Admissibility and Bayesian direct inference: no HOPe against ubiquitous defeaters

Zalán Gyenis, Leszek Wroński

Jagiellonian University

January 29, 2020

Abstract

In this paper we discuss the "admissibility troubles" for Bayesian accounts of direct inference proposed in Wallmann & Hawthorne (2018), which concern the existence of surprising, unintuitive defeaters even for mundane cases of direct inference. We first show that one could reasonably suspect that the source of these troubles was informal talk about higher-order probabilities: for cardinality-related reasons, classical probability spaces abound in defeaters for direct inference. We proceed to discuss the issues in the context of the rigorous framework of Higher Probability Spaces (HOPs, Gaifman, 1988). However, we show that the issues persist; we prove a few facts which pertain both to classical probability spaces and to HOPs, in our opinion capturing the essence of the problem. In effect we strengthen the message from the admissibility troubles: they arise not only for approaches using classical probability spaces—which are thus necessarily informal about metaprobabilistic phenomena like agents having credences in propositions about chances—but also for at least one respectable framework specifically tailored for rigorous discussion of higher-order probabilities.

1 Introduction: defeaters for direct inference

In this paper we will be concerned with the notion of defeater as used in the context of direct inference (Levi, 1977): roughly, if a subject knows that the chance of A is x, and if he or she knows no other relevant information, then he or she should assign to A the degree of belief x. That is, with the "relevant information" proviso in mind, the rational degree of belief in A conditional on that the chance of A is x should be x. A defeater is then a proposition such that if it is additionally conditionalized upon, that is, if it is added to the proposition about chance to form the body of propositions given which the conditional subjective probability of A is considered, then that probability becomes something else than x. In other words, we call a piece of evidence a defeater if a certain conditional probability is not resilient under it.[1] This is how Hawthorne et al. (2017) formulate the idea (where P is a subjective probability function):

> "[Suppose] X says that the chance (...) of A is x (...) We shall take the claim that E is not a defeater to hold just when $P(A \mid XE) = x = P(A \mid X)$." (Hawthorne et al., 2017, p. 123-124)

It is typical to discuss the notion in the context of the Principal Principle ("PP", Lewis (1986)), which is where the notoriously vague notion of admissibility comes into play: using the notation as before, $P(A \mid XE) = x$ provided E is *admissible*. Without going into the jungle of details about admissibility, we hope it is evident at this point that whatever the notion really comes down to, we can agree that

[1] This is essentially how Lewis (1986) uses the notion of resiliency, see e.g. p. 85-86.

(keeping the notation introduced so far and being somewhat charitable with implicit quantification) if something is a defeater, then it is inadmissible, which is exactly the route taken in the recent paper by Wallmann & Hawthorne (2018).

The prevailing intuition among the authors writing about admissibility seems to be that inadmissibility should be somewhat rare, since it involves impacting the credence in A by some other way than impacting the credence about the chance of A (Lewis, 1986, p. 92). For examples, barring some very special cases, historical information and hypothetical information about chance should be admissible. The clear examples of inadmissible evidence occurring in the literature invariably involve some sort of soothsaying device. This is from where the results in (Wallmann & Hawthorne, 2018) receive at least a part of their bite: seemingly, defeaters abound! A key example concerns Maria, a craps player, who by way of direct inference assigns credence $1/6$ to the proposition that the outcome of the next toss of two fair dice will be seven, based on her knowledge of the chances involved. It turns out, Wallmann & Hawthorne claim, that something so seemingly innocent as the proposition uttered by John, who is standing nearby and says "I'll buy you dinner this evening if and only if the next toss comes up seven", is a defeater for that credence: given Maria's knowledge about the relevant chances, and that the next toss comes up seven if and only if John buys her dinner this evening, her credence that the next toss comes up seven is not equal to $1/6$. And so in this situation this biconditional is a defeater—an inadmissible proposition.

If Wallmann & Hawthorne—who proceed to generalize the above example to a theorem about inadmissible biconditionals—are right, then inadmissibility is a lot more frequent than everyone assumed. This may pose some troubles for Bayesian accounts of direct inference (though, it has to be mentioned, the two authors disagree on the import of their results).

We will begin by investigating the following worry: perhaps the fact that defeaters for direct inference are so easy to come by should come as no surprise? It may actually, the worry continues, be a consequence of using the traditional, Kolmogorovian notion of probability whereas a different approach should better be employed. The key problem is that the event algebra[2] of a classical probability space does not contain events which would in any formal sense be related to the idea that the probability of some event has some value. That is, for any event A, the probability space assigns it some probability $P(A)$, but if we want to use that space to model a degree of belief function of an agent who not only has a credence in A but also in that the probability of A is, say, .3, we seemingly have no structural features to turn to which would help us in identifying which of the elements of the event algebra is the proposition "that the probability of A is .3". Some additional structure involving higher order probabilities seems needed at this point.[3]

Still, many authors in modern formal epistemology proceed to, in a completely informal move, claim that some element of the event algebra *is* "the proposition that the chance of A is x" and then prove theorems which at their core are simply results about various conditional dependences in classical probability spaces and not really about propositions about probabilities (see e.g. the discussion of the Hawthorne et al. (2017) paper in Gyenis & Wroński (2017)).

In the following Section we will show that if we stick to the classical probability space approach then in quite frequent cases we should expect defeaters to abound, and so we should consider it doubtful whether the results by Wallmann & Hawthorne (2018) should really surprise us. This suggests that a serious discussion of probabilistic defeaters for direct inference should take place in some non-classical setting. In Section 3 we recall one such proposal by Gaifman, that of Higher Order Probability spaces (HOPs). Then, in Section 4, we return to Wallmann & Hawthorne's argument about biconditionals. It

[2] We will be using the notions "event" and "proposition" interchangeably in this paper, always refering to an element of the second element of some probability space (the field of sets).

[3] As one of the anonymous referees remarked, one could also go for, say, a different understanding of direct inference, such as the one proposed in Hall (1994).

turns out that their result actually carries over to HOPs. We eventually show what we believe to be the mathematical essence of the problem (Corollary 2 below), which pertains both to classical probabilities and to HOPs.

