FORMAL CONSISTENCY OF THE PRINCIPAL PRINCIPLE REVISITED

ACCEPTED IN THE EUROPEAN JOURNAL FOR PHILOSOPHY OF SCIENCE

Leszek Wroński*, Zalán Gyenis, Mariangela Zoe Cocchiaro

Jagiellonian University

January 28, 2025

Abstract

We rigorously describe the relation in which a credence function should stand to a set of chance functions in order for these to be compatible in the way mandated by the Principal Principle. This resolves an apparent contradiction in the literature, by means of providing a formal way of combining credences with modest chance functions so that the latter indeed serve as guides for the former.

Along the way we note some problematic consequences of taking admissibility to imply requirements involving probabilistic independence. We also argue, *contra* [12], that the Principal Principle does *not* imply the Principal of Indifference.

Keywords: Principal Principle, chance, credence, admissibility, consistency, Principle of Indifference.

1 INTRODUCTION

The Principal Principle ("the PP") has been proposed as a condition of rationality on agents with credences about chances. It requires, roughly, and with caveats we will investigate below, that, for any proposition A, given that the chance of A is x, the agent's credence in A be x. While David Lewis proposed it in the context of his general philosophical framework of Humean Supervenience, and in fact intended it to explicate the core of the concept of chance, the main idea had been around for a long time. The idea in question – of a direct inference principle which identifies values of some probabilistic credences with values of objective chances or frequencies – goes by various names and has been articulated in many different ways. Peirce dubbed it "probable deduction" [24]. Some logicians refer to it as a "statistical syllogism".[1] But out of the many alternatives the PP is quite likely its most widely known example and is currently one of the few generally accepted norms of Bayesian rationality [26].

The authors writing about the PP seem to share the assumption that both the credences of rational agents and the objective chances are classical probabilities; we will also require that. A rational agent's credence (degree of belief) function is, then, a

*leszek.wronski@uj.edu.pl

[1] See, among others: [33, 28, 29, 14, 15, 16, 27, 32].

classical probability function.[2] Postulating the PP means that we require more: not just any probability functions are permitted as rational, but only those in which credences are coordinated in a particular way; where, specifically, *conditional* credences achieve the values uniquely determined by the two propositions involved (e.g., by A and *the chance of A is .3*).

For decades now, the philosophical community has discussed the credences satisfying the PP, without making sure that such probability functions exist (in other words: whether the PP is consistent[3]). As far as we know, Lewis himself never produced an example of such a function in writing. Perhaps most of the authors involved thought the issue was trivial, and that one could always suitably extend the agent's probability space as needed. However, the lack of attention given to PP's consistency may seem surprising when we note that the principle is, indeed, still stronger a condition than just the coordination requirement outlined above. Not only should a rational agent's conditional credence (at some fixed time) in A given that *the chance of A is* .3 be .3; this credence should supposedly be unperturbed if we choose one of a host of propositions "admissible" at that moment, and append it to the given chance proposition. If, in other words, E is an admissible proposition, then the PP requires that one's credence in A given E and *that the chance of A is* .3 be .3. Assuming that Lewis and his followers can suitably fix the domain of admissible propositions, the PP seems to be an even more substantive condition on probabilistic credence functions and the question arises again as to whether functions satisfying it even exist in the first place.

One could entertain the thought that, as guides to credences, some chance functions are more forthcoming than others. That is, that there might indeed be some chance functions which, if the propositions about their values are in the domain of the given agent's credence function, allow for the coordination required by the PP. This would be enough to show the consistency of the PP in probably the weakest useful sense. Perhaps, though, in the case of some *other* chance functions such a coordination was impossible. A newcomer to the field might be quite surprised to see that, after four decades of discussion, the literature on the problem of PP's consistency seems to be contradictory. A group of authors, Lewis among them, believes that the situation just described in fact occurs: only so-called "immodest" chance functions can play the role of guiding our credences as the PP mandates.[4] A different group (see e.g. [1] and [9]), after approaching the issue with formal rigor, concluded that any credence can be coordinated with any chance function in the way required by the PP. One goal of the current paper is to address this seeming contradiction. Along the way we will contribute to the formal investigations regarding the consistency problem. However, perhaps the more surprising contribution stems from our discoveries concerning the notion of admissibility: we argue that, if we put the probability spaces satisfying the PP under careful scrutiny, we should arrive at the conclusion that the notion of admissibility should either be abandoned or given a different formal treatment than the one usually adopted.

The paper is structured as follows. Section 2 introduces the PP in roughly the same semi-formal way as the one used by Lewis. In Section 3 we point out surprising restrictions on agent's credence functions stemming from the use of the notion of admissibility as intended by Lewis and, it seems, a large portion of writers on the topic. In Section 4 we tackle the contradiction just outlined, offering our take on the consistency problem. The style of the body of the paper is intended to be semi-rigorous.

[2] A probability space is a tuple $(W, \mathcal{F}, P)$, where W, the sample space, is a non-empty set, $\mathcal{F}$ is a sigma algebra over W, and P, the probability function, is a sigma additive mapping from $\mathcal{F}$ into $[0, 1]$, such that $P(W) = 1$.

[3] Among the few exceptions are the papers [1, 9]; see below.

[4] For a concise formal paper exploring this topic, see [25] (but do read Section 4 below); the issue stems from a philosophical problem outlined in [18].

All formal details, including the definitions of the employed notions and proofs of theorems, are in the Appendix. The readers interested only in the PP's consistency may safely skip Section 3.

Let us spoil the takeaway. Discussing the PP's consistency requires determining:

1. whether the probability spaces in question exist; and
2. if they do, whether they are fit for their intended philosophical role.

As for point 1, the answer is "yes"; we hope that after reading this paper the reader will agree that we now understand more about what such probability spaces look like. As for point 2, the answer is "no, if we do not fundamentally revise our understanding of admissibility, or remove any mention of admissibility from the formulation of the PP".

2 THE INITIAL SETUP

In a typical formulation, the PP requires from an agent with a credence function[5] P, that regarding a proposition A, assuming that X says that the chance of A is x,

$$P(A \mid XE) = x, \tag{1}$$

provided E is "admissible". We will focus on admissibility in Section 3; for now, the following gloss from Lewis himself should suffice: "Admissible propositions are the sort of information whose impact on credence about outcomes comes entirely by way of credence about the chances of those outcomes" [17, p. 92].

This formulation is still present in the literature, having been recently used e.g. by [12] in an influential article on the relationship between the PP and the Principle of Indifference. We shall adopt it as one of the two main formulations for the current paper, with implicit universal quantification over A, E, and x, and with the caveat that the proofs will eventually require us to be more rigorous. We do have to note, though, the existence of at least two additional ways in which the Principle has been expressed.

One of these is due to Lewis himself and requires us to refer to:

- "the complete theory of chance for world w", T_w;
- "the complete history of world w up to time t", H_{tw};
- "a reasonable initial credence function", P;
- "the chance distribution that obtains at t and w", ch_{tw};

and, given all that, to say that the PP requires that for any time t, world w, and proposition A,

$$ch_{tw}(A) = P(A \mid H_{tw}T_w).$$[6]

We will not use this formulation in the current paper since our goal here is to engage the philosophical discussion of the PP with rigorous formal results, to which this way of putting things does not lend itself easily.[7] We do not have information about the language in which the history of the world and its theory of chance will be specified: it is probably supposed to be that of the Best System (see Section 2 of [3]). The theory T_w is supposed to consist of all conditionals "from history to chance" that hold at w[8].

[5] This P is usually taken to be the agent's *initial* credence (see e.g. the persuasive arguments in [22]). This matter is orthogonal to our discussion and so will not be covered here.

[6] All quotations come from [17, p. 97]; we have introduced small notational changes.

[7] Note, however, the opposing opinion of [30], who presents a "proof" of the PP in this formulation.

[8] [17, p. 97].

The matter of the logic of conditionals is difficult in itself, and doubly so if we do not know the language in which they are expressed. In any case, some more extended discussion is needed before we assume that H_{tw}'s and T_w's "behave" formally in a way that would allow them to be in the domain of a probability function.[9] If we adopted this formalism and wanted to propose a rigorous mathematical argument, we would need to add substantial formal assumptions that would perhaps not be justified according to Lewis and others who had used it. Therefore it will be safer to avoid the "history and theory of chance" way of phrasing the PP from now on.

The first way of formulating the PP that we've introduced crucially requires the use of some proposition which specifies the chance of some other proposition. We will call propositions of this sort *local chance propositions*; in the literature on the PP they are usually referred to by the capital letter X. They should be contrasted with propositions which specify the whole chance function, which we will call *global chance propositions*.[10] A part of the PP literature uses global chance propositions in its formulation of the principle. Informally, the principle then relates, for each proposition A and chance function ch, the credence in A and the credence that the actual chance function is ch, so that the credence in the former given the latter is $ch(A)$. The reference to a global chance proposition C_{ch} ("that the chance is given by the function ch") appears e.g. in [25]; there (with small notational changes), the Principle is given as relating a time-indexed credence P^t with evidence the agent has gathered up to t as

$$P^t(A \mid C_{ch}) = ch(A \mid E_t), \tag{2}$$

with the original Principle recovered as the special case for tautologous evidence and initial credence P:

$$P(A \mid C_{ch}) = ch(A). \tag{3}$$

The similarity of (1) and (3) should be evident. One difference is that the former concerns the chance proposition as specifying only the chance of A. Another is that the latter omits any reference to admissibility. Why this might be a good choice will be one of the topics of Section 3 below. Lewis himself used a similar formulation of the PP in his 1994 paper. This will be the second formulation of the PP we will use here; eventually, we will touch the issue of the relationship between the two.[11]

When discussing the consistency problem, we will, in a nutshell, be considering whether a credence function P on some body of propositions $\mathcal{F}$ can be combined with a set CH of chance functions in the right way. By this we mean that the resulting credence function P', defined both on $\mathcal{F}$ and on the chance propositions relating to the chance functions from CH, agrees with P on the members of $\mathcal{F}$ and meets the requirements given by the PP: then we will say that P is *CH-PP-compatible.* The definition is somewhat intricate (especially since we will consider two variants of the

[9] Note that this is true even if we straightforwardly identify propositions with sets of possible worlds.

[10] Of course, it might very well happen that a local chance proposition *is* a global chance proposition (if propositions are sets of possible worlds), or at least it is true at the same set of worlds. (Suppose there are three worlds, and only one of them, w_1, is governed by a chance function, ch^1, which says that the chance of A is .2 (according to the two other chance functions this chance is different). Then if we consider the global chance proposition C_{ch^1} and the local chance proposition X defined as *that the chance of A is .2*, we will see that they are both true at exactly w_1. If propositions are sets, then both of these propositions will literally be the singleton $\{w_1\}$.) Also, depending on one's preferred metaphysical account of chance, it might very well be the case that every proposition is in some sense a chance proposition: if a proposition is true at the set of worlds S, then the actual chance function belongs to the set of the chance functions governing the worlds from S. However, in this paper we aim to stay away from metaphysics, and so will not discuss such issues further.

[11] Let us note one doubt one might have about expressing the Principle as (2): the PP was intended to be a synchronic rule for coordinating one's credences with one's credences about chances, while (2) makes it look as a belief update rule, a constraint on agents learning from evidence [37, p. 508].

compatibility relation corresponding to the two formulations of the PP mentioned above) and we decided to leave the details out from this introductory Section (please consult Definitions B.1.7 and B.1.8 in the Appendix, and in particular section B.1 which collects the main notions employed in those Definitions). However, the gist just given should be enough to see that there are various questions that could be asked regarding the consistency problem, which employ different quantification. For example, is every credence function CH-PP-compatible with every CH? If not, is every credence function CH-PP-compatible with *some* CH? What, then, are the necessary and sufficient conditions on P and CH so that P would be CH-PP-compatible? These are the issues we discuss in Section 4.

