to get unsharp chances in the special sciences.

The view that probabilistic versions of the special science laws derive from some Mentaculus-like Globalist axiomatisation of statistical mechanics has been dubbed 'Statistical Mechanical

Imperialism' (Callender and Cohen (2010); see also Callender (2011) and Weslake (2013)). The view is one that has been challenged by Callender and Cohen (2010), Callender (2011), Dunn (2011), Frigg and Hoefer (2013), Frisch (2011, 2013) and Weslake (2013). In particular, it has been argued that there is not much evidence – let alone proof – that probabilistic versions of the special science laws can be derived in this way.

As we saw in Section 6, one alternative to the approach of Albert and Loewer is the view that Localist axiomatizations of SM are superior to their Globalist counterparts (Callender 2011, Frigg and Hoefer 2013). In Section 6, we noted Callender's argument that it simply doesn't follow from the fact that a uniform distribution over the sub-region of the universe's phase space associated with the low-entropy initial macro-state specified by the PH makes entropy increase for the universe as a whole very likely, that entropy increase in non-equilibrium sub-systems of the universe is very likely (Callender 2011, 100-101). If Callender is correct then, since the special sciences are themselves concerned with subsystems of the universe (e.g. ecology is concerned with ecosystems, meteorology with atmospheric systems, biology with organisms, economics with markets, etc.), it appears that one cannot infer probabilities for these systems (and probabilistic laws for the corresponding sciences) from a probability distribution over the initial conditions of the universe as a whole.

But, even if one is a Localist about SM, one might still think that probabilistic approximations to the generalizations of the special sciences derive from SM (while thinking that the standard SM probabilities derive, not from a probability distribution over some region of the phase space of the universe associated with its initial low-entropy initial macrostate, but rather from distributions over regions of the phase spaces of subsystems of the universe associated with *their* initial macrostates). As we saw in Section 6, it is plausible that there is a tie between various Localist axiomatizations of SM, with various of the tied-for-best Localist axiomatizations endowed with different probability functions. If so, and if the special science chances do indeed derive from SM, then it appears that, once again, we have reason to believe the special science chances to be unsharp.[47]

[47]Indeed, Callender's favored Localist approach to SM effectively treats SM itself as a special science (see Callender (2011, 96, 106, 110); cp. also Dunn (2011)).

Even if probabilistic approximations to the generalizations of the special sciences don't derive from SM, whether it is given a Globalist or a Localist axiomatization, it needn't follow that there are no special science laws or chances. One possibility is that the strength of a system that already entails SM probabilities is augmented, with minimal cost in simplicity, by the addition of special science laws (or axioms that entail them) to its set of axioms (see Frisch (2013); cp. also Frigg and Hoefer (2013)).[48] If such laws are probabilistic, then the probabilities that are entailed will be special science chances. And if there aren't uniquely best axiomatizations of the special sciences (more on this below), then there may well be tied-for-best systems for the universe as a whole that entail different special science probabilities, so that special science chances are unsharp.

An alternative approach to special science laws and chances, proposed by Cohen and Callender (2009) and Callender and Cohen (2010)[49] is to modify the BSA in an attempt to ensure that the simple, informative generalizations of the special sciences come out as genuine laws.[50] Cohen and Callender (2009) and Callender and Cohen (2010) call the resulting, modified BSA, a 'Better Best Systems Account (BBSA)' of laws.

Their proposal draws upon Lewis's observation that a system's simplicity depends upon the vocabulary in which it is expressed.[51] But rather than following Lewis in restricting the systems under consideration to those whose axioms contain only perfectly natural kind predicates, and rather than following the approach that we have been considering so far where naturalness-of-vocabulary-in-which-axioms-are-expressed is a competing theoretical virtue to be weighed alongside simplicity and strength/fit, their idea is that Best Systemhood should be taken to be *relative* to a set of kinds K (or predicates P_K). Relative to different sets of kinds, different axiom systems strike the best balance between simplicity, strength, and