The take-away message of the paper is twofold: first, the discussions of defeaters taking place in formal epistemology should not employ classical probability spaces; and second, the admissibility troubles identified by Wallmann & Hawthorne persist even after making the move to HOPs, in our opinion casting doubts on the prospects of any theory modelling direct inference by conditional probability.

2 Defeaters in classical probability spaces

We will first show that if one wishes to use classical probability spaces to non-trivially speak about defeaters, one should better consider certain cardinality features (virtually never discussed in the literature) of the events in question. We are assuming that the traditional notion of a classical probability space is to be used, that is, a triple $(W, \mathcal{F}, P)$ consisting of a nonempty set, a field of its subsets, and a probability measure[4]. Wallmann & Hawthorne could perhaps respond that the phenomena we will be describing disappear if we consider probability functions defined not on fields of sets, but on languages, which they seem to suggest (see e.g. p. 2 of their paper). However, carrying out this move would require its proponents to precisely define an algebra of statements on which the probability function is defined on, and this in turn would require a formulation of some condition linking the proposition A with the proposition that the probability of A is (say) .3. None of this is present in the discussed paper.

When it comes to notation, assume the following: AB means $A \cap B$; $|A|$ is the cardinality of A; A' is the complement of A.

We will now define a general notion of defeater, applicable also outside of the context of direct inference. (To reiterate the point, from the perspective of a classical probability space unimbued with any additional structure the phenomenon of direct inference is essentially invisible, because no events are really events about probabilities: so, there is — at least not until the structure is carefully interpreted or extended — no defeat for direct inference, there is just non-resilient conditional probability.)

Definition 1. Assume a probability space $(W, \mathcal{F}, P)$ is given. For any $A, B \in \mathcal{F}$ such that $P(B) > 0$ an event $D \in \mathcal{F}$ is called a:

- *non-trivial defeater for* $P(A \mid B)$ if $P(BD) > 0$ and $P(A \mid B) \neq P(A \mid BD)$;
- *trivial defeater* for $P(A \mid B)$ if $P(BD) = 0$.

D is a defeater for $P(A \mid B)$ if it is either a trivial or a non-trivial defeater for $P(A \mid B)$. If D is neither a trivial nor a non-trivial defeater for $P(A \mid B)$ we say that it is a *non-defeater for* $P(A \mid B)$. □

If there is no risk of confusion about the measure, we will use the terms "defeater for $P(A \mid B)$" and "defeater for A and B" interchangeably. Just to be sure concerning how negation works here: if something is not a non-defeater for A and B then it is a defeater for those events; also, any defeater for A and B is not a non-defeater for those events.

[4] We assume the traditional Kolmogorov axioms: all probabilities are nonnegative, $P(W) = 1$, and for any disjoint $A, B \in \mathcal{F}$, $P(A \cup B) = P(A) + P(B)$. (This last condition needs to be modified in infinite structures, which we do not explicitly consider in this paper.) Note also that we assume the textbook definition of conditional probability: $P(A|B)$ is defined as $P(AB)/P(B)$ if $P(B) > 0$ and is undefined otherwise.

Are defeaters rare? Should we expect conditional probabilities to be resilient? We will now argue that in finite uniform probability spaces, for (literally) most pairs of events, (literally) most events are defeaters. Therefore, in general, we should be surprised if something is *not* a defeater for some given two events, and not when it is.

Claim 1. Suppose a finite probability space $(W, \mathcal{F}, P)$ is given with P uniform, $|W| = N$. Take two events $A, B \in \mathcal{F}$, such that $P(B) > 0$. Let $\mathrm{NonDef}(N, k, l)$ denote the number of non-defeaters for A and B, where $k = |AB|$ and $l = |A'B|$. Then

$$\mathrm{NonDef}(N,k,l) = \begin{cases} 2^{N-l}(2^l - 1) & \text{if } k = 0, l > 0; \\ 2^{N-k}(2^k - 1) & \text{if } k > 0, l = 0; \\ 2^{N-(k+l)} \sum\limits_{n \in G(k,l)} \binom{k}{n}\binom{l}{\frac{nl}{k}} & \text{if } k > 0, l > 0. \end{cases} \tag{1}$$

(Note that $\mathrm{NonDef}(N, 0, 0)$ is undefined as we require $P(B) > 0$.)

We exile the proof to the Appendix.

Let us return to the original question: for given N, k, l, which of the two, $\mathrm{NonDef}(N, k, l)$ or $\mathrm{Def}(N, k, l)$, is higher? ($\mathrm{Def}(N, k, l)$ is the respective number of defeaters.)

In the case of disjoint A and B (i.e. $k = 0$) or the case of $B \subseteq A$ (i.e. $l = 0$), almost all events are non-defeaters. If one of the sets $|AB|$ and $|A'B|$ is large enough (but the other is empty), then the proportion of the cardinality of the set of nondefeaters to 2^N is very close to 1. (It is approximately 0.97 already if that cardinality is as low as 5.) However, we can dismiss these cases as unimportant to our argument, for two reasons.

First, in the context of direct inference, we are interested in nontrivial conditionalization, where some proposition A and some proposition specifying the chance of A are such that, first, they are in general not mutually exclusive, and second, the latter does not logically imply the former. We are pursuing here a general mathematical question about the resiliency of conditional probability, but the two cases outlined above will not be important for the philosophical goal.

Second, such cases are in a precise sense rare. Suppose we chose two subsets (at random with uniform probability) A and B of W. What is the probability that they are disjoint (respectively, $B \subseteq A$)? We have 3^N possibilities to choose a disjoint pair, as for each element in W we have three choices: in A; in B; in neither A nor B (respectively, 3^N possibilities to choose a pair with $B \subseteq A$, as for each element in W we have three choices again: in A but not in B; in both; in neither). There are 2^{2N} possibilities in total to choose a pair of events, therefore the probability that A and B are disjoint (respectively, $B \subseteq A$) is ${}^{3^N}\!/_{2^{2N}} = ({}^3\!/_4)^N$. As ${}^3\!/_4 < 1$, for large enough N this probability gets very close to 0. (It equals approximately 0.06 for $N = 10$ already.) This means it is very unlikely to pick sets at random that are disjoint (resp. one contains the other), provided the sample space is large enough.