Before we talk about admissibility, let us get one more mundane thing out of the way. Typically, we speak about a probability measure P understood as an element of the probability space $(W, \mathcal{F}, P)$. Usually, if W is finite, the $\mathcal{F}$ is its power set.

3 BEING SERIOUS ABOUT ADMISSIBILITY

The Principal Principle says that a credence in a proposition A given some proposition E and a chance proposition which states that the objective chance of A is x should be x, providing that the conditional probability is well-defined and E is "admissible" at t. As even such a rough formulation makes clear, some account of admissibility is crucial for the viability of the Principle: its scope and power depend on it. If any piece of evidence is admissible, then the Principal Principle applies in any setting but is inconsistent [17, p. 92]. If nothing is admissible, then the Principle is vacuous. For perhaps the easiest source of examples of inadmissibility, consider the cases where E entails A: then we cannot demand that $P(A \mid XE) = x$ (for x other than 1) on pain of contradiction with the fact that P is a classical probability function, and so $P(A \mid XE) = 1$ (if defined at all)[12]. One of the roles of admissibility is to help us restrict the scope of the PP so that it does not apply to these cases. Anyway, the authors employing admissibility in their discussions of the PP have to believe, of course, that admissibility can indeed happen; we believe it a fair, if vague, assessment that according to them admissibility **is not very rare**. Still, we have not found in the literature a controversy around the PP holding only very rarely.

In spite of the importance of the notion, however, spelling out what admissibility consists of has turned out to be a controversial undertaking. Apart from the criterion that deems inadmissible a proposition E whenever it entails A as per the previous example, even Lewis's sufficient (or almost sufficient) conditions for admissibility have not been exempt from doubt.[13] In this essay we will not attempt to spell out a substantive account of the notion that would tell us exactly *which* propositions are inadmissible [34].

For our purposes it is enough to assume the informal characterization of the notion that comes from Thau's work [31] and on which many scholars [19, 13] have converged together with a formal rendering of it that is just as popular [11, 34, 21, 12, 35]. According to the rough characterization in question, a proposition is admissible at t "if it does not provide direct information about the outcome of a chancy event that occurs subsequently to t" [31, p. 500]. A formal counterpart of this rough characterization of

[12] More on this in Section 3.1 below.

[13] [23] has recently argued that not all historical propositions are admissible, and from a Humean perspective even hypotheses about objective chances might not be such (see, for example, Chapter 3 of [36]).

admissibility is the idea that a necessary[14] condition for a proposition to be admissible at t is probabilistic independence: when X says that the chance of A is x, then if some E is to be admissible, then $P(A \mid XE) = x$ ([35]). Since by the PP (assuming, noncontroversially, that tautologies are admissible) $P(A \mid X) = x$, this means that $P(A \mid XE) = P(A \mid X)$; that is, in the induced probability space[15] with the measure P_X, A and E are probabilistically independent: $P_X(A \mid E) = P_X(A)$. Catchphrase: **admissibility implies independence.** (Cf. [35], p. 959: "failure of probabilistic independence is a sure-fire way for admissibility to fail".) This way of thinking assumes that admissibility is parametrized only by time; we should note here that it will soon become clear that this idea is problematic, and if admissibility is to be fruitfully retained in the formulation of the PP, it should be relativised to more parameters.

We will introduce the problems with the idea that admissibility implies independence by means of a few examples involving conditional probability in small structures. Imagine, first, that you are throwing a fair, standard six-sided die and you are interested in whether or not the result of the next throw is in the set $\{4, 5\}$. The die is six-sided; therefore the chance of A (the proposition that the result of the next throw is either a 4 or a 5), $ch^6(A)$, is $1/3$, since there are six atomic events, each with the chance $1/6$ of occurring, two of which make A true. It is quite easy to find an event probabilistically independent from A: for example, if B is "the result of the next throw is an odd number", then $ch^6(A \mid B) = ch^6(A)$, since B is irrelevant for the chance of A: the chance of A given B is exactly the same as the chance of A.

However, things change drastically as soon as we tweak our example and consider a scenario in which you are still interested in obtaining a 4 or a 5, but your fair die is now five-sided (e.g., you use a six-sided one and reroll sixes until you get something else) and the chance of A, $ch^5(A)$, is now $2/5$ (again, due to there being five atomic events, each with the chance $1/5$ of occurring). Try as you might, you will not find a nontrivial event B such that $ch^5(A \mid B) = ch^5(A)$. In fact, if ch^5 is the uniform measure assigning $1/5$ to all atomic events, then there are *no* pairs of nontrivial events independent by the light of ch^5. This is an instance of a general phenomenon:

3.0.1 Fact. Assume $(W, \mathcal{F}, P)$ is a finite probability space with P uniform. If the number of atoms of $\mathcal{F}$ is a prime, then there are no events $A, B \in \mathcal{F}$ such that:

- both A and B are nonempty;
- both A and B are nonuniversal;
- $P(A \mid B) = P(A)$.

For a proof, see [6]. (Still more is true: even if $|W|$ is a composite number (as is typical in certain applications, where the probability space is a product of spaces), if $|B|$ is relatively prime to $|W|$, then no non-trivial A independent of B exists. For a proof see [6, Theorem 2].)

Notice that this has important consequences for anyone willing to consider the PP as possibly holding in finite probability spaces and attempting to follow Lewis in employing the notion of admissibility. If the agent's probability space is such that a local chance proposition X is true at a prime number of worlds, while the agent is indifferent as to which of those is the actual one, in fact we have to conclude that there is no admissible E: it is impossible to find non-trivial[16] A and E so that $P(A \mid X) = P(A \mid XE)$.

[14] Some scholars in the literature [11, 34] take it to be sufficient as well, but there is disagreement on this [21, 35].

[15] For a probability space $(W, \mathcal{F}, P)$ and an event $X \in \mathcal{F}$, the probability space induced by the event X is defined to be the probability space $(X, \mathcal{F} \upharpoonright X, P_X)$, where $\mathcal{F} \upharpoonright X$ is the algebra $\mathcal{F}$ restricted to X (also called the relativization of $\mathcal{F}$ to X, see [2, Def. 1.5, p. 132]), and $P_X(\cdot) = P(\cdot \mid X)$. When writing P_X, we assume that $P(X) > 0$.

[16] That is, different from the full and the empty set, but also different from X.

For an example of this, consider the above two-dice example. Suppose you're told that four things can happen, each with probability[17] $1/4$:

- a throw of a fair, standard six-sided die;
- a throw of a fair six-sided die with two sides labelled "4" and none labeled "5";
- a throw of a fair six-sided die with two sides labelled "5" and none labeled "4";
- a sequence of throws of a fair six-sided die terminating when a result different from "6" is obtained.

Notice that in the first three options the chance of A is $1/3$, and in the last one it is $2/5$. Suppose that each option corresponds to a possible world in the set on which the credence P is defined. The crucial thing is that if X is the local chance proposition "the chance of A is $1/3$", then X is true at a prime number of worlds, so if P indeed assigns $1/4$ to each of the four above options, there will be no non-trivial E such that $P(A \mid XE) = P(A \mid X) = 1/3$. If admissibility implies independence of this sort, then the inescapable conclusion is that no non-trivial E is admissible. This might be a reason to relativise admissibility not only to times (with which credence functions might be, implicitly or explicitly, indexed), but also to chance propositions. In an example in which the first three above options stay as they are, but the local chance proposition $Y =$ "the chance of A is $2/5$" is true at **four** worlds, there could be a proposition E such that $P(A \mid YE) = P(A \mid Y) = 2/5$, even though there would still be no E such that $P(A \mid XE) = P(A \mid X) = 1/3$. There would then exist an E for which it would seem natural to say that it is, at that moment, admissible relative to Y, but inadmissible relative to X.

In any case, in finite spaces, we should be wary about considering any proposition as admissible *tout court*: the existence of even a single nontrivial proposition of this sort implies that **for any real number x and for any proposition A, the number of worlds at which the proposition "that the chance of A is x" is true is not prime.** We do not think this surprising restriction has been noted in the literature. The more general issue, however, is to which set of parameters admissibility should be relativised so that one could retain some hopes for the notion's fruitfulness.

3.1 ADMISSIBILITY: PARAMETRIZED?

Lewis assumed that admissibility is admissibility "at t", that the credence function is implicitly indexed by a moment, and whether a proposition is admissible or not depends on the choice of that moment (which determines, for instance, whether that proposition is about the past or not). He seems to have changed his mind in [18] due to philosophical arguments concerning chance by [31]. However, let us note that a simple formal reasoning should persuade one that the original approach was untenable. That is, admissibility is not just admissibility "at t", but something more akin to "at t, with respect to A".[18]

We will proceed by *reductio*. Assume the credence function P is implicitly indexed by the time t, and suppose admissibility is admissibility "at t". Suppose E is admissible. Take an x different from 1 and consider the proposition X saying that the chance of E is x. Assume $P(XE) > 0$. Then:

[17] Take these probabilities to be well-considered *credences* of a rational Bayesian agent, to be contrasted with the *chances* involved in actual die-throwing. We would like to thank one of the anonymous reviewers for pressing us on this point.

[18] This is not an original insight; see e.g. [31] and [10]. Note also that if you believe the PP only concerns *initial* credences, when reading this subsection you of course need to fix the value of t as the initial moment of the epistemic life of the agent you are considering.

- $P(E \mid XE) = x$, by the PP;
- $P(E \mid XE) = 1$, by the fact that P is a classical probability function.

Therefore $x = 1$. Contradiction.

There are, it seems, three ways out of this which would not have us abandon anything crucially important such as P being probabilistic or satisfying the PP. We can, first, conclude that $P(XE) = 0$. But this assumption is made at the point where we have chosen an x different from 1, ending with the proposition X. If this move was to work, then, it would have to be the case that for any x, if X is the proposition that the chance of E is x, then $P(XE) = 0$. However, it is very natural to assume that these Xs form a partition (since every possible world gives rise to some chance of E), and so we cannot escape the conclusion that $P(E) = 0$. Our agent would then be forced to hold 0 credence in any admissible proposition.

The second move could be to say that admissibility "admits of degree", and then say that if E is, say, *almost perfectly* admissible, then $P(E \mid XE)$ could diverge a little from x. One could cite the somewhat *ad hoc* remarks from Section 8 of [18] in support of this position. Note, however, that we cannot put much hope into this as a tool for removing the above contradiction in general; it will not work if x is significantly lower than 1. The idea that the PP might in general be expected to hold only loosely will be considered as "Option 3" on p. 10 below.

The only solution, then, seems to be the third option of outright *denying* that E is admissible. Then admissibility, as parametrized only by time, is empty. (Or, more carefully, the only admissible propositions are those for which chance is undefined in at least one possible world.) This would make the notion useless. To avoid this, we should then consider propositions to be admissible, at t, *with respect to other propositions*: and notice tangentially that no proposition is admissible with respect to itself.

At this point one might think that, assuming we fix some time t as the one at which the credence P is considered, relativisation of admissibility to propositions is enough. Instead of asking "is E admissible?" we should consider questions of the sort "is E admissible with respect to A?". A positive answer might presumably require, then, that for any proposition X stating the chance of A, $P(A \mid XE) = P(A \mid X)$. However, note that admissibility so understood would be a very strong condition: the existence of even a single chance proposition X true in a prime number of possible worlds of which none are deemed privileged by the agent (whose credence is thus uniformly distributed among them) makes E inadmissible with respect to A.