[48]The resulting systems may, however, incur the cost that many of their axioms are formulated in imperfectly natural vocabulary. Whether they are nevertheless among the systems that are tied-for-best (at least given the limited precision of notions like simplicity and naturalness, and of the exchange rate between the theoretical virtues) will depend upon the strength/fit that this sacrifice of naturalness-of-vocabulary buys.
[49]For similar proposals, see Schrenk (2008) and Dunn (2011, 88-90).
[50]Of course, following the recent literature, I have already been assuming a version of the BSA that departs somewhat from the original in allowing axioms to be framed in imperfectly natural vocabulary. But Cohen and Callender's proposed modification to the BSA is more radical still.
[51]They also point out that strength and balance are vocabulary-relative (Cohen and Callender 2009, 6).

fit. A generalization is a law relative to K just in case it is a theorem of the Best System relative to K. In the case of a set of systems that are tied-for-best relative to K, Cohen and Callender claim that a generalization is a law relative to K just in case it is a theorem of *all* of the systems that are tied-for-best relative to K (Cohen and Callender (2009, 21); Callender and Cohen (2010, 440); see also Callender (2011, 111-112)).[52] Applying their BBSA to chances as well as laws, a probability is a chance relative to K just in case it is entailed by a probabilistic generalization that is a law relative to K (Cohen and Callender 2009, 27-30).[53]

The BBSA is particularly conducive to counting special science generalizations as laws. In particular, on the BBSA, the generalizations of a special science (such as biology, economics, or chemistry) count as laws of that science if they are theorems of the best system relative to the science's proprietary kinds or predicates (e.g. the biological, economic, or chemical kinds).[54]

[52]This is clearly analogous to Lewis's earlier construal of the BSA according to which, in case of ties, those theorems entailed by *all* of the tied-for-best systems count as laws (Lewis 1983, 367).

[53]Cohen and Callender (2009, 27-29) spend some time wrestling with the question of how different chances, each relativized to a different set of kinds K, for the same proposition A can play the Principal Principle (PP) role in guiding rational credence. But this issue appears acute for the proponent of BBSA only if she takes chances to be fundamentally *un*conditional. But the chances that the various sciences entail for any given proposition are inherently conditional. For example, think of the different chances that are entailed by SM, on the one hand, and by Newtonian mechanics, on the other, for a particular classical isolated system evolving to thermodynamic equilibrium (cp. Cohen and Callender (2009, 28)). The different chances that these theories entail are chances *conditional upon different propositions concerning the system.* Newtonian mechanics entails a chance conditional upon the proposition that the system occupies such-and-such a point in phase space, while SM entails a chance conditional upon the proposition that the system is in the non-point-sized region of phase space so-and-so. There is no conflict between divergent chances for a proposition *A conditional upon different propositions* $B_1, \ldots, B_n$. Principles like COND and COND* tell us precisely how rational credence is constrained by various existent conditional chances. In their later paper, Callender and Cohen (2010, 444n) give some indication of recognizing this.

[54]Cohen and Callender (2009, 10, 28) and Callender (2011, 106-112) treat thermodynamics and SM as at least on a par with the special sciences in the sense that they are likely upshots of Best System competitions conducted in their own proprietary vocabularies, which include predicates – such as *entropy* – that don't refer to fundamental natural kinds. Similarly, Dunn (2011, 80-81, 91) explicitly treats thermodynamics as just one special science among many (cp. Loewer (2012b, 15, 18)). Winsberg (2008, 884) objects that there is no distinctive proprietary vocabulary for SM: there is only the thermodynamic language and the microphysical language. Weslake (2013) suggests that the correct response is to see SM as a best system for the *conjunction* of the fundamental kinds and the thermodynamic kinds. Frisch (2013) (cp. also Frigg and Hoefer (2013)) suggests an alternative variant of the Best Systems approach that makes it plausible that the laws of the special sciences, together with those of fundamental physics, are part of a single 'big' (non-vocabulary relative) best system. This seems suited to accommodating axiom systems that draw upon multiple scientific vocabularies. Frisch's 'Big' Best System approach is likely to result in ties between systems