To reiterate the point, in Section 3 we move away from classical probability spaces towards structures tailored for meta-level probabilistic phenomena. There A and B are not chosen independently: B will be the event $pr(A, \alpha)$ expressing "the probability of A is α" (cf. Definition 2). It can be shown that A and $pr(A, \alpha)$ cannot be disjoint, provided Miller's principle holds, which will be the axiom (VI_w) on p. 7 below. In the remainder of this section we concentrate, then, on the "non-trivial" case only, when $k, l > 0$.

Let us then proceed, then, under the assumption that A and B are not disjoint. We know how many nondefeaters there are, namely, $\mathrm{NonDef}(N, k, l)$. Is this number lower or higher than the number of defeaters, $\mathrm{Def}(N, k, l)$? Let us now assume that for given A and B (with $|AB|, |A'B| > 0$) we randomly choose an event D with uniform probability Prob over all events. Which is more likely: that D is a defeater, or that it is a non-defeater? Let n be such that $|ABD| = n$. Our counting arguments above allow us to express the precise value of the probability that D is a non-defeater for A and B, namely,

$$\text{Prob}(D \text{ is a non-defeater for } A, B) = \frac{\sum_{n \in G(k,l)} \binom{k}{n}\binom{l}{\frac{nl}{k}}}{2^{k+l}}. \tag{2}$$

Note that the probability of D being a (non)-defeater (for A and B) depends only on the size of AB and $A'B$. How small is this probability? The general answer to this question seems hard[5], so we discuss below a single easy-to-handle case which turns out to be more likely than not.

Assume k and l are coprime. Then $G(k,l) = \{k\}$, and thus

$$\sum_{n \in G(k,l)} \binom{k}{n}\binom{l}{\frac{nl}{k}} = \binom{k}{k}\binom{l}{l} = 1.$$

The only (and all) D's which are *not* defeaters for A and B are such that $B \subseteq D$. The proportion of non-defeaters for A and B is then

$$\text{Prob}(D \text{ is a non-defeater for coprime } k, l) = \frac{\sum_{n \in G(k,l)} \binom{k}{n}\binom{l}{\frac{nl}{k}}}{2^{k+l}} = \frac{1}{2^{k+l}} = \frac{1}{2^{|B|}}.$$

If B is a singleton, then $\frac{1}{2^{|B|}}$ equals .5. But in this case every D avoiding B is a trivial defeater and every D containing B is a non-defeater, thus, there are no non-trivial defeaters for singleton B's. In all other cases ($|B| > 1$) the value of the fraction is at most $\frac{1}{2}$. The value "quickly" diminishes if the cardinality (of $|B|$) increases.

Among randomly selected integers two co-primes are *not* so hard to come by: If we consider sets of integers $\{1, \ldots, N\}$, then with N approaching infinity the probability of randomly selecting two co-primes from such a set is $6/\pi^2$ (Hardy & Wright, 2008, Theorem 332, p. 354), that is, about .61. Elaborated number-theoretical results[6] show that this probability is always strictly larger than .5. The situation, then, is more likely than not!

To recap: if you are choosing two events at random from a finite probability space (say, you go over all elements of the sample space and individually decide to include them in the given event or not depending on a fair coin toss) with the uniform measure, you should expect ending up with two events such that whatever third event you draw, it is more often than not a defeater for them: if you are not lucky, you will end up with two events such that a vast majority of events in the space will be defeaters for them.

There is at least one obvious generalization which the Reader might feel to be needed at this point, namely, one could loosen the uniformity assumption regarding the measure. We do not wish to consider these issues here[7] because our results above already persuasively suggest the main philosophical point we would like to make: before investigating any defeater-related phenomena in classical probability spaces,

[5] Let us note that $G(k,l)$ can be rewritten as $G(k,l) = \{x \cdot k/\text{gcd}(k,l) : \ 0 < x \leq \text{gcd}(k,l), x \in \mathbb{N}\}$, thus $|G(k,l)| = \text{gcd}(k,l)$. Assuming $k = l$ the sum $\sum_{n \in G(k,l)} \binom{k}{n}\binom{l}{nl/k}$ simplifies to $\sum_{n \leq k} \binom{k}{n}^2$ which is equal to $\binom{2k}{k}$. Calculus shows $\lim_{k \to \infty} \binom{2k}{k}/2^{k+k} = 0$, consequently for large enough $k = l$ the probability Prob(D is non-defeater for A, B) is very close to 0, way much smaller than $1/2$. We conjecture that $\lim_{k+l \to \infty} \sum_{n \in G(k,l)} \binom{k}{n}\binom{l}{nl/k} = 0$ and thus the argument that Prob(D is non-defeater for A, B) is negligible can be extended to all the cases, irrespective of the value of k and l. We do not pursue this issue any further here.

[6] Namely, results by Mertens 1874 and Erdős and Shapiro 1951. Mertens proves that the number of coprimes in $\{1, \ldots n\}$ is $S(n) = \frac{6}{\pi^2}n^2 + O(n \log n)$. The probability of two coprimes is then $P(n) = \frac{S(n)}{n^2}$. According to Erdős and Shapiro, the error function $S(n) - \frac{6}{\pi^2}n^2$ is in between $\pm cn \log\log\log\log n$ for some constant c. Then, the error of $P(n)$ is smaller than $\frac{\log\log\log\log n}{n}$, which is smaller than .005 for all $n > 1$. Thus $P(n)$ is always larger than $\frac{6}{\pi^2} - .005 > .5$.