We have already seen an instance of this phenomenon. Our dice example illustrates a case where one chance proposition has prime cardinality, hence admits no probabilistic independence: there are no nontrivial events A and E which would be independent in the induced space $P_{C_{ch^5}}$. However, the other chance proposition in that scenario, C_{ch^6}, allows the existence of such propositions: indeed, E, "the result is even", is probabilistically independent from A, "the result is in $\{4, 5\}$", in the space $P_{C_{ch^6}}$. There, getting to know E does not influence your credence in A, and in particular does not do it while bypassing your credences about chances. It seems, then, that there is no obstacle to regarding the E as admissible with respect to A and C_{ch^6}, and as inadmissible with respect to A and C_{ch^5}. If we are using local chance propositions, we may say that admissibility is always relativised to (time and) a chance proposition X about the chance of A. If we wish to use global chance propositions, we need three parameters: time, a proposition A, and a chance proposition C_{ch}.

3.2 THE GENERAL RARITY OF ADMISSIBILITY

We believe the prime-number-related issue with admissibility is something to be concerned about. The issue does not arise only in small spaces, of course. Fix some A and a local chance proposition X about A such that you're indifferent as to which of the possible worlds in X is the actual one. If your X has 4482 atoms, there will be propositions admissible with respect to it and A; if it has 4481 or 4483 atoms, there won't be any. We find this puzzling to the extent that we would indeed be fine with abandoning the admissibility talk whatsoever—or taking this as an argument to use infinite spaces. However, let us notice that the above phenomenon, that is, the total lack of pairs of nontrivial independent events, which arises in the very specific spaces to which we've pointed, has a perhaps more important general companion: the larger the space, the 'rarer' the independent pairs.

To be precise: as the size of a finite uniform probability space increases, the probability that two (uniformly randomly picked) events are independent converges to zero. In large finite uniform spaces, then, most pairs of events are not independent. For details, including the proof of this result, see Appendix A.1. Admissibility, in other words, can be described as 'rare' in such spaces.

3.3 ADMISSIBILITY, DEFEATERS AND THE PRINCIPLE OF INDIFFERENCE

We can apply the above discussion to partially assess the argument from [12] to the effect that the PP implies the Principle of Indifference (PoI). (Partially, at this point, since later we will show that the argument collapses in general—by presenting a counterexample, that is, a probability space satisfying the PP in which the PoI fails;[19] see Section 4 and, for some detailed exegesis, Appendix C.).

Instead of talking about admissibility, the authors speak of "defeaters": "We shall take the claim that E is not a defeater to hold just when $P(A \mid XE) = x = P(A \mid X)$" [12, p. 124]. We have shown that in some cases such non-defeaters simply do not exist. The argument needs at least one to exist to get off the ground. The preliminary verdict is, then, that the argument needs an additional assumption (to the effect that if X is true at a prime number of worlds, then P is not uniform, since otherwise no non-defeaters would exist). That said, we will provide a counterexample to the argument anyway below.

3.4 ADMISSIBILITY: THE TAKEAWAY

Given the above, what should we do with admissibility and the role it may play in the formulation of the PP? It seems we have three options.

Option 1: abandon. Let us formulate the PP without any mention of admissibility. The PP will then require from an agent with a credence function P, regarding a proposition A, assuming that the local chance proposition X says that the chance of A is x, that $P(A \mid X) = x$; a version using global chance propositions will also be easily obtainable. While the original formulation may seem to be logically stronger (since this one is a special case, implicitly using a tautologous proposition after the conditioning bar as an example of an admissible one), we believe this is indeed only a seeming, not reality, due to the murkiness of the notion of admissibility. This is the option we prefer.

Option 2: go infinite. If prime numbers seem to be a problem, and admissible propositions are harder and harder to come by as the cardinality of the space rises, perhaps it would be fruitful to demand that every probability space satisfying the PP

[19] Note that the example will come with a *caveat*.

be infinite. This invites a host of issues which taken together would surely deserve a separate paper. Spelling out the PP in infinite contexts is decidedly nontrivial and there are many variants to be considered.[20] It would be natural, for example, to consider local chance propositions according to which chances belong to *segments* of real numbers[21], instead of *being equal* to a real number. If X_x is "the chance of A is x", then an agent might have a credence such that for any x their credence in X_x is 0, and so $P(A \mid X_x E)$ is undefined (regardless of what the E is); however, at the same time their credence in Y, that is, say, "the chance of A belongs to $[.1, .2]$" might be positive. Such an agent might satisfy the condition $P(A \mid Y) \in [.1, .2]$, which, suitably generalized, would perhaps lead us to some principle intuitively in the spirit of the original PP.

Option 3: go imprecise. Perhaps the PP shouldn't be expected to work *perfectly*: we should accept slight deviations from strict identities. In Section 8 of [18] Lewis explicitly considers "degrees of admissibility". It is not obvious how one should think of them. For example, [21] takes the position that if E almost satisfies the philosophical *desideratum* of not revealing anything about our future history, then (assuming Y says "the chance of A is y") the value of $P(A \mid YE)$ may be constrained to y "even though E is not perfectly admissible" [21, p. 53]. For Masterton, then, the imperfectness of admissibility of a proposition has no influence on the strict identities between relevant probabilities. However, Lewis himself writes that in such cases these identities should hold "if not exactly, [then] at least to a very good approximation" [18, p. 486]; the PP "applied loosely will very often come very close" (*ibid.*). Such a claim is problematic for a number of reasons. The lack of any argument for it is one; it might be intuitive for some, but it would certainly need to be backed up by a theorem. Moreover, if admissibility is to be a nonempty notion, then agents whose credence is uniformly distributed among a prime number of worlds sharing the same chance function just *cannot* satisfy the PP perfectly; the credences of such agents, then, cannot be guided by chances as well as the credences of those agents for whom the relevant cardinalities are not prime. This demands explanation; its weirdness strengthens our preference for the conceptually clean Option 1.

We believe, then, that the best way to proceed for someone trying to investigate the details of the PP is—at least, before the intricacies of the infinite contexts are figured out—to abandon the admissibility talk whatsoever.

4 SHOWING THE PP'S CONSISTENCY

Let us now tackle the issue of the PP's consistency. What do we mean by this? Recall that in first-order logic, a theory is consistent iff it has a model; informally, iff there exists something it describes. We have a similar intention here: the PP, a requirement on credence functions, will be taken to be consistent if there is a credence function that satisfies it. This is, however, the weakest imaginable (to us) notion of consistency to be of note. If the PP was not consistent in this sense, there would literally be nothing for us to talk about regarding it. Still, first things first—and so we will present an example of a probability function P such that for any A on which it is defined there is a partition of events which play the role of chance propositions about A: for any $x \in [0, 1]$ there is an event X in that partition such that $P(A \mid X) = x$ (if $P(X) > 0$; see Definition B.1.1).

A norm of rational credence should be satisfiable. But it should also be satisfiable *in principle*, which in the case of the PP we take to suggest that any credence P in some

[20] Some may use integrals—on that front, see footnotes 2 and 3 in [1] for a start.
[21] Such an idea might, of course, also be considered in finite contexts.

body of propositions $\mathcal{F}$ should be compatible with any set of chance functions CH on $\mathcal{F}$ so that the requirements of PP in its local and / or global variants (see equations (1) and (3) above) would be met. Otherwise, we could conceive of a metaphysical coincidence whereupon the possible worlds are governed by chance functions in a way which prohibits some agents from satisfying the PP; effectively, chance could only perform its guiding role for a select privileged group of ideal agents. The PP would then be highly suspect as a candidate for a norm of rationality. Fortunately, this is not the case. It might happen that already the probability space containing the credence P itself and the relevant chance propositions displays the fortuitous coherence—we will then say that the P *satisfies* the PP locally / globally with respect to the set CH (Defs. B.1.7 and B.1.8 below). We should expect, however, that such coherence is only achievable in a larger probability space, in which case we will say that the P is locally/globally-PP-*compatible* with the set CH. Our main result (Theorem 4.0.1) is that indeed, if a simple formal condition is satisfied, for a body of propositions $\mathcal{F}$, any credence P on $\mathcal{F}$ is PP-compatible with any set CH of chance functions on $\mathcal{F}$. The status of the PP as a candidate for a norm of rationality is not in question, then.

As already mentioned, we will first make sure that there *are* probability spaces satisfying the principle, and then we will investigate the limits of combining various credences and chances, that is, CH-PP-compatibility as sketched in Section 2. Along the way we will refute the argument by [12] that the PP implies the PoI. It is an immediate conclusion from that argument that if a probability space satisfies the PP, it has to contain at least one event with probability .5. This seems surprising: if satisfying the PP is a condition of rationality, then it follows that every rational agent has to hold credence .5 in at least one proposition. This seems to be a mistaken requirement, and we will shortly present a counterexample, that is, a probability space satisfying the PP in which no event has probability .5.

As for the first task: given a finite probability space on a field of propositions $\mathcal{F}$, we can construct the local chance propositions "the chance of A is x", for all $x \in [0,1]$, and for all $A \in \mathcal{F}$, so that the PP is satisfied. The procedure requires only a careful analysis of conditional probabilities. This shows the PP's consistency, in a very weak sense. We will show an example.

Consider a probability space $(W = \{a,b,c\}, \mathcal{F} = 2^W, P)$, with $P(\{a\}) = {}^1\!/_7$, $P(\{b\}) = {}^2\!/_7$, and $P(\{c\}) = {}^4\!/_7$. We will assume that the agent has in mind two chance functions; worlds a and b are governed by an indeterministic ch_1, while c is governed by a deterministic ch_2, as shown in Table 1.

	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\top$	$\bot$
$ch_1(\cdot)$	$1/3$	$2/3$	0	1	$1/3$	$2/3$	1	0
$ch_2(\cdot)$	0	0	1	0	1	1	1	0

Table 1: Two chance functions considered by the agent with the credence function P.

Now define the local chance propositions as in Table 2. The Table was created simply by inspecting all conditional probabilities, using the assumption about which worlds are governed by which of the two chance functions. Notice that **for any choice of A and** x, if X is the proposition that the chance of A is x as defined in Table 2, $P(A \mid X) = x$ (if defined). Therefore in this probability space the PP is satisfied ("locally", in the sense of Definition B.1.1 below[22]). Note that no proposition has probability 0.5; therefore,

[22] In Section B.1 of the formal Appendix we use the notation "C_x^A" instead of "X" to refer to the local chance proposition saying that the chance of A is x; this is simply more convenient. The function P under discussion here also "satisfies the PP locally and globally with respect to the set $\{ch_1, ch_2\}$"; see Definitions B.1.2 and B.1.3 below.

if the Principle of Indifference demands the existence of such propositions, which we believe to be an implication of [12],[23] then we have shown by means of this structure that the PP does *not* imply the PoI.[24]

x \ A	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\top$	$\bot$
1	$\emptyset$	$\emptyset$	$\{c\}$	$\{a,b\}$	$\{c\}$	$\{c\}$	W	$\emptyset$
0	$\{c\}$	$\{c\}$	$\{a,b\}$	$\{c\}$	$\emptyset$	$\emptyset$	$\emptyset$	W
$1/3$	$\{a,b\}$	$\emptyset$	$\emptyset$	$\emptyset$	$\{a,b\}$	$\emptyset$	$\emptyset$	$\emptyset$
$2/3$	$\emptyset$	$\{a,b\}$	$\emptyset$	$\emptyset$	$\emptyset$	$\{a,b\}$	$\emptyset$	$\emptyset$
other	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Table 2: The local chance propositions "the chance of A is x", for all possible values of A and x.