Callender and Cohen (2010, 437-438) and Callender (2011, 103, 111-112) suggest that the best axiomatization for various special sciences will include probability distributions over underlying state-spaces (these need not be phase spaces however). On their view, the probabilistic theorems generated by the resulting axiom systems will be probabilistic laws of the sciences in question, and the probabilities that those laws entail will be chances of that science. Specifically, Callender and Cohen's idea appears to be that the probability distributions over the underlying state-spaces of the systems that the special sciences seek to characterize can be justified in terms of the *frequency*, among the large number of such systems, of systems whose initial conditions lie in various sub-regions of that the underlying state space.[55] It is natural to construe this in a manner similar to Frigg and Hoefer's Localist approach to axiomatizing SM (Frigg and Hoefer 2013): the idea being that, if we consider systems whose state spaces have (roughly) the same dimensionality, then the precise actual initial conditions of those various systems will be distributed over that state space. A probability distribution *fits well* if it matches (or comes close to matching) the actual frequency with which the initial conditions of actual systems are to be found in the various sub-regions of the state space.

The trouble with this proposal – as was noted in our earlier discussion of Frigg and Hoefer's Localist approach to SM – is that it is extremely plausible that, in many cases, there are too few systems that have state spaces with exactly the same number of dimensions. *Too few* in the sense that there aren't enough to yield a frequency distribution of actual initial conditions within such a space that nails down a unique probability distribution as that

and unsharp chances for reasons very similar to those discussed in connection with Cohen and Callender's BBSA in the main text below.

[55]Like Callender and Cohen, Ismael (2009, 2012) is skeptical about the derivability of the probabilistic generalizations of the high-level sciences from a distribution over the possible microphysical initial conditions of the universe (a la Albert and Loewer), but argues that a probability measure over underlying state space is essential to their derivation (cp. Glynn (2010, 60-62)). Which measure is appropriate depends upon the relative frequencies with which macrostates are realized in various microphysical ways (Ismael (2009, 96); Ismael (2012, 433, 438); cp. Glynn (2010, 61)): a distribution will be preferable if it closely matches those frequencies. While Ismael doesn't commit to (and has "reservations about" – Ismael (2012, 432)) the Best System approach to laws and objective probability, she nevertheless takes a probability distribution over underlying state space to be part of the "objective content" (Ismael 2009, 91) of any theoretical package from which probabilistic high-level generalizations can be derived, and takes such high-level generalizations to be laws and the probabilities that they entail to be objective.

which strikes the robustly best balance between simplicity and fit. It is plausible that, instead, we have a large (perhaps continuous) range of probability distributions over such spaces that fit the actual initial condition frequencies within them reasonably well, with some of the better fitting (such as distributions that have peaks at each actual initial condition that lies within the space in question) being more complex, and the worse fitting (e.g. the uniform distribution) being simpler. If so, then axiomatizations of the special sciences that incorporate different such distributions will plausibly yield a set of tied-for-best systems that will entail different special science probabilities because of the different distributions over the underlying state spaces that they incorporate.

More generally, ties between special science systems that strike different and good, but not robustly *best*, balances between the theoretical virtues would not be surprising, especially given the limits of precision of the notions of simplicity, strength, and balance. In a discussion of what he calls 'Multiple Models Idealization' (MMI) in science, Weisberg (2007) points out that sciences commonly draw upon "multiple related but incompatible models" (Weisberg 2007, 645). He gives as examples the multiple models that ecologists have for phenomena such as predation, and the multiple models that meteorologists use to predict the weather (Weisberg 2007, 646). He notes that one particularly important justification for using multiple models "is the existence of *tradeoffs*" (ibid.). The idea is precisely that there are various theoretical desiderata, such as strength/predictive power and simplicity, and that "these desiderata . . . can trade off with one another in certain circumstances, meaning that no single model can have all of these properties to the highest magnitude" (Weisberg 2007, 646). Weisberg observes that "[o]ur cognitive limitations, the complexity of the world, and constraints imposed by logic, mathematics, and the nature of representation, conspire against simultaneously achieving all of our scientific desiderata." (Weisberg 2007, 647). He claims that this has been taken to justify the notion that "communities of scientists should construct multiple models, which collectively can satisfy our scientific needs" (Weisberg 2007, 647). He also claims that:

> "if tradeoffs exist between theoretically important desiderata in a particular domain, then we should not expect MMI to abate with further progress. These tradeoffs are consequences of logic and mathematics and thus present a permanent justification for MMI." (Weisberg 2007, 648)

It is notable that some of Weisberg's main examples are drawn from the special sciences, strongly suggesting that there fail to be unique best models relative to the proprietary kinds of these sciences. For those (such as Cohen and Callender (2009), Callender and Cohen (2010), Schrenk (2008), and Frisch (2011)) who are sympathetic to a Best System-style approach to special science laws and chances, it is natural to think that the various models that are equally good upshots of trading off competing theoretical virtues against one another are genuine tied-for-best systems for the sciences in question.[56] Where such systems come endowed with differing probability functions – as competing models in meteorology, ecology, etc. – very often do, then it is natural to regard the set of such functions as constituting an unsharp chance for the science in question.