[7] It is worth mentioning that the number of defeaters goes up if the probability function P is not uniform: with other

and certainly before we allow ourselves to be surprised by a general theorem regarding the apparent abundance or lack of defeaters, we should make sure that the relationships between the events involved are not such that they would trivialize the matters. Who knows? Maybe for some deep metaphysical reason the cardinalities of the proposition A and any proposition about the chance of A are such that almost any event will be a defeater for them (assuming uniformity of the measure; we already know that most of the events are defeaters for them). Can we exclude that? Should we? We don't know, but in our opinion it seems a lot more fruitful to move away from classical probability spaces and towards structures tailored for meta-level probabilistic phenomena. An example of such an approach is that of Higher Order Probability spaces from Gaifman (1988), to which we now turn.[8]

3 Defeaters in Higher-Order Probability Spaces

We will abbreviate "Higher-order probability space" with "HOP". To quote (Gaifman, 1988, p. 197), a HOP is a 4-tuple $(W, \mathcal{F}, P, pr)$—where $\mathcal{F}$ is a field of subsets of W, to be called "events", P is a probability over $\mathcal{F}$ and pr is a mapping associating with every $A \in \mathcal{F}$ and every real closed interval Δ an event $pr(A, \Delta)$—which satisfies axioms (I)-(V) below. The initially intended interpretation is that P is the agent's subjective probability and $pr(A, \Delta)$ is the event that "the expert probability of A lies in Δ".

We will adopt the convention that $\mathbf{1} = W$ and $\mathbf{0} = \emptyset$ and will omit the curly brackets when dealing with singletons without further commentary (e.g., $pr(w_1, [.5, .6])$ means $pr(\{w_1\}, [.5, .6])$). Also, closed intervals may be single points, and when talking about such cases we will use α instead of Δ: $pr(A, \alpha)$ makes sense for $\alpha = .3$ and means then the same thing as $pr(A, [.3, .3])$.

Each HOP satisfies the following five axioms, which are actually axiom schemes, with implicit universal quantification over A (see Gaifman (1988)):

(I) $pr(A, [0, 1]) = pr(W, [1, 1]) = \mathbf{1}$;

(II) $pr(A, \emptyset) = \mathbf{0}$;

(III) If $\Delta_1 \cup \Delta_2$ is an interval, then

$$pr(A, \Delta_1 \cup \Delta_2) = pr(A, \Delta_1) \cup pr(A, \Delta_2);$$

(IV) $\cap_n pr(A, \Delta_n) = pr(A, \cap_n \Delta_n)$;

(V) If, for all $n \neq m$, $A_n \cap A_m = \emptyset$, then:

$$\cap_n pr(A_n, [\alpha_n, \beta_n]) \subseteq pr(\cup_n A_n, [\textstyle\sum_n \alpha_n, \sum_n \beta_n]).$$

Let us briefly mention two points which might not be evident on first reading of Gaifman's paper:

- even though the author writes about extending the function pr so that it would deal with arbitrary (Borel) subsets of $\mathbb{R}$ (p. 199), the axioms only work for convex sets, and in fact the only generalization explicitly considered is one which allows the Δ to be a half-open or open interval (p. 201);

priors P one would expect to have many more defeaters. We thank this comment to one of the anonymous referees. Further, that a randomly chosen event should be expected to be a defeater seems obvious "in the limit": when one increases the cardinality of the sample space (i.e. N approaches infinity) the proportion of pairs that are uncorrelated goes to zero. Thus, intuitively, it is very improbable to stumble upon a non-defeater. We owe this intuitive explanation to Bartosz Wcisło. We do not pursue these issues any further here.

[8] Note that the paper, despite the identical title, is not a simple reprint of Gaifman (1986). We recommend reading the later version, even if it is the earlier one which has been carefully LaTeXed and recently reprinted in Arló-Costa et al. (2016).

- Gaifman calls anything which satisfies axioms (I)-(V) a HOP, even if it does not satisfy (VI) or (VI_w), that is, even if no condition connecting P and pr is stipulated to hold in the structure under consideration.

HOPs become workable when we deal with their *kernels* (see (Gaifman, 1988, p. 199)). A kernel of a HOP (having finite W) is a $|W| \times |W|$ matrix of real numbers interpreted as a mapping p which associates with each $x \in W$ a probability p_x over $\mathcal{F}$ such that

$$pr(A, \Delta) = \{x : p_x(A) \in \Delta\}. \tag{3}$$

In a kernel, a row corresponding to some $x \in W$ contains the values of p_x for all singletons of elements of W (in effect defining p_x as a probability on $\mathcal{F}$; a kernel is, then, a *stochastic* matrix, that is, each entry is non-negative and the sum of each row is 1). That axioms (I) - (V) suffice for the existence of kernels so understood, with p_x's connected with pr as stated in equation (3), is the subject of Gaifman's Theorem 1. For later purposes we introduce the notion of a *reduced kernel*. Let W' contain $x \in W$ if and only if $P(x) = 0$. The reduced kernel is obtained from the kernel matrix by deleting columns and rows corresponding to $x \in W'$. The reduced kernel is thus the kernel "modulo P-zero".

For a formally trivial, but perhaps conceptually beneficial observation, note that for each choice of A and Δ, $pr(A, \Delta)$ is a single event (possibly the empty set). Therefore, facts such as "the expert probability of B lies in $[.2, .4]$" are represented by one and only one event in $\mathcal{F}$.

We will now introduce formally the notion of a defeater in the context of HOPs—a "HOP-defeater".

Definition 2. Assume a HOP $(W, \mathcal{F}, P, pr)$ is given. For any $A \in \mathcal{F}$ and closed real interval Δ such that $P(pr(A, \Delta)) > 0$ an event $D \in \mathcal{F}$ is called a

- *non-trivial HOP-defeater for A and Δ* if $P(pr(A, \Delta) \cap D) > 0$ and

$$P(A \mid pr(A, \Delta) \cap D) \notin \Delta; \tag{4}$$

- *trivial HOP-defeater for A and Δ* if $P(pr(A, \Delta) \cap D) = 0$.

D is a HOP-defeater for A and Δ if it is either a trivial or a non-trivial HOP-defeater for A and Δ. □

Note that the conventions we adopted dictate that a D is a HOP-defeater for A and α if $P(A \mid pr(A, \alpha) \cap D) \neq \alpha$.