Recall that we have assumed that the set $\{a,b\}$ is the global chance proposition C_{ch_1}, while $\{c\}$ is the global chance proposition C_{ch_2}. Note that for any $A \in \mathcal{F}$, for any $i \in \{1,2\}$, $P(A \mid C_{ch_i}) = ch_i(A)$ (if defined). Therefore P satisfies the PP "globally with respect to the set $CH = \{ch_1, ch_2\}$", in the sense of Definition B.1.3 below.

The procedure will work for any probability space with a finite W (with obvious modifications if some singleton sets have probability 0). First, we decide how to partition W; the number of elements of the partition corresponds to the number of chance functions the agent considers. For any singleton element of the partition the corresponding chance function will be deterministic; for non-singleton elements the situation will usually be different. We collect all the conditional probabilities on the elements of our chosen partition (that is, on the global chance propositions) as in Table 1. This is then used to generate all local chance propositions "the chance of A is x", for arbitrary A and x, as in Table 2. That the probability of A given that the chance of A is x is indeed x (when defined) is then guaranteed. Similarly, the probability of A is $ch(A)$ given that the global chance proposition C_{ch} holds.

This, as already mentioned, shows that the PP is consistent in a weak sense: every finite probability space, understood as containing some agent's credences in a body of propositions, can be understood as containing also credences in propositions about the chance of those propositions, so that the two are coordinated exactly as the PP requires. At this point we know that examples of the structures Lewis wrote about indeed exist, and so the whole literature on the PP is not a vacuous exercise. However, the deeper consistency problem is roughly as follows: can any credence function be combined with any set of chance functions in the way required by the PP? Less roughly, is it true that for any probability space $(W, \mathcal{F}, P)$, and for any set of chance functions on $\mathcal{F}$, there exists a credence function P' which agrees with P on $\mathcal{F}$, but which is also defined on the propositions about the chance functions, so that the credences as given by P' are

[23] This implication can be resisted—however, such resistance comes with a significant cost. For an attempt at an exegesis of [12], see Appendix C below.

[24] Some may think admissibility should be addressed before we claim that Hawthorne et al. should agree with us that the PP is satisfied here. The shortest way would be to say that no proposition is admissible with respect to any proposition and chance proposition. However, we could also calculate, for each A and X, the set of E's such that $P(A \mid X) = P(A \mid XE)$, and say that the elements of that set are admissible w.r.t. A and X.

Another *caveat:* despite the title, [12] actually argues that the PP implies the PoI under two additional conditions, the first of which supposedly is provable. One could then take the route of saying that our run-of-the-mill probability space under discussion here is a counterexample to [12]'s Condition 2. Be that as it may, this only reinforces the point that [12] do *not* show the PoI to be implied by the PP. See Appendix C below for a detailed analysis.

coordinated as the PP requires? The answer is **yes**, with some formal caveats. We will first show an intuitive construction fit for countable probability spaces and countable sets of chance functions. It will use global chance propositions; in the Appendix we prove the needed theorems also using local ones.

For illustration let us start with the probability space just discussed: $(W = \{a, b, c\}, \mathcal{F} = 2^W, P)$, with $P(\{a\}) = 1/7$, $P(\{b\}) = 2/7$, and $P(\{c\}) = 4/7$. Suppose the agent would like to consider two chance functions defined on $\mathcal{F}$ as in Table 3, in such a way that the PP is satisfied. We extend the initial probability space to a $(W', \mathcal{F}', P')$ which is *atomless*. This means that whenever $P'(A) = x$, then for any $y < x$ there exists in $\mathcal{F}'$ a B such that $B \subset A$ and $P'(B) = y$. This makes for some convenient freedom in defining the propositions relevant to our goals. We shall choose W' to be the the $[0, 1]$ real segment, $\mathcal{F}'$ to be set of its Borel subsets, and the P' to be the Lebesgue measure. We can extend our $(W, \mathcal{F}, P)$ to the atomless space $(W', \mathcal{F}', P')$ by means of a boolean homomorphism h which "keeps all the old probabilities", that is, such that for any $A \in \mathcal{F}$, $P'(h(A)) = P(A)$.[25] The details of the extension feature in Figure 1—for example, $h(\{b\}) = [1/7, 3/7)$.[26] Then, in the extended space, we define the two global chance propositions, again e.g. as displayed in Figure 1. We end up with combining the credences as given by P with the two specified chance functions by means of the credence P' which is such that for any proposition A on which P is defined, $P'(h(A) \mid C_{ch_3}) = ch_3(A)$ and $P'(h(A) \mid C_{ch_4}) = ch_4(A)$. That is, P is $\{ch_3, ch_4\}$-PP-compatible: the credence function can be combined with the two chance functions as the PP mandates.

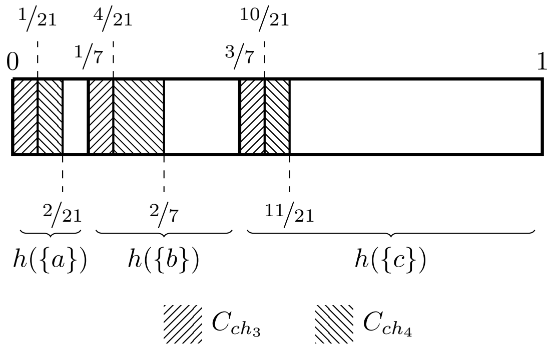

Figure 1: The two global chance propositions in the extended probability space which shows that P is $\{ch_3, ch_4\}$-PP-compatible.

	$\{a\}$	$\{b\}$	$\{c\}$
$ch_3(\cdot)$	1/3	1/3	1/3
$ch_4(\cdot)$	1/4	1/2	1/4

Table 3: Two chance functions we are "fitting together" with a credence function defined on $\mathcal{F} = \mathcal{P}(\{a, b, c\})$.

This generalizes almost fully, as we show in the following Theorem (where $a \lessdot b$ is true iff both a and b are zero or a is strictly smaller than b):

[25] For details on why this is generally possible, see e.g. Theorem 2 from [20] and the discussion leading up to it.

[26] No, you cannot see in the figure which ends of the relevant intervals are open and which are closed.

4.0.1 Theorem. Let $(W, \mathcal{F}, P)$ be an atomic probability space, and let $CH = \{ch_1, ch_2, \ldots\}$ be a countable set of probability functions over $\mathcal{F}$. The following statements are equivalent.

(1) P is CH-PP-compatible in the sense of Definition B.1.8.

(2) For each $ch \in CH$ there is a constant $c > 0$ such that

$$c \cdot ch(A) \lessdot P(A) \quad \text{for all } A \in \mathcal{F}\,. \tag{4}$$

For the proof, see Appendix, Section B.2. (The condition (4) might be seen to be reminiscent of the one used in Theorem 2.1 of [4].)

We have established that in exactly those cases where each chance function in CH can be 'lifted above' the credence P by means of multiplication by a constant, P is CH-PP-compatible; that is, there is a credence function P' which relates credences as given by P with chances from CH as mandated by the PP. Note that, crucially, we should expect this P' to be defined also on propositions outside of the domains of the chance functions from CH.

We started this section with an example of a probability space satisfying the PP, illustrating the general strategy of creating such spaces by "reading off" the chance functions from the conditional probabilities. Take a look at Table 1 again. Notice that $\{a, b\}$ is the global chance proposition C_{ch_1}, and $ch_1(C_{ch_1}) = 1$; likewise, $\{c\}$ is the global chance proposition C_{ch_2}, and $ch_2(C_{ch_2}) = 1$. Both chance functions assign chance 1 to themselves being the actual chance function; in other words; they are *immodest*. This is not a coincidence: as shown by [25], a credence function P can satisfy the PP with respect to a set of chance functions CH[27] only if all chance functions from CH are immodest. However, as we show in the Appendix (Theorems B.2.1 and B.3.7), a P will be CH-PP-*compatible* as long as the 'lift-up' condition is satisfied, even if some, or all, members of CH are modest. The key technical feature here is that the compatibility of P and CH is witnessed by an *extension* of the original space, in which the P is indeed combined with the members of CH in the PP-appropriate way, but which contains also events on which the members of CH are not defined. The apparent contradiction mentioned in the Introduction is then resolved by the following subtlety: P can *satisfy* the PP only if CH contains exclusively immodest chance functions, but, supposing the "lift-up" condition is satisfied, P is CH-PP-*compatible* also if some chances in CH are modest.

5 CONCLUSIONS

We have proposed here a rigorous solution to a problem which we believe to be foundational for that corner of formal epistemology which is interested in chance-credence norms: that is, after charting several ways in which the Principal Principle can be formally stated (and discussing some pitfalls concerning the notion of admissibility in Section 3), we showed that it can be satisfied by an arguably large class of agents. In particular, we have specified the relationship between the agent's credences in a body of propositions and chance functions defined on the same domain which is necessary and sufficient for these to be combined in the way the PP mandates, so that the chances can be seen to guide the agent's credences (Theorems B.2.1 and B.3.7). We have also shown that the formulation of the PP which uses local chance propositions is, in atomic probability spaces, equivalent to the one which employs global chance propositions (Corollary B.3.8).

[27] [25]'s definition of this is closest to our "satisfying the PP globally w.r.t. CH", that is, Def. B.1.3 below.

We admit we are puzzled that the issue of the PP's consistency has not garnered the attention of researchers: after all, if the Principle is to be a norm of rationality, shouldn't we make sure that satisfying it is possible in general (even for ideal agents)? [5], noticing the problem, suggests that we could employ higher-order probability spaces introduced by [7] to extend arbitrary probability spaces so that they include propositions about chances in the way suggested by the PP (Gaifman himself suggests the connection between his approach and Lewis' notion). However, Gaifman's framework is remarkably intricate, and so doing this in detail is a matter for a separate paper.

FUNDING

The research was supported by the grant 2019/34/E/HS1/00044, 'Epistemic inaccuracy and foundational issues in formal epistemology', of the National Science Centre (Poland).

LW's and ZG's work on the paper was also supported by the grant 2022/47/B/HS1/01581, 'Epistemic inaccuracy – what's next?', of the National Science Centre (Poland).

The OA statement required by the National Science Centre:

REFERENCES

[1] Bana, G. (2016). On the Formal Consistency of the Principal Principle. *Philosophy of Science*, 83(5):988–1001.

[2] Burris, S. and Sankappanavar, H. P. (1981). *A Course in Universal Algebra*. Springer.

[3] Carroll, J. W. (2020). Laws of Nature. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2020 edition.

[4] Diaconis, P. and Zabell, S. L. (1982). Updating Subjective Probability. *Journal of the American Statistical Association*, 77(380):822–830.

[5] Dziurosz-Serafinowicz, P. (2016). *The Double Life of Probability. A Philosophical Study of Chance and Credence*. University of Groningen.

[6] Eisenberg, B. and Ghosh, B. K. (1987). Independent Events in a Discrete Uniform Probability Space. *The American Statistician*, 41(1):52–56.

[7] Gaifman, H. (1986). A theory of higher order probabilities. In Halpern, J. Y., editor, *Theoretical Aspects of Reasoning About Knowledge. Proceedings of the 1986 Conference.*, pages 275–292.

[8] Gyenis, B. and Wroński, L. (2017). Is it the Principal Principle that implies the Principle of Indifference? In Hofer-Szabó, G. and Wroński, L., editors, *Making it Formally Explicit: Probability, Causality, and Determinism*, pages 35–42. Springer.

[9] Gyenis, Z. and Rédei, M. (2016). Measure theoretic analysis of consistency of the Principal Principle. *Philosophy of Science*, 83(5):972–987.