Yet, as was noted above, Cohen and Callender say that in the event of ties the laws relative to K are those theorems common to all of the systems that are tied-for-best relative to K, and that the chances relative to K are the probabilities entailed by the laws relative to K. This seems to me to be the wrong thing to say, just as it would be wrong for the defender of the non-vocabulary-relativized BSA to deny that there are chances in the case of ties between systems. In particular, in the event of ties (and, for reasons already given, it seems that in worlds like ours ties are ubiquitous), it is not the case that there are no quantities that play the chance role in guiding rational expectations and decision-making. If, for example, I know that an ecosystem is in the initial ecological state E (where that state is characterized by

[56]It is plausible that many special science models don't alone constitute systems for the whole of the science in question. For example, the Lotka-Volterra predator-prey model is merely intended to capture one important aspect of ecosystem dynamics, and isn't intended as an axiomatization of ecology as a whole. At this point there are important questions that arise about how unified a special science like ecology is, and whether various models for specific ecological phenomena can be integrated into a single 'system' for ecology. But, even if axiomatizations of ecology as a whole can be produced, the underlying reasons for the existence of ties between models of specific ecological phenomena (namely tradeoffs between theoretical virtues) will also apply at the level of candidate axiomatizations for the science as a whole.

the values of various ecological variables) and I know that all of the tied-for-best systems for ecology entail a high probability ($\gg 0.5$), given E, for the bunny population growing at rate $> r$, and if I know nothing else of relevance, then it would be wise for me not to bet against the bunny population growing at rate $> r$ and indeed wise to bet in favor of it. This is so even if there is no unique probability x such that the tied-for-best systems all imply that the probability of growth at rate $> r$, given E, is equal to x, and so I am not rationally compelled to have a sharp credence x in a growth rate $> r$.

9. Conclusion

It has been argued that, if we take some variant of the BSA of laws and chances to be correct, then we have good reason to think that there are unsharp chances in our world. This is for the following reasons. *Firstly*, it is implausible that there is a single axiom system for our world that strikes the uniquely best balance between the theoretical virtues of simplicity, strength, and fit. It is also implausible even to think that there is a unique best system *relative* to the proprietary kinds of each science. Part of the reason for this is that notions such as simplicity and balance are imprecise. *Secondly*, it appears that the tied-for-best systems come endowed with different probability functions, which entail divergent probabilities for the same events (given the same chance setup). But, *thirdly*, it is not true that there is simply nothing that plays the chance role in (for example) guiding rational credence and decision when such chance setups are instantiated. Rather, the *set of probabilities* entailed by the set of probability functions associated with the tied-for-best systems plays this role, and so such sets of probabilities constitute unsharp chances.

Acknowledgments

I am extremely grateful to Radin Dardashti, Matthais Frisch, and Karim Thebault for a series of very detailed discussions of the topic of this paper. For helpful discussion of the ideas in it, I would also like to thank Claus Beisbart, Seamus Bradley, and Leszek Wroński. I gratefully acknowledge the support of the Alexander von Humboldt Foundation during my time at the Munich Center for Mathematical Philosophy, where I began work on this paper.

References

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

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

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

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

Beisbart, C. (2013). Good just isn't good enough – Humean chances and Boltzmannian statistical physics. In M. Galavotti, D. Dieks, W. Gonzalez, S. Hartmann, T. Uebel, and M. Weber (Eds.), *New Directions in the Philosophy of Science*. Dordrecht: Springer. Forthcoming.

Bradley, S. (2013). Weak rationality and imprecise choice. Manuscript.

Callender, C. (2011). The past histories of molecules. In C. Beisbart and S. Hartmann (Eds.), *Probabilities in Physics*, pp. 83–113. New York: Oxford University Press.