What questions can be meaningfully asked about defeaters in a given HOP depends on whether it satisfies any of the following two axioms (Gaifman, 1988, p. 200), of which (VI) logically implies (VI_w):

(VI) If C is a finite intersection of events of the form $pr(B_i, \Delta_i)$, and if $P(pr(A, \Delta) \cap C) \neq 0$, then:

$$P(A \mid pr(A, \Delta) \cap C) \in \Delta.$$

(VI_w) If $P(pr(A, \Delta)) \neq 0$, then:

$$P(A \mid pr(A, \Delta)) \in \Delta.$$

If we consider single point intervals, (VI_w) becomes Miller's Principle: $P(A \mid pr(A, \alpha)) = \alpha$.

Note that if a HOP satisfies (VI), then its reduced kernel is an idempotent matrix (i.e. it is equal to its own square)[9]. The reduced kernel of a HOP satisfying (VI) is, then, an idempotent stochastic matrix.[10]

The literature citing the Gaifman paper, and indeed the original paper itself, contains very few concrete examples of worked-out HOPs. It might be beneficial to consider one at this point.

Example 1. Suppose we would like to model the credence function of an agent who is interested in the outcome of a coin toss, where the coin might be of one of four types: it might have two Heads, two Tails, it might be fair, or it might slightly favour Tails. Suppose the agent is initially indifferent when it comes to the type to which the coin belongs. Consider a HOP $(W, \mathcal{F}, P, pr)$ with $W = \{w_i\}_{i\in\{1,\ldots,6\}}$ and $\mathcal{F} = \mathcal{P}(W)$, in which the pr given by the following kernel:

worlds:	w_1	w_2	w_3	w_4	w_5	w_6
weight \ toss	H	T	H	T	H	T
.25	1	0	0	0	0	0
.25	0	1	0	0	0	0
.125	0	0	.5	.5	0	0
.125	0	0	.5	.5	0	0
.125	0	0	0	0	.4	.6
.125	0	0	0	0	.4	.6
P:	**.25**	**.25**	**.125**	**.125**	**.1**	**.15**

The credence function P is generated by taking a mixture of the rows of the kernel with the weights displayed in the leftmost column. (In the other examples in the paper we will adopt this convention where both the columns and rows of the kernel correspond to consecutively numbered worlds from W.) This HOP satisfies the axiom (VI). In w_1 the coin has two Heads; in w_2 it has two Tails. In both w_3 and w_4 it is fair, but in w_3 the result of the toss is Heads, and in w_4 it is Tails. In w_5 and w_6 the coin slightly favours Tails. For any type of the coin, the agent's credence that the coin belongs to that type is .25.

The agent's prior credence in Heads is $P(H) = .475$. Consider now the proposition that the coin favours Tails, either extremely or slightly. This is referred to by the expression $pr(H, [0, .5))$ and is the set $\{w_2, w_5, w_6\}$. Given that information, the agent's credence in H is modified in accordance with axiom (VI): as the Reader can easily verify, $P(H|pr(H, [0, .5)) = .2$. ♣

Lets us return to the issue of defeaters. Now, if an event C is a finite intersection of events of the form $pr(B_i, \Delta_i)$ and the HOP satisfies (VI), then C cannot be a non-trivial HOP defeater for any event A. It will already be interesting to consider which events are not of the form $pr(A, \Delta)$ for any possible choice of A and Δ:

Definition 3. Assume a HOP $H = (W, \mathcal{F}, P, pr)$ is given. We will say that an event $E \in \mathcal{F}$ is *metarepresentable in* H if and only if there exist $A \in \mathcal{F}$ and a closed real interval Δ such that $E = pr(A, \Delta)$. □

[9] We stress that the result predicates idempotence of the reduced kernel only. Example 2 (Gaifman, 1988, p. 206) features a HOP which satisfies (VI) and has a kernel which is not idempotent, but its reduced version is.

[10] If the HOP satisfies axiom (VI_w), then P (considered as a row-vector) is a left-eigenvector of the kernel matrix corresponding to eigenvalue 1. There is a serious typo in (3_d) of Lemma 1 in Gaifman (1988). The correct equation is $P(y) = \sum_x p(x, y) \cdot P(x)$, thus P is a left-eigenvector of the matrix p, instead of being a (right-)eigenvector. A HOP satisfies axiom (VI) if and only if its *reduced* kernel is idempotent and P is a left-eigenvector of the kernel (Gaifman, 1988, Theorem 2).

Example 2. Consider a HOP $(W, \mathcal{F}, P, pr)$ with $W = \{w_1, w_2, w_3\}$, $\mathcal{F} = \mathcal{P}(W)$, P uniform, and pr given by the following kernel:

$$\begin{array}{ccc} .5 & .5 & 0 \\ 0 & .5 & .5 \\ .5 & 0 & .5 \end{array}$$

(This is Gaifman's Example 1, p. 208). A direct check will ensure the Reader that in this HOP all events are metarepresentable. Axiom (VI) fails, while axiom (VI$_w$), and so Miller's principle, holds. ♣

It turns out that satisfaction of axiom (VI) is enough for some events not to be metarepresentable, and what's more, for some events not to be intersections of families of metarepresentable events. In what follows we recall (Högnäs & Mukherjea, 2011, Theorem 1.16) about the shape of idempotent stochastic matrices. Let M be a $d \times d$ idempotent stochastic matrix of rank $k > 0$. Then there is a unique partition of $\{1, \dots, d\}$ into classes $\{T, C_1, \dots, C_k\}$[11] such that the following hold (see Figure 1):