[10] Hall, N. (1994). Correcting the Guide to Objective Chance. *Mind*, 103(412):505–517.

[11] Hall, N. (2004). Two Mistakes About Credence and Chance. *Australasian Journal of Philosophy*, 82(1):93–111.

[12] Hawthorne, J., Landes, J., Wallmann, C., and Williamson, J. (2017). The Principal Principle Implies the Principle of Indifference. *The British Journal for the Philosophy of Science*, 68(1):123–131.

[13] Hoefer, C. (2007). The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance. *Mind*, 116(463):549–596.

[14] Kyburg, H. E. (1961). *Probability and the Logic of Rational Belief*. Middletown, CT, USA: Wesleyan University Press.

[15] Kyburg, H. E. (1974). *The Logical Foundations of Statistical Inference*. Dordrecht and Boston: Reidel.

[16] Levi, I. (1977). Direct Inference. *Journal of Philosophy*, 74(1):5–29.

[17] Lewis, D. (1986). A Subjectivist's Guide to Objective Chance (with postscripts). In *Philosophical Papers, Volume II*, pages 83–132.

[18] Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, 103(412):473–90.

[19] Loewer, B. (2004). David Lewis' Humean Theory of Objective Chance. *Philosophy of Science*, 71(5):1115–1125.

[20] Marczyk, M. and Wroński, L. (2014). Completion of the Causal Completability Problem. *The British Journal for the Philosophy of Science*, 66(2):307–326.

[21] Masterton, G. (2010). *Objective Chance. A Study in the Lewisian Tradition*. Filosofiska institutionen Uppsala.

[22] Meacham, C. J. (2010). Two Mistakes Regarding the Principal Principle. *British Journal for the Philosophy of Science*, 61:407–431.

[23] Nissan-Rozen, I. (2018). On the Inadmissibility of Some Historical Information. *Philosophy and Phenomenological Research*, 97(2):479–493.

[24] Peirce, C. S. (1883). A theory of probable inference. In *Studies in Logic by Members of the Johns Hopkins University*, pages 126–181. Little, Brown and Co, New York, NY, US.

[25] Pettigrew, R. (2015). What Chance-Credence Norms Should Not Be. *Noûs*, 49(1):177–196.

[26] Pettigrew, R. (2016). Epistemic Utility Arguments for Probabilism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2016 edition.

[27] Pollock, J. L. (1990). *Nomic Probability and the Foundations of Induction*. Oxford University Press, Oxford, New York.

[28] Reichenbach, H. (1949). *The Theory of Probability: An Inquiry Into the Logical and Mathematical Foundations of the Calculus of Probability*. Berkeley, CA, USA: University of California Press.

[29] Salmon, W. C., Jeffrey, R. C., and Greeno, J. G. (1971). *Statistical Explanation and Statistical Relevance*. University of Pittsburgh Press.

[30] Schwarz, W. (2014). Proving the Principal Principle. In Wilson, A., editor, *Chance and Temporal Asymmetry*, pages 81–99. Oxford University Press.

[31] Thau, M. (1994). Undermining and Admissibility. *Mind*, 103(412):491–503.

[32] Thorn, P. D. (2012). Two Problems of Direct Inference. *Erkenntnis (1975-)*, 76(3):299–318.

[33] Venn, J. (1876). *The Logic of Chance*. Mineola, N.Y.: Dover Publications.

[34] Vranas, P. B. (2002). Who's Afraid of Undermining? Why the Principal Principle Might Not Contradict Humean Supervenience. *Erkenntnis*, 57(2):151–175.

[35] Wallmann, C. and Hawthorne, J. (2020). Admissibility Troubles for Bayesian Direct Inference Principles. *Erkenntnis*, 85:957–993.

[36] Wroński, L. (2018). *In Good Form. Arguing for Epistemic Norms of Credence*. Jagiellonian University Press.

[37] Wroński, L., Gyenis, Z., and Cocchiaro, M. Z. (2022). What Chance-Credence Norms Should Be. *Synthese*, 200(508).

A APPENDIX

A.1 PROOF THAT THE LARGER THE UNIFORM SPACES THE RARER THE INDEPENDENT PAIRS

We consider here finite, uniform probability spaces and calculate the probability that two uniformly randomly picked events are independent. We will see that as the size of the space increases the probability that two random events are independent tends to zero. Thus, in large finite uniform spaces most pairs of events are not independent.

Consider the probability space with a sample space of size n having the uniform probability measure. Two events A, B are independent if and only if

$$\frac{|A|}{n} \cdot \frac{|B|}{n} = \frac{|A \cap B|}{n}. \tag{5}$$

Call a triple of numbers (a, b, c) *good* if $0 \leq a, b, c \leq n$ and $a/n \cdot b/n = c/n$. Then A and B are independent iff $(|A|, |B|, |A \cap B|)$ is good. For a good triplet (a, b, c) let us write

$$D_{a,b,c} = |\{(A, B) \text{ independent} : \ |A| = a, |B| = b, |A \cap B| = c\}|. \tag{6}$$

The number of all pairs (A, B) is 2^{2n}, therefore the probability that a randomly picked pair is independent is

$$\mathrm{IND}_n = \frac{1}{2^{2n}} \sum_{(a,b,c) \text{ good}} D_{a,b,c}. \tag{7}$$

Combinatorics shows that

$$D_{a,b,c} = \binom{n}{a}\binom{a}{c}\binom{n-a}{b-c}. \tag{8}$$

We pick first the a-element set A; then we pick c elements from A that will be in $A \cap B$; then pick the rest of the elements of B from the complement of A. We have the following estimates for IND_n:

$$\frac{1}{2^{2n}} \sum_{\substack{(a,b,c) \\ \text{is good}}} D_{a,b,c} \overset{(1)}{\leq} \frac{1}{2^{2n}} \sum_{a=0}^{n} \sum_{c=0}^{a} D_{a, \frac{cn}{a}, c} \leq \frac{1}{2^{2n}} \sum_{a=0}^{n} \sum_{c=0}^{a} \binom{n}{a}\binom{a}{c}\binom{n-a}{\frac{cn}{a} - c} \tag{9}$$

$$\overset{(2)}{\leq} \frac{1}{2^{2n}} \sum_{a=0}^{n} \binom{n}{a} \sum_{c=0}^{a} \binom{a}{c}\binom{n-a}{\frac{n-a}{2}} = \frac{1}{2^{2n}} \sum_{a=0}^{n} \binom{n}{a}\binom{n-a}{\frac{n-a}{2}} \sum_{c=0}^{a} \binom{a}{c} \tag{10}$$

$$\overset{(3)}{=} \frac{1}{2^{2n}} \sum_{a=0}^{n} \binom{n}{a}\binom{n-a}{\frac{n-a}{2}} 2^a \overset{(4)}{\leq} \frac{1}{2^{2n}} \sum_{a=0}^{n} \binom{n}{a} 2^a \frac{2^{n-a}}{\sqrt{\frac{\pi}{2}(n-a)}} \tag{11}$$

$$= \frac{1}{2^n} \sum_{a=0}^{n} \binom{n}{a} \frac{1}{\sqrt{\frac{\pi}{2}(n-a)}} \overset{(5)}{=} \frac{1}{2^n} \sum_{a=0}^{n} \binom{n}{n-a} \frac{1}{\sqrt{\frac{\pi}{2}(n-a)}} \tag{12}$$

$$= \frac{1}{2^n \sqrt{\pi/2}} \sum_{k=0}^{n} \binom{n}{k} \frac{1}{\sqrt{k}}, \tag{13}$$

where we have used the facts that:

(1) not all triplets are good;

(2) $\binom{x}{x/2}$ is always maximal among the $\binom{x}{k}$'s;

(3) $\sum_{c=0}^{a} \binom{a}{c} = 2^a$;

(4) $\binom{k}{k/2} \leq \frac{2^k}{\sqrt{\frac{\pi}{2}k}}$ (under the convention that $\frac{1}{\sqrt{0}} = 1$);

(5) $\binom{n}{a} = \binom{n}{n-a}$.

A.1.1 Theorem. The probability IND_n that a randomly picked pair of events from an n-element uniform space is independent converges to zero as n tends to infinity:

$$\lim_{n\to\infty} \mathrm{IND}_n \;=\; 0\,. \tag{14}$$

Proof. By the previous estimates it is enough to prove that

$$\frac{1}{2^n}\sum_{k=1}^{n}\binom{n}{k}\frac{1}{\sqrt{k}} \;\longrightarrow\; 0 \qquad \text{as } n\to\infty. \tag{15}$$

Hölder's inequality with $p = 2$ and $q = 2$ yields

$$\frac{1}{2^n}\sum_{k=1}^{n}\binom{n}{k}\frac{1}{\sqrt{k}} \;\overset{(1)}{\leq}\; \frac{1}{2^n}\sqrt{\sum_{k=1}^{n}\binom{n}{k}^2}\sqrt{\sum_{k=1}^{n}\frac{1}{k}} \;\overset{(2)}{\leq}\; \frac{1}{2^n}\sqrt{\binom{2n}{n}}\sqrt{\log(n)} \tag{16}$$

$$\overset{(3)}{\leq}\; \frac{1}{2^n}\sqrt{\frac{2^{2n}}{\sqrt{\pi n}}}\sqrt{\log(n)} \;=\; \sqrt{\frac{\log(n)}{\sqrt{\pi n}}} \;\longrightarrow\; 0 \qquad \text{as } n\to\infty \tag{17}$$

where we used

(1) Hölder's inequality with $p = 2$ and $q = 2$,

(2) $\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}$, and $\sum_{k=1}^{n}\frac{1}{k} \leq \int_1^n \frac{1}{x}dx = \log(n)$,

(3) $\binom{2n}{n} \leq \frac{2^{2n}}{\sqrt{\pi n}}$, and finally that $\lim_{n\to\infty}\frac{\log(n)}{\sqrt{n}} = 0$.

■

B PP's consistency: the formal details

B.1 The main notions employed

B.1.1 Definition (satisfying the PP locally). Let $(W, \mathcal{F}, P)$ be a probability space. We will say that P *satisfies the PP locally* iff for any $A \in \mathcal{F}$ there exists a family of propositions $\{C_x^A\}_{x\in[0,1]}$ from $\mathcal{F}$ such that:

- the nonempty elements of $\{C_x^A\}_{x\in[0,1]}$ form a partition of W;
- for any $x \in [0,1]$, $P(A \mid C_x^A) = x$ (if defined). □

The justification for the first condition from B.1.1 is that every world assigns some single chance to A; therefore, every world belongs to some local chance proposition about A, and no world belongs to two different local chance propositions about A. For some x there might be no worlds at which the chance of A is x—hence the qualification.

The final "(if defined)" is needed so that we allow an agent to satify the PP locally even if for some A and x their credence in that the chance of A is x equals 0. However, variants of the notion without this relaxing clause might be fruitfully investigated (see B.4 below); note, though, that requiring (with the A fixed) $P(C_x^A)$ to be nonzero for all $x \in [0,1]$ would imply that $C_x^A \cap C_y^A \neq \emptyset$ for some $x \neq y$. This is because there cannot exist more than countably many pairwise disjoint, non-zero-measure sets in a probabilty space. If the events C_x^A ought to represent the proposition that the "chance of A is x", and events are sets of possible worlds, then requiring $P(C_x^A)$ to be nonzero for all $x \in [0,1]$ would yield the counterintuitive situation that the chance of A is x and y for $x \neq y$ in some world. To avoid such cases, in the above definition we are speaking of a "family of propositions" only.[28]

Below we omit the justifications of similar conditions.