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

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

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

Earman, J. (2006). The "Past Hypothesis": Not even false. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 37*(3), 399–430.

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.

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

Frigg, R. and C. Hoefer (2010). Determinism and chance from a Humean perspective. In D. Dieks, W. Gonzalez, S. Hartmann, M. Weber, F. Stadler, and T. Uebel (Eds.), *The Present Situation in the Philosophy of Science*, pp. 351–371. Berlin and New York: Springer.

Frigg, R. and C. Hoefer (2013). The best humean system for statistical mechanics. *Erkenntnis*. Forthcoming.

Frisch, M. (2011). From Arbuthnot to Boltzmann: The Past Hypothesis, the Best System, and the special sciences. *Philosophy of Science 78*(5), 1001–1011.

Frisch, M. (2013). Why physics can't explain everything. In A. Wilson (Ed.), *Asymmetries of Chance and Time*. Oxford: Oxford University Press.

Glynn, L. (2010). Deterministic chance. *The British Journal for the Philosophy of Science 61*(1), 51–80.

Goldstein, S. and J. Lebowitz (2004). On the (Boltzmann) entropy of non-equilibrium systems. *Physica D: Nonlinear Phenomena 193*, 53–66.

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

Hájek, A. (2003a). Conditional probability is the very guide of life. In H. Kyburg Jr and M. Thalos (Eds.), *Probability is the Very Guide of Life: The Philosophical Uses of Chance*, pp. 183–203. Chicago: Open Court.

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*, 563–585.

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

Hall, N. (1994). Correcting the guide to objective chance. *Mind 103*(412), 505–518.

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

Hughes, R. I. G. and B. van Fraassen (1984). Symmetry arguments in probability kinematics. *Philosophy of Science (Proceedings) 1984.2*, 851–869.

Ismael, J. (2009). Probability in deterministic physics. *Journal of Philosophy 106*(2), 89–108.

Ismael, J. (2012). A modest proposal about chance. *The Journal of Philosophy 108*(8), 416–442.

Joyce, J. M. (2005). How probabilities reflect evidence. *Philosophical Perspectives 19*(1), 153–178.

Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives 24*(1), 281–323.

Lavis, D. A. (2005). Boltzmann and Gibbs: An attempted reconciliation. *Studies in History and Philosophy of Modern Physics 36*, 245–273.

Lewis, D. (1980). A subjectivist's guide to objective chance. In R. Jeffrey (Ed.), *Studies in Logic and Inductive Probability*, Volume 2, pp. 267–297. Berkeley: University of California Press.

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. New York: Oxford University Press.

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

Lewis, D. (1999). Why conditionalize? In D. Lewis (Ed.), *Papers in Metaphysics and Epistemology*, Volume 2, pp. 403–407. New York: Cambridge University Press.

Loewer, B. (2001). Determinism and chance. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 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. *Causation, physics, and the constitution of reality: Russell's republic revisited*, 293–326.

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

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.

Maudlin, T. (2007a). A modest proposal concerning laws, counterfactuals, and explanations. In *The Metaphysics within Physics*, pp. 5–49. New York: Oxford University Press.

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

Maudlin, T. (2011). Three roads to objective probability. In C. Beisbart and S. Hartmann (Eds.), *Probabilities in Physics*, pp. 293–319. New York: Oxford University Press.

Schaffer, J. (2003). Principled chances. *The British journal for the philosophy of science 54*(1), 27–41.

Schaffer, J. (2007). Deterministic chance? *The British journal for the philosophy of science 58*(2), 113–140.

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

Sturgeon, S. (2008). Reason and the grain of belief. *Noûs 42*(1), 139–165.

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

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

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

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy 104*, 639–659.

Weslake, B. (2013). Statistical mechanical imperialism. In A. Wilson (Ed.), *Asymmetries of Chance and Time*. Oxford: Oxford University Press.

White, R. (2010). Evidential symmetry and mushy credence. *Oxford studies in epistemology 3*, 161–186.

Winsberg, E. (2004). Can conditioning on the "Past Hypothesis" militate against the reversibility objections? *Philosophy of Science 71*(4), 489–504.

Winsberg, E. (2008). Laws and chances in statistical mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 39*(4), 872–888.

Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.