1. $T = \{i : \text{ the } i^{\text{th}} \text{ column of } M \text{ is a zero column}\}$
2. $M \restriction_{C_s \times C_s}$ has identical positive rows of sum 1 and $M \restriction_{C_s \times C_t} = 0$ for $s \neq t$
3. If $i \in T$, then

$$\frac{M_{ij}}{M_{jj}} = \frac{M_{ih}}{M_{jh}}, \quad j, h \in C_s \tag{5}$$

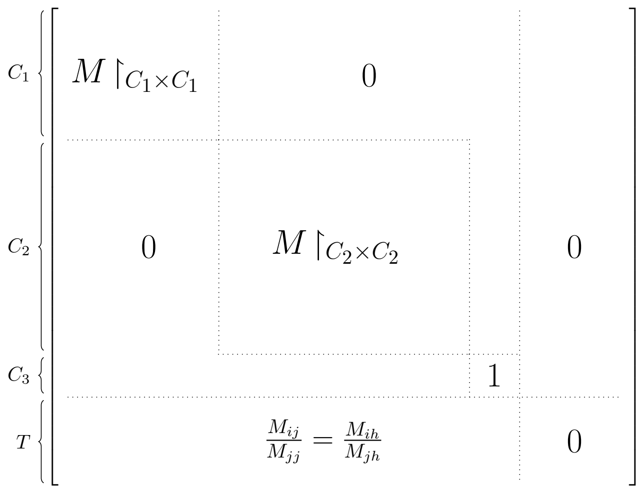

Figure 1: The form of an idempotent stochastic matrix of rank 3

(Notice that T or the C_i's might contain non-consecutive numbers (rows). In this sense Figure 1 might be a little misleading, as the blocks $M \restriction_{C_1 \times C_1}$ might not be "connected". On the other hand, there always exists a permutation of $\{1, \dots, d\}$ such that after it is applied to the rows and columns of the matrix M, the matrix achieves the shape as displayed in Figure 1.)

The "partition" $\{T, C_1, \dots, C_k\}$ can be of course thought of as a set of sets of rows of M, which will be the reading we ask the Reader to adopt in the proof of the following lemma.

Lemma 1. Assume a HOP $(W, \mathcal{F}, P, pr)$ is given. If this HOP satisfies (VI) and its reduced kernel is not the identity matrix, then some event E in $\mathcal{F}$ with $P(E) > 0$ is not metarepresentable.

[11] *Caveat*: as will be clearly visible soon, T may be empty, so this is not a partition in the strict set-theoretic sense.

Proof. Since the HOP satisfies (VI), then, as already mentioned, its reduced kernel M is an idempotent stochastic matrix. Take the partition $\{T, C_1, \ldots, C_k\}$ of the reduced kernel according to (Högnäs & Mukherjea, 2011, Theorem 1.16). We note first that T of this partition must be empty. As axiom (VI) is satisfied, P is the mixture of the rows of the kernel. Suppose the i^{th} column of the kernel is filled with 0's. Then $P(w_i)$ must be zero as it is a mixture of 0's. It follows that when creating the reduced kernel we delete the i^{th} row and column of the kernel.

As the reduced kernel is not the identity matrix, it is impossible that $|C_s| = 1$ for all s. Thus, there is at least one C_s having at least two elements, say $i, j \in C_s$. As rows of C_s are identical, it is immediate to note from equation (3) that for any A and Δ, if w_i belongs to $pr(A, \Delta)$, then $w_j \in pr(A, \Delta)$, too. Thus $E = \{w_i\}$ is not metarepresentable. ∎

(Note that, in contrast with (VI), (VI$_w$) does *not* imply non-metarepresentability: Example 2 describes a HOP which satisfies (VI$_w$), whose kernel contains no linearly dependent rows, and in which all events are metarepresentable.)

We now know that it follows from (VI) that some events are not metarepresentable. But to assess whether the notion of a HOP-defeater is not trivialised, we need to check whether some events are not finite intersection of metarepresentable events (which is a weaker condition, since *a priori* some events which are not themselves metarepresentable might be finite intersections of metarepresentable ones). Fortunately this is also the case in general.

Fact 1. Assume a HOP $(W, \mathcal{F}, P, pr)$ is given. If this HOP satisfies (VI) and its reduced kernel is not the identity matrix, then some event E in $\mathcal{F}$ with $P(E) > 0$ is not an intersection of metarepresentable events.

Proof. Define a function $pr : W \to \mathcal{F}$ as follows:

$$pr(w) := \{pr(A, \Delta) : w \in pr(A, \Delta)\}. \tag{6}$$

$pr(w)$ is the collection of metarepresentable events that contain w. If in the given HOP's kernel rows corresponding to distinct worlds w and v are identical, then $p_w = p_v$ and so for any $A \in \mathcal{F}$ and closed real interval Δ it holds that $w \in pr(A, \Delta)$ iff $v \in pr(A, \Delta)$. That is, when w and v correspond to identical rows, $pr(w) = pr(v)$. Then the singleton $\{w\}$ is not only not metarepresentable but also is not an intersection of metarepresentable events: if it was, then for some family of events of the form $pr(B_i, \Delta_i)$ the singleton $\{w\}$ would be the set of all elements belonging to each event from the aforementioned family, but then that singleton would also contain v: contradiction. We know, then, that if a HOP satisfies (VI), then if its kernel contains identical rows, it follows that some events are not intersections of metarepresentable events.

For the remaining case, assume a HOP satisfies (VI) but its kernel does not contain identical rows. Consider its reduced kernel and its partition $\{T, C_1, \ldots, C_k\}$ from (Högnäs & Mukherjea, 2011, Theorem 1.16) which we have just used. Recall that in this case, since we are talking about the reduced kernel, T is empty. Since no rows are identical, the reduced kernel is the identity matrix. ∎

On the one hand, Fact 1 brings a positive message: assuming (VI)—which, if P is interpreted as degree of belief, is a requirement of rationality, for which a coherence argument in the form of a Dutch Book construction is given in Gaifman's paper on p. 201-204, and which is similar in spirit to Lewis' Principal

Principle (see p. 201 of Gaifman's paper and its 4^{th} footnote)—guarantees that problems regarding defeaters will not be trivialised just by structural infelicities. To be sure, some cardinality-related troubles similar to the ones we talked about in the previous section might probably be reformulated in the context of HOPs. Such problems, if they are found, will not, however, originate from the label "the proposition that the chance of A belongs to Δ" being slapped on some more or less arbitrarily chosen event. On the contrary, in the context of HOPs we can say something substantial about the relationship of A and $pr(A, \Delta)$. A big part of it is captured by (VI$_w$): P has to satisfy a certain conditional probability requirement. Axiom (VI) adds to this the requirement that certain events cannot be HOP-defeaters: namely, events of the form $pr(B_i, \Delta_i)$ for some B_i and Δ_i—that is, metarepresentable events—and finite conjuctions of them. (In the Lewisian parlance: chance information is always admissible.) It seems the interested parties can continue philosophizing assured that they have a formal grasp of how a degree of belief function of a subject who satisfies the Principal Principle looks like, without resorting to hand-waving about admissibility or being informal when talking about how some events are actually "about" chances of other events.