To tackle the consistency problem in general, we need definitions capturing the idea of coordinating a credence function P with a set of chance functions CH in the spirit of PP. There will be four definitions in total. Choosing whether to use local or global chance propositions makes for the initial division into two types of the relevant formal concepts. Finally, B.1.2 and B.1.3 describe the fortunate situation in which the original probability space with the measure P already contains the chance propositions in question; while B.1.7 and B.1.8 deal with the general situation where extending the original space might be needed.

B.1.2 Definition (satisfying the PP locally w.r.t. CH). Let $(W, \mathcal{F}, P)$ be a probability space. Let CH be a set of probability functions with the same domain as P. We will say that P *satisfies the PP locally with respect to CH* if and only if for every $A \in \mathcal{F}$ there are propositions C_{ch}^A for $ch \in CH$ such that

- (assuming the A is fixed) the nonempty elements of $\{C_{ch}^A\}_{ch \in CH}$ form a partition of W;
- for any $ch \in CH$, $P(A \mid C_{ch}^A) = ch(A)$ (if defined). □

B.1.3 Definition (satisfying the PP globally w.r.t. CH). Let $(W, \mathcal{F}, P)$ be a probability space. Let CH be a set of probability functions with the same domain as P. We will say that P *satisfies the PP globally with respect to CH* if and only if for every $ch \in CH$ there exists a proposition $C_{ch} \in \mathcal{F}$ such that:

- the family of all nonempty C_{ch}'s forms a partition of W;
- for any $A \in \mathcal{F}$, $P(A \mid C_{ch}) = ch(A)$ (if defined). □

B.1.4 Fact. For every P with a finite domain there exists a nonempty set CH such that P satisfies the PP globally with respect to CH.

To prove Fact B.1.4, inspect the strategy used in creating the example displayed in Tables 1 and 2 above.

B.1.5 Fact. For any P and CH, if P satisfies the PP globally with respect to CH, then P satisfies the PP locally with respect to CH.

Proof. Suppose P satisfies the PP globally w.r.t. CH. For every $A \in \mathcal{F}$ and $ch \in CH$ take $C_{ch}^A := C_{ch}$. Then P satisfies the PP locally with respect to CH, as witnessed by the events C_{ch}^A. ■

B.1.6 Conjecture. For any P and CH, P satisfies the PP globally with respect to CH iff P satisfies the PP locally with respect to CH.

B.1.7 Definition (local PP compatibility). Let $(W, \mathcal{F}, P)$ be a probability space, and let CH be a set of probability functions over $\mathcal{F}$. We say that P is *CH-PP-locally-compatible* if there exists an extension $(W', \mathcal{F}', P')$ of $(W, \mathcal{F}, P)$ and a σ-algebra embedding $h : \mathcal{F} \to \mathcal{F}'$ such that

[28] In a still different variant, if we kept just the second bullet point, but deleted the first—perhaps because we would want to be as minimal as possible with additional assumptions about chance propositions—it would be possible for an agent to vacuously satisfy the PP locally by setting all credences in local chance propositions to 0.

- P and P' agree on events of $\mathcal{F}$, that is, for all $A \in \mathcal{F}$ we have

$$P'(h(A)) = P(A)\,; \tag{18}$$

- For every $A \in \mathcal{F}$ and $ch \in CH$ there is an event C^A_{ch} in $\mathcal{F}'$ such that
  - (keeping A fixed) the nonempty elements of $\{C^A_{ch}\}_{ch \in CH}$ are pairwise disjoint;
  - $P'(h(A) \mid C^A_{ch}) = ch(A)$. □

B.1.8 Definition (global PP compatibility). Let $(W, \mathcal{F}, P)$ be a probability space, and let CH be a set of probability functions over $\mathcal{F}$. We say that P is *CH-PP-globally-compatible* if there exists an extension $(W', \mathcal{F}', P')$ of $(W, \mathcal{F}, P)$ and a σ-algebra embedding $h : \mathcal{F} \to \mathcal{F}'$ such that

- P and P' agree on events from $\mathcal{F}$, that is, for all $A \in \mathcal{F}$ we have

$$P'(h(A)) = P(A)\,; \tag{19}$$

- For every $ch \in CH$ there is an event C_{ch} in $\mathcal{F}'$ such that
  - $P'(h(A) \mid C_{ch}) = ch(A)$ for all $A \in \mathcal{F}$;
  - if $ch, ch' \in CH$, $ch \neq ch'$, then $C_{ch} \cap C_{ch'} = \emptyset$.

□

A comment on definitions B.1.2/7 and B.1.3/8. Despite the "global" and "local" labels, both variants of these definitions use chance functions defined on the whole $\mathcal{F}$. The difference is in how these chance functions are used when considering conditional credences in some A. In the local variants we are dealing, for any value x, with credences in A given that the chance of A is x, that is, with credences in A given *local* chance propositions (see Section 2 above). In contrasts, the global variants concern credences in A conditional on that the chance function is given by some particular ch, that is, credences in A given *global* chance propositions.

B.2 CONSISTENCY: GLOBAL COMPATIBILITY

For convenience we make use of the following notation. For non-negative numbers a and b we write

$$a \lessdot b \;\Leftrightarrow\; \begin{cases} a < b & \text{if } 0 < b \\ a = 0 & \text{if } b = 0. \end{cases} \tag{20}$$

Thus, $a \lessdot b$ expresses that a is strictly smaller than b, except when b equals zero, in which case a is zero too.

B.2.1 Theorem. (4.0.1 in the main text.) Let $(W, \mathcal{F}, P)$ be an atomic probability space, and let $CH = \{ch_1, ch_2, \ldots\}$ be a countable set of probability functions over $\mathcal{F}$. The following statements are equivalent.

(1) P is CH-PP-globally-compatible.

(2) For each $ch \in CH$ there is a constant $c > 0$ such that

$$c \cdot ch(A) \lessdot P(A) \quad \text{for all } A \in \mathcal{F}\,. \tag{21}$$

Proof. (1)⇒(2). Suppose P is CH-PP-globally-compatible as witnessed by the extension $(W', \mathcal{F}', P')$ and the embedding $h : \mathcal{F} \to \mathcal{F}'$. Pick a $ch \in CH$. Applying (B.1.8) we get, for each $A \in \mathcal{F}$,

$$ch(A) = P'(h(A) \mid C_{ch}) = \frac{P'(h(A) \cap C_{ch})}{P'(C_{ch})}\,, \tag{22}$$

and thus

$$P'(C_{ch}) \cdot ch(A) \leq P'(h(A) \cap C_{ch}) \leq P'(h(A)) = P(A)\,. \tag{23}$$

Note that it follows that if $P(A) = 0$, then $ch(A)$ should equal zero. Thus, taking $c = {}^{P'(C_{ch})}/_{2}$ will do.

(2)⇒(1). Let $(W, \mathcal{F}, P)$ be an atomic probability space, and let $CH = \{ch_1, ch_2, \ldots\}$ be a countable[29] set of probability functions over $\mathcal{F}$. Except for the measure zero atoms, without loss of generality, we can assume that $W = \{a_n : n \in |W|\}$, where $0 < P(\{a_n\})$ for all $n \in |W|$. Here $|W|$ is either $\mathbb{N}$, or some finite natural number $N \in \mathbb{N}$ depending on whether W is countably infinite or finite.[30] For each $ch_n \in CH$ let c_n be the constant in (21). Let λ be the Lebesgue measure over the unit interval $[0, 1]$. We will embed everything into this space.

For simplicity we write $P(a_n)$ in place of $P(\{a_n\})$, etc.

Step 1. Take any measurable partition A_n ($n \in |W|$) of $[0, 1]$ such that $P(a_n) = \lambda(A_n)$ holds for all $n \in |W|$. For example, each A_n can be an interval of length $P(a_n)$. The mapping $h(a_n) = A_n$ extends to an algebra embedding such that $P(A) = \lambda(h(A))$ for all $A \in \mathcal{F}$.

Step 2. For each $n \in |W|$ pick an arbitrary partition $C_{n,i}$ ($i \in |CH|$) of A_n[31] such that $\lambda(C_{n,i}) = \frac{1}{2^i}\lambda(A_n)$ for each i and n. By non-atomicity of the Lebesgue measure, this is possible. As $c_i \cdot ch_i(a_n) < P(a_n)$, and $P(a_n) = \lambda(A_n)$, we get that

$$\frac{1}{2^i} \cdot c_i \cdot ch_i(a_n) < \frac{1}{2^i}\lambda(A_n) = \lambda(C_{n,i}) \tag{24}$$

for $i \in |CH|$, $n \in |W|$.

Step 3. For each $i \in |CH|$ and $n \in |W|$ pick $X_{n,i} \subseteq C_{n,i}$ such that

$$\lambda(X_{n,i}) = \frac{c_i}{2^i} \cdot ch_i(a_n) \tag{25}$$

Let $X_i = \bigcup_n X_{n,i}$.

Calculations. First, let us calculate $\lambda(X_i)$:

$$\lambda(X_i) = \sum_n \lambda(X_{n,i}) = \sum_n \frac{c_i}{2^i} \cdot ch_i(a_n) = \frac{c_i}{2^i} \cdot \sum_n ch_i(a_n) = \frac{c_i}{2^i}\,. \tag{26}$$

Then,

$$\lambda(A_n \mid X_i) = \frac{\lambda(A_n \cap X_i)}{\lambda(X_i)} = \frac{\lambda(X_{n,i})}{\lambda(X_i)} = \frac{\frac{c_i}{2^i} \cdot ch_i(a_n)}{\frac{c_i}{2^i}} = ch_i(a_n)\,. \tag{27}$$

[29] That is, countably infinite or finite.

[30] To avoid cumbersome notation we rely on the convention that $0 = \emptyset$ and $n = \{0, \ldots, n-1\}$. Thus, for example, if $|W| = 3$, then $n \in |W|$ means $n = 0$ or $n = 1$ or $n = 2$.

[31] To be sure: $C_{n,i}$ is a partition of A_n whose elements are indexed by i.

Writing C_{ch_i} for the event X_i we get

$$\lambda(h(a_n) \mid C_{ch_i}) = ch_i(a_n) \quad \text{for all } n \in |W| \,. \tag{28}$$

As the atoms a_n generate every element of $\mathcal{F}$ by taking disjoint unions, we also get

$$\lambda(h(A) \mid C_{ch_i}) = ch_i(A) \quad \text{for all } A \in \mathcal{F} \,, \tag{29}$$

completing the proof.

We finally note that $C_{ch_i} \cap C_{ch_j} = \emptyset$ for $i \neq j$. ■

B.3 CONSISTENCY: LOCAL COMPATIBILITY

The results in this Section gradually lead to the main theorem (B.3.7).

We first show that (30) in the following remark is a necessary condition for local compatibility.

B.3.1 Remark. Let $(W, \mathcal{F}, P)$ be a probability space and p an other probability function over $\mathcal{F}$. Suppose that P is $\{ch\}$-PP-locally-compatible. Then there is a constant $a \in (0, 1)$ such that

$$a \cdot ch(A) \lessdot P(A) \quad \text{for all } A \in \mathcal{F} \,. \tag{30}$$

Proof. Let $\{ch\}$-PP-local-compatibility of P be witnessed by the extension $(W', \mathcal{F}', P')$ and the embedding $h : \mathcal{F} \to \mathcal{F}'$. Write X for C^A_{ch}. Then

$$ch(A) = P'(h(A) \mid X) = \frac{P'(h(A) \cap X)}{P'(X)} \,, \tag{31}$$

and thus

$$P'(X) \cdot ch(A) \leq P'(h(A) \cap X) \leq P'(h(A)) = P(A) \,. \tag{32}$$

Taking $a = 1/2 \cdot P'(X)$ will do. ■

The next theorem and the remark following it and their proofs are essentially Theorem 4.1 of Bana [1], slightly generalized and adopted to our context.