On the other hand, if, as suggested by Gaifman himself (p. 193), we take rows to signify objective chance functions at various possible worlds, we seem to have stumbled on a weird consequence: there is an *a priori* constraint on the possible relationships between different objective chance functions. We are planning to revisit these issues in a paper devoted exclusively to the details of HOPs, and in the current article we will now return to the "admissibility troubles" which according to Wallmann & Hawthorne plague Bayesian accounts of direct inference.

4 Admissibility troubles revisited and strengthened

Recall the problem posed in (Wallmann & Hawthorne, 2018): Bayesian accounts of direct inference seem to be in trouble, since seemingly innocent propositions turn out to be defeaters. In Section 2.1 of their paper the authors tell a story about supposedly inadmissible biconditionals. If a fair pair of dice is tossed on a flat surface in a fair way, the chance that the outcome of the toss is seven is $1/6$. Maria, the subject, forms a direct inference, and sets her credence in that the outcome of the next toss is seven conditional on that setup to $1/6$. Nothing out of ordinary so far.

But now John says to Maria "I'll buy you dinner this evening if and only if the next toss comes up seven". A surprising claim of Wallmann & Hawthorne is that if Maria conditionalises additionally on *that*, her credence in that the next toss comes up seven moves away from $1/6$. The result is that the seemingly innocent biconditional ends up being a defeater for the probability that the next toss comes up seven given that the chance setup is like described above—which is highly unintuitive. Such biconditionals should not be defeaters for such probabilities; in this simple case, because we can assume that Maria believes that having dinner with John in the future is not probabilistically relevant for the outcome of the toss. We will now see how we can attempt to model this situation using HOPs.

Example 3. For the first model we will assume that Maria initially gives credence .5 to the proposition that John buys her dinner that evening. 4 worlds will suffice; in this and all later examples, the event algebra is the powerset of the set of worlds. The worlds are as follows:

	next toss is 7?	John buys dinner?
w_1	yes	yes
w_2	no	yes
w_3	yes	no
w_4	no	no

and the kernel is as follows:

$$\begin{matrix} 1/12 & 5/12 & 1/12 & 5/12 \\ 1/12 & 5/12 & 1/12 & 5/12 \\ 1/12 & 5/12 & 1/12 & 5/12 \\ 1/12 & 5/12 & 1/12 & 5/12 \end{matrix}$$

with P being any mixture of the rows, ending up of course as the vector $(1/12, 5/12, 1/12, 5/12)$. Call the proposition "next toss comes up seven" N and "John buys Maria dinner" D. From the setup we see immediately that $N = \{w_1, w_3\}$, $D = \{w_1, w_2\}$, $N \leftrightarrow D = \{w_1, w_4\}$, $pr(N, 1/6) = W$, and so $P(N \mid pr(N, 1/6)) = 1/6$, in accordance with (VI_w). What's more, $N \leftrightarrow D$ is not a HOP-defeater for N and $1/6$, since $P(N \mid pr(N, 1/6) \cap N \leftrightarrow D) = 1/6$, just like it should. ♣

This approach can be generalised in some natural ways. For example, Maria can entertain various hypotheses about the chance function, say, that the dice can be fair or skewed in a specific way. (We do keep the assumption that Maria's credence in John buying her dinner is .5 no matter the chance function—this will, unfortunately, turn out to be crucial.)

Example 4. Assume the setup is similar, but now Maria considers two chance hypotheses: that the dice are fair (which she gives credence $2/3$) or that they are skewed so that the chance of the next toss coming up seven is $1/4$ (let us call it the "skewed chance function"). We will use 8 worlds, which are as follows:

	next toss is 7?	John buys dinner?	chance?
w_1	yes	yes	fair
w_2	no	yes	fair
w_3	yes	no	fair
w_4	no	no	fair
w_5	yes	yes	skewed
w_6	no	yes	skewed
w_7	yes	no	skewed
w_8	no	no	skewed

and the kernel is as follows:

$$\begin{matrix} 1/12 & 5/12 & 1/12 & 5/12 & 0 & 0 & 0 & 0 \\ 1/12 & 5/12 & 1/12 & 5/12 & 0 & 0 & 0 & 0 \\ 1/12 & 5/12 & 1/12 & 5/12 & 0 & 0 & 0 & 0 \\ 1/12 & 5/12 & 1/12 & 5/12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/8 & 3/8 & 1/8 & 3/8 \\ 0 & 0 & 0 & 0 & 1/8 & 3/8 & 1/8 & 3/8 \\ 0 & 0 & 0 & 0 & 1/8 & 3/8 & 1/8 & 3/8 \\ 0 & 0 & 0 & 0 & 1/8 & 3/8 & 1/8 & 3/8 \end{matrix}$$

with P being a mixture of the rows of the kernels with weights $1/6$ (for the first four rows) and $1/12$ (for the next four rows), ending up with the vector $(1/18, 5/18, 1/18, 5/18, 1/24, 3/24, 1/24, 3/24)$. As before, call the proposition "next toss comes up seven" N and "John will buy Maria dinner" D. From the setup we see immediately that $N = \{w_1, w_3, w_5, w_7\}$, $D = \{w_1, w_2, w_5, w_6\}$, $N \leftrightarrow D = \{w_1, w_4, w_5, w_8\}$, $pr(N, 1/6) = \{w_1, w_2, w_3, w_4\}$, $pr(N, 1/4) = \{w_5, w_6, w_7, w_8\}$ and so $P(N \mid pr(N, 1/6)) = 1/6$ and $P(N \mid pr(N, 1/4)) = 1/4$, in accordance with ($\mathrm{VI}_w$). What's more, $N \leftrightarrow D$ is not a HOP-defeater for N and $1/6$, since $P(N \mid pr(N, 1/6) \cap N \leftrightarrow D) = 1/6$; it is also not a HOP-defeater for N and $1/4$, since $P(N \mid pr(N, 1/4) \cap N \leftrightarrow D) = 1/4$, just like it should. ♣

In our opinion all the intuitive independencies arising in the situation are modelled by this HOP and no unintuitive dependencies are introduced by conditionalising on the $N \leftrightarrow D$ proposition. Consider, still, that D is initially probabilistically independent of whether the chance function is fair or skewed, but ends up being dependent after conditionalization on $N \leftrightarrow D$. This is as it should be: Maria may initially think that whether John buys her dinner has no bearing on the toss result, and *vice versa*, but once John makes his promise, the details of the chances start being relevant to the dinner plans!