B.3.2 Theorem. Let $(W, \mathcal{F}, P)$ be a probability space, and let $CH = \{ch_1, ch_2, \ldots\}$ be a countable set of probability functions over $\mathcal{F}$. Suppose there is a non-negative sequence a_n such that

$$\sum_n a_n ch_n(A) = P(A) \quad \text{for all } A \in \mathcal{F} \,, \tag{33}$$

that is, P is a mixture of the probabilities in CH. Then P is CH-PP-locally-compatible.

Proof of Theorem B.3.2. Define

$$W^{ch} \stackrel{\text{def}}{=} \{p : \mathcal{F} \to [0, 1] : \ p \text{ is a probability}\}, \tag{34}$$

$$D^A_r \stackrel{\text{def}}{=} \{p \in W^{ch} : \ p(A) = r\}, \quad \text{for } A \in \mathcal{F}, r \in [0, 1] \,, \tag{35}$$

and let $\mathcal{F}^{ch}$ be the smallest σ-algebra over W^{ch} that contains the sets D^A_r. Define $\mu : \mathcal{F}^{ch} \to [0, 1]$ as

$$\mu \stackrel{\text{def}}{=} \sum_n a_n \delta_{ch_n} \,, \tag{36}$$

where δ_p is the Dirac-measure: for $X \subseteq W^{ch}$

$$\delta_p(X) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } p \in X \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

By (33), μ is a probability measure on $\mathcal{F}^{ch}$.

B.3.3 Claim. $P(A) = \sum_r r \cdot \mu(D_r^A)$ for $A \in \mathcal{F}$. (In particular, the sum exists).

Proof of Claim B.3.3. By definition

$$\sum_r r \cdot \mu(D_r^A) = \sum_r r \cdot \left(\sum_n a_n \delta_{p_n}(D_r^A) \right). \tag{38}$$

Here $\delta_{ch_i}(D_r^A)$ is 1 if $ch_i(A) = r$, and 0 otherwise. Therefore, δ_{ch_i} adds to the sum only when $r = ch_i(A)$, hence

$$\sum_r r \cdot \left(\sum_n a_n \delta_{ch_n}(D_r^A) \right) = \sum_n a_n ch_n(A). \tag{39}$$

The latter sum is $P(A)$ by (33). ■

Consider now the product space

$$(W', \mathcal{F}') \stackrel{\text{def}}{=} (W \times W^{ch}, \mathcal{F} \otimes \mathcal{F}^{ch}) \tag{40}$$

and define P' as the probability measure that extends

$$P'(A \times E) \stackrel{\text{def}}{=} \sum_r r \cdot \mu(D_r^A \cap E) \qquad \text{for } A \in \mathcal{F},\ E \in \mathcal{F}^{ch} \tag{41}$$

B.3.4 Claim. The mapping $h(A) = A \times W^{ch}$ is a σ-algebra embedding and $P'(h(A)) = P(A)$ for every $A \in \mathcal{F}$.

Proof of Claim B.3.4. That h is an embedding that preserves the Boolean-algebra operations is straightforward. To check $P'(h(A)) = P(A)$, by definition we have

$$P'(A \times W^{ch}) = \sum_r r \cdot \mu(D_r^A \cap W^{ch}) \tag{42}$$

$$= \sum_r r \cdot \mu(D_r^A) \stackrel{\text{Claim B.3.3}}{=} P(A) \tag{43}$$

■

Proof of Theorem B.3.2 continued. Suppose now that $\mu(D_x^A) \neq 0$ for some $x \in [0,1]$. Then

$$P'(A \times D_x^A) \stackrel{\text{def}}{=} \sum_r r \cdot \mu(D_r^A \cap D_x^A) = x \cdot \mu(D_x^A), \tag{44}$$

$$P'(W \times D_x^A) \stackrel{\text{def}}{=} \sum_r r \cdot \mu(D_r^W \cap D_x^A) = 1 \cdot \mu(D_x^A), \tag{45}$$

because in (44), $D_r^A \cap D_x^A$ is non-empty only when $r = x$, and in (45), D_r^W is non-empty only when $r = 1$. It follows that

$$P'(A \times W^{ch} \mid W \times D_x^A) = \frac{P'(A \times D_x^A)}{P'(W \times D_x^A)} = \frac{x \cdot \mu(D_x^A)}{1 \cdot \mu(D_x^A)} = x. \tag{46}$$

Observe that

$$\mu(D_x^A) \stackrel{\text{def}}{=} \sum_n a_n \delta_{ch_n}(D_x^A) > 0 \tag{47}$$

if and only if there is i such that $ch_i(A) = x$. This is exactly the case when x is chosen to be one of the $ch_i(A)$. Therefore, for every $A \in \mathcal{F}$ we obtained

$$P'(A \times W^{ch} \mid W \times D_{ch_i(A)}^A) = ch_i(A) \tag{48}$$

Letting

$$C_{ch_i}^A \stackrel{\text{def}}{=} W \times D_{ch_i(A)}^A, \tag{49}$$

we conclude the proof with noting that as Definition B.1.7 requires, first, for a fixed A the nonempty elements of $\{C_{ch}^A\}_{ch \in CH}$ are pairwise disjoint,[32] and also, second, that

$$P'(h(A) \mid C_{ch_i}^A) = ch_i(A)\,. \tag{50}$$

■

B.3.5 Remark. Theorem B.3.2 remains true in the stronger form:

$$P'(h(A) \mid C_{ch}^A \cap \bigcap_{i<n} C_{ch}^{A_i}) = ch(A)\,, \tag{51}$$

for any finitely many $A, A_i \in \mathcal{F}$ $(i < n)$ and $ch \in CH$.

Proof. The proof of Theorem B.3.2 can be slightly modified to get the result. The proposition $C_{ch}^{A_i}$ was encoded by the event $W \times D_{ch(A_i)}^{A_i}$. Then in place of (44) and (45) can have

$$P'(A \times (D_x^A \cap \cap_{i<n} D_{x_i}^{A_i})) \stackrel{\text{def}}{=} \sum_r r \cdot \mu(D_r^A \cap D_x^A \cap \cap_{i<n} D_{x_i}^{A_i}) \tag{52}$$

$$= x \cdot \mu(D_x^A \cap \cap_{i<n} D_{x_i}^{A_i}), \tag{53}$$

$$P'(W \times (D_x^A \cap \cap_{i<n} D_{x_i}^{A_i})) \stackrel{\text{def}}{=} \sum_r r \cdot \mu(D_r^W \cap D_x^A \cap \cap_{i<n} D_{x_i}^{A_i}) \tag{54}$$

$$= 1 \cdot \mu(D_x^A \cap \cap_{i<n} D_{x_i}^{A_i}), \tag{55}$$

because in (53), $D_r^A \cap D_x^A \cap \cap_{i<n} D_{x_i}^{A_i}$ is non-empty only when $r = x$, and for some probability mapping p we have $p(A) = x$ and $p(A_i) = x_i$ for $i < n$; and in (55), D_r^W is non-empty only when $r = 1$. It follows that

$$P'(h(A) \mid C_{ch}^A \cap \bigcap_{i<n} C_{ch}^{A_i}) = \frac{P'(A \times (D_{ch(A)}^A \cap \cap_{i<n} D_{ch(A_i)}^{A_i}))}{P'(W \times (D_{ch(A)}^A \cap \cap_{i<n} D_{ch(A_i)}^{A_i}))} \tag{56}$$

$$= \frac{ch(A) \cdot \mu(D_{ch(A)}^A \cap \cap_{i<n} D_{ch(A_i)}^{A_i})}{1 \cdot \mu(D_{ch(A)}^A \cap \cap_{i<n} D_{ch(A_i)}^{A_i})} \tag{57}$$

$$= ch(A)\,. \tag{58}$$

■

[32] Indeed, more is achieved: these elements form a partition of W'.

B.3.6 Theorem. Let $(W, \mathcal{F}, P)$ be a probability space, and let $CH = \{ch_1, ch_2, \ldots\}$ be a countable set of probability functions over $\mathcal{F}$. Suppose there is a non-negative sequence a_n such that

$$\sum_n a_n ch_n(A) \lessdot P(A) \quad \text{for all } A \in \mathcal{F}\,. \tag{59}$$

Then P is CH-PP-locally-compatible.

Proof. Let a_i be as in the statement of the theorem. Applying (59) to the event $A = W$ we get

$$a \stackrel{\text{def}}{=} \sum_n a_n \;<\; 1, \tag{60}$$

$$q(A) \stackrel{\text{def}}{=} \sum_n a_n ch_n(A) \lessdot P(A) \quad \text{for all } A \in \mathcal{F}\,. \tag{61}$$

Define a "dummy" ch_0 by

$$ch_0(A) \stackrel{\text{def}}{=} \frac{1}{1-a}(P(A) - q(A)) \quad \text{for all } A \in \mathcal{F}\,. \tag{62}$$

By (61), ch_0 is non-negative for all $A \in \mathcal{F}$, and

$$ch_0(W) = \frac{1}{1-a}\big(P(W) - q(W)\big) = \frac{1}{1-a}(1-a) = 1, \tag{63}$$

thus p_0 is a probability over $\mathcal{F}$. By reordering (62) we get

$$P(A) = (1-a) \cdot ch_0(A) + q(A), \tag{64}$$

that is,

$$P(A) = (1 - \sum_n a_n) \cdot ch_0(A) + \sum_n a_n ch_n(A), \tag{65}$$

for every $A \in \mathcal{F}$. Thus, the conditions of Theorem B.3.2 are fulfilled with the set $CH' = \{ch_0\} \cup CH$. ∎

B.3.7 Theorem. Let $(W, \mathcal{F}, P)$ be a probability space, and let $CH = \{ch_1, ch_2, \ldots\}$ be a countable set of probability functions over $\mathcal{F}$. Suppose there is a non-negative sequence a_n such that for all $n \geq 1$ the following condition is satisfied:

$$a_n \cdot ch_n(A) \lessdot P(A) \quad \text{for all } A \in \mathcal{F}\,. \tag{66}$$

Then P is CH-PP-locally-compatible. By Remark B.3.1 the condition (66) is necessary.

Proof. It is enough to prove that (66) implies (59) (with a sequence a'_n). Indeed, take $a'_n = {}^{a_n}\!/_{2^n}$. Then

$$\sum_n a'_n ch_n(A) = \sum_n \frac{a_n}{2^n} ch_n(A) \lessdot P(A) \quad \text{for all } A \in \mathcal{F}\,. \tag{67}$$

∎

B.3.8 Corollary. Let $(W, \mathcal{F}, P)$ be an atomic probability space, and let $CH = \{ch_1, ch_2, \ldots\}$ be a countable set of probability functions over $\mathcal{F}$. P is CH-PP-locally-compatible iff it is CH-PP-globally-compatible.

B.4 SATISFYING THE PP "STRONGLY"

Definitions B.1.1, B.1.2 and B.1.3 feature "if defined" clauses to get around the cases in which the agent has 0 credence in some chance propositions. Here we gather definitions and results interesting if we wish to consider the case in which *all* the PP-relevant conditional probabilities are defined (which will by necessity for each A make some two different local chance propositions about A not disjoint).

B.4.1 Definition (satisfying the PP locally in the strong sense). Let P be a probability function defined on $\mathcal{F}$. We will say that P *satisfies the PP locally in the strong sense* if and only if for every $A \in \mathcal{F}$ with $0 < P(A) < 1$ and every $x \in [0,1]$ there is an event C_x^A such that $P(A \mid C_x^A) = x$.

B.4.2 Fact. For any P and CH, if P satisfies the PP locally in the strong sense, and $P(A) = 0$ implies $ch(A) = 0$ for every $ch \in CH$ and $A \in \mathcal{F}$, then P satisfies the PP locally with respect to CH.

B.4.3 Fact. Every non-atomic P satisfies the PP locally in the strong sense.

Proof. Suppose P is non-atomic, and $0 < P(A) < 1$. Let $x \in [0,1]$ be arbitrary. Take events $B_0 \subseteq A$ and $B_1 \subseteq \bar{A}$. By non-atomicity of P, any values $0 \leq P(B_0) \leq P(A)$ and $0 \leq P(B_1) \leq 1 - P(A)$ can be realized by appropriate choices of B_0 and B_1. If $C_x^A = B_0 \cup B_1$, then

$$P(A \mid C_x^A) = \frac{P(A \cap C_x^A)}{P(C_x^A)} = \frac{P(B_0)}{P(B_0) + P(B_1)}\,. \tag{68}$$

Appealing to continuity, this latter fraction can take any value from $[0,1]$ by suitable choices of B_0 and B_1. ■

B.4.4 Definition (strong local PP compatibility). Let $(W, \mathcal{F}, P)$ be a probability space. We say that P is *strongly PP-locally-compatible* if there exists an extension $(W', \mathcal{F}', P')$ of $(W, \mathcal{F}, P)$ and a σ-algebra embedding $h : \mathcal{F} \to \mathcal{F}'$ such that

- P and P' agree on events of $\mathcal{F}$, that is, for all $A \in \mathcal{F}$ we have

$$P'(h(A)) = P(A)\,; \tag{69}$$

- For every $A \in \mathcal{F}$ with $0 < P(A) < 1$ and any $x \in [0,1]$ there is an event C_x^A in $\mathcal{F}'$ such that

$$P'(h(A) \mid C_x^A) = x\,. \tag{70}$$

B.4.5 Theorem. Every probability space $(W, \mathcal{F}, P)$ is strongly PP-locally-compatible.

Proof. The space $(W, \mathcal{F}, P)$ can be embedded into the product of $(W, \mathcal{F}, P)$ with the Lebesgue probability space over $[0,1]$ via cylinder sets; and this product space is non-atomic, thus by Fact B.4.3 it satisfies the PP locally in the strong sense. ■

C THE PP AND THE PRINCIPLE OF INDIFFERENCE: A DETAILED ANALYSIS OF THE ARGUMENT FROM [12]

Our goal in this section is to clarify why we believe that producing an example of a probability space which satisfies the PP and exhibits no events of probability 0.5 shows that Hawthorne et al. have not proven that the PP implies the Principle of

Indifference (PoI). We will argue that satisfying their version of the PoI implies either that the probability space in questions contains 0.5-probability events or that it satisfies a counter-intuitive condition (**) below. Arguing this point requires the reconstruction of the whole argument from [12]; perhaps because that paper was intended to be just a short note, the argument is quite sketchy, especially with regard to quantification, and below we have present as charitable a version of it as we could. (We refrain from commenting on admissibility here.)

Reconstruction of [12]'s argument

Assumption:

"Principal principle: $P(A|XE) = x$, where X says that the chance at time t of proposition A is x and E is any proposition that is compatible with X and admissible at time t." (p. 123)

The contingency and logical atomicity assumption:

"[H]enceforth, F is any proposition that is contingent (neither necessarily true nor necessarily false) and atomic (not logically complex)" (p. 124)

Comment: We have no problems with this meaning of 'contingency'. However, the assumption of F being logically atomic (i.e. not logically complex) is – in that paper – deeply mysterious. One thing that's clear is that this atomicity is *not* intended to be the atomicity in the domain of P. However, what Hawthorne et al. mean by 'logically atomic' is not explained and, more importantly, *the notion plays no part in the supposed derivation of the Principle of Indifference from the Principal Principle*, apart from the suggestion in Footnote 1 that for a logically complex F, Conditions 1 and 2 (see below) may fail. This is inconsistent with the fact that Hawthorne et al. establish Condition 1 as a theorem (in their Proposition 1) by reasoning which makes no use of the internal structure (or lack thereof) of F.

Since Hawthorne et al.'s claims about the logical atomicity of F and its role in the reasoning are inconsistent, it would be natural to search for some positive information as to what they mean by the notion. However, the only thing we were able to find is the example from Footnote 1 of a logically complex proposition, "ticket number ninety-seven won a fair thousand-ticket lottery with one winner". We would not hazard a guess as to what the Hawthorne et al.'s definition of logical atomicity is based on that. If we were talking about some specified language, and assumed, say, that we had two propositional letters p and q, with the usual host of logical connectives, we could start by saying that there are two logically atomic sentences, p and q, while all other sentences are logically complex. But in [12] there are no assumptions about the language, and the authors talk about 'propositions'. Well, if propositions, as is often the case in formal epistemology, are (or at least correspond to) sets of possible worlds, then we're at a loss as to which of those sets are supposed to be logically atomic.[33] If pressed, we'd probably go in the direction of saying that the singletons, or, more generally, the atoms of the Boolean Algebra of propositions in the domain of P, are logically atomic, since all other nontrivial propositions are expressible using them, while they themselves are not divisible into other nonempty propositions. However, it is clear that Hawthorne et al. do not go in that direction. The most charitable reading of the argument with regard to the logical complexity issue seems, then, to be to:

- agree with Hawthorne et al.'s that their Proposition 1 is correct, and thus Condition 1 below is a theorem;
- disregard their statement that for a logically complex F Condition 1 might fail, as it is inconsistent with the above;
- to assume that, in the domain of the P we're talking about, there *are* some logically atomic propositions, whatever they ultimately are...
- ...and that the scope of the argument is supposed to be limited to them, since otherwise, the Authors claim, Condition 2 might fail.

[33]And if propositions are supposed to be, say, Russelian abstract n-tuples, well, Hawthorne et al. should tell us this!

Two additional assumptions:

"We shall take the claim that E is not a defeater to hold just when $P(A|XE) = x = P(A|X)$. Moreover, we shall take the supposition that XE contains no information that renders F relevant to A to imply that $P(A|FXE) = P(A|XE)$." (p. 124)

Comment: These assumptions only make sense if A and X are held fixed: some propositions are defeaters for some propositions, but not for others. We thus continue under the assumption that we are indeed in the context of some particular A and X.

Note that the second of these two assumptions is equivalent to saying that if XE contains no information that renders F relevant to A, then XE screens off F from A. There seems to be no justification in the paper for such a strong claim!

Two conditions:

"Condition 1: If E is not a defeater and XE contains no information that renders F relevant to A, then EF is not a defeater.

Condition 2: If E is not a defeater and XE contains no information relevant to F, then $E(A \leftrightarrow F)$ is not a defeater." (p. 124)

Comment: Condition 1 follows from Proposition 1, which seems to be correct given the definitions. There is no proof of Condition 2 offered; it is supposed to "encapsulate core intuitions about defeat" (p. 126) and footnote 1 suggests that it might fail if F is logically complex. The implicature here is that if F is *not* logically complex, then Condition 2 should hold. While we agree with [8] that Condition 2 ultimately fails, we do not rely on this in our arguments.

Proposition 2 (p. 125): Suppose that E, EF and $E(A \leftrightarrow F)$ are non-defeaters and that $0 < x < 1$. Then the Principal Principle implies that $P(F|XE) = 0.5$.

Comment: We agree that the proof of this Proposition 2 is correct, given the above definitions of the Principal Principle and of "being a non-defeater".

"Now suppose that E is a non-defeater and XE contains no information relevant to F or that renders F relevant to A. By Conditions 1 and 2, neither EF nor $E(A \leftrightarrow F)$ are defeaters. Hence, by Proposition 2, $P(F|XE) = 0.5$." (p. 125)

Comment: Correct.

"[S]ince XE contains no information relevant to F, $P(F|XE) = P(F)$, so $P(F) = 0.5$ too." (p. 125)

Comment: Note that this last step does not appeal to any previously stated assumptions, definitions, Conditions or Propositions; it seems to rely rather on our supposed intuition that for a rational initial credence P, for any two propositions B and C, if B contains no information relevant to C, then $P(C|B) = P(C)$.

End of the reconstruction of [12]'s argument

To sum up, the reasoning is as follows. Take an initial credence function P. Fix a proposition A and a proposition X which specifies the chance of A. Take a contingent, logically atomic proposition F. Assume the Principal Principle, the two Conditions and the definition of being a non-defeater. Now, if:

(*) there exists a non-defeater E such that XE contains no information relevant to F or that renders F relevant to A,

then – by using the two Conditions and Proposition 2 – $P(F) = 0.5$. Generalizing this, P should assign the value 0.5 to all the contingent logically atomic propositions for which the (*) condition holds, and since P is an initial credence, we arrive at a form of the Principle of Indifference.

Let us now home in on the goal of this Section, namely, on showing that upholding Hawthorne et al.'s claim while admitting the existence of PP-satisfying probability spaces with no propositions of probability 0.5 requires the satisfaction of a very unintuitive condition.

Assume, then, that you have on your hands a probability space satisfying the Principal Principle which includes no 0.5-probability propositions. For this not to be a counterexample to the reasoning under discussion, it has to be the case that for all contingent logically atomic propositions the existential condition (*) fails. That is, for any contingent logically atomic proposition F, there does not exist a non-defeater E which satisfies two additional requirements relating to X and A; recall that the definition of a non-defeater is also made in the context of that same fixed X and A. But to get to the desired conclusion of $P(F) = 0.5$ Hawthorne et al. may choose some other $B \neq A$ and a proposition Y specifying the chance of B; therefore, to avoid *that*, we need to stipulate that for any contingent logically atomic proposition F the condition (*) fails also when 'B' is substituted for 'A', 'Y' is substituted for 'X', and 'non defeater' is understood in the context of B and Y. And so on...

To generalize Hawthorne et al.'s non-defeater definition, say that E is a non-defeater for some arbitrary proposition A and chance proposition X which specifies the chance of A to be x just when $P(A|XE) = x = P(A|X)$. Then, in general, to uphold their reasoning in the face of probability spaces satisfying the Principal Principle in which there are no 0.5-probability propositions, the authors would have to claim that the following condition holds in such spaces:

(**) For any contingent, logically atomic proposition F, for any proposition A and a proposition X which specifies the chance of A, there does not exist a proposition E such that:

- E is a non-defeater for A and X;
- XE contains no information relevant to F;
- XE does not render F relevant to A.

It is only when this condition is satisfied that Hawthorne et al.'s reasoning is made vacuous, and then the fact of the existence of a probability space satisfying the Principal Principle without exhibiting any 0.5-probability propositions is indeed compatible with the claim that the Principal Principle entails the (Hawthorne et al.'s version of the) Principle of Indifference.

But we do not know why Hawthorne et al. would want to claim anything like this, that is, why should all probability spaces satisfying the Principal Principle which do not include 0.5-probability propositions satisfy the condition (**). We see no justification for it (even given that in some cases, which we discuss in Section 3, it is trivially satisfied, because the probability spaces in question feature no non-trivial non-defeaters at all). It seems arbitrary to us. But, to conclude, if that condition is not assumed, then a probability space satisfying the PP and not exhibiting any proposition with probability 0.5 does not satisfy the Principle of Indifference as modelled by Hawthorne et al., and thus their argument does not go through.