At this point the prospects of using HOPs to model direct inference might seem bright: employing higher-order structures gives as a real handle on propositions about probabilities of propositions. We will now see, though, that the phenomena uncovered by Wallmann & Hawthorne persist even after we leave behind the purely classical language. The issue can be seen already in a 4-world example:

Example 5. Let us attempt to generalize the structure from Example 3 so that Maria is permitted an arbitrary prior in that John buys her dinner that evening. Again, call the proposition "next toss comes up seven" N and "John buys Maria dinner" D. Assume that Maria's initial credence in D equals some $d \in (0, 1)$ and also that she considers D to be probabilistically independent of the outcome of the toss. The worlds are as follows:

	N	D
w_1	yes	yes
w_2	no	yes
w_3	yes	no
w_4	no	no

and the kernel is as follows:

$$\begin{array}{cccc} d/6 & 5d/6 & (1-d)/6 & 5(1-d)/6 \\ d/6 & 5d/6 & (1-d)/6 & 5(1-d)/6 \\ d/6 & 5d/6 & (1-d)/6 & 5(1-d)/6 \\ d/6 & 5d/6 & (1-d)/6 & 5(1-d)/6 \end{array}$$

with P being any mixture of the rows, ending up as the vector $(d/6, 5d/6, (1-d)/6, 5(1-d)/6)$. From the setup we see immediately that $N = \{w_1, w_3\}$, $D = \{w_1, w_2\}$, $N \leftrightarrow D = \{w_1, w_4\}$, $pr(N, 1/6) = W$, and so $P(N \mid pr(N, 1/6)) = 1/6$, in accordance with ($\mathrm{VI}_w$). But notice that $P(N \mid pr(N, 1/6) \cap N \leftrightarrow D) = d/(5-4d)$. This equals $1/6$ iff $d = .5$. Ergo, $N \leftrightarrow D$ is not a defeater for N and $1/6$ in the exclusive case when Maria's prior credence in D equals .5. ♣

And so the biconditional-admissibility problem remains after the move to HOPs: the additional structure which enables us to rigorously talk about propositions about probabilities of propositions turns out not to be enough to dispel it. In the example above, Maria may be perfectly justified in having initial credence in D other than .5: she might for various reasons consider John buying her dinner to be quite improbable or, say,almost certain. In such cases, $N \leftrightarrow D$ will defeat her direct inference; modelling the situation using HOPs does not help.

A similar generalisation of Example 4 is also, of course, impossible: for it to succeed we need not only assume that Maria's prior credence in D equals .5, but that it does so conditional on any of the two chance hypotheses. We leave the details to the uninterested Reader. Let us, then, move to the core of the biconditional inadmissibility problem: we now know that it persists even after we make the move to a proper, higher-order framework.

Formally, the issue boils down to the following simple Fact and Corollaries[12]. The proof of Fact 2 is in the Appendix. Note that, intentionally, the four results just to be formulated refer to a probability function P and algebra of events $\mathcal{F}$: that is, they apply both to classical probability spaces and to HOPs.

Fact 2. Suppose P is a probability function on an event algebra $\mathcal{F}$. Assume $P(AD) > 0$ and $P(A) < 1$. Then $P(A|A \leftrightarrow D) = P(A)$ if and only if $P(D|A) = P(\neg D|\neg A)$.

Let us now add the assumption that A and D are probabilistically independent, that is, that $P(D|A) = P(D)$. Note that this implies that $\neg D$ and $\neg A$ are also probabilistically independent.

Corollary 1. Suppose P is a probability function on an event algebra $\mathcal{F}$. Assume $P(D|A) = P(D)$, $P(AD) > 0$ and $P(A) < 1$. If $P(A|A \leftrightarrow D) = P(A)$, then $P(D) = {}^1\!/_2$.

Proof. Assume $P(A|A \leftrightarrow D) = P(A)$. Use Fact 2 to infer that (*) $P(D|A) = P(\neg D|\neg A)$. From $P(D|A) = P(D)$ infer that $P(\neg D|\neg A) = P(\neg D)$. Substitute this into (*), establishing that $P(D) = P(\neg D)$. Therefore $P(D) = {}^1\!/_2$. ∎

Now consider assuming that A and D are probabilistically independent given B, that is, that $P(D|AB) = P(D|B)$.

Corollary 2. Assume $P(D|AB) = P(D|B)$, $P(BAD) > 0$ and $P(\neg AB) > 0$. If $P(A|B \cap (A \leftrightarrow D)) = P(A|B)$, then $P(D|B) = {}^1\!/_2$.

Proof. Suppose P is a probability function on an event algebra $\mathcal{F}$. Recall that any B such that $P(B) > 0$ gives rise to a probability function P_B on $\mathcal{F}$, defined as $P_B(\cdot) := P(\cdot|B)$. Then rewrite the proof of Corollary 1 using $P_B(\cdot)$. ∎

The implications of Corollary 2 for defeaters in classical probability spaces and HOPs are as follows (assuming all conditional probabilities are defined):

- in classical probability spaces, if A and D are probabilistically independent given B, then if $A \leftrightarrow D$ is not a defeater for $P(A|B)$, it follows that $P(D|B) = {}^1\!/_2$;
- in HOPs, if A and D are probabilistically independent given $pr(A, \Delta)$, then if $A \leftrightarrow D$ is not a HOP-defeater for $P(A|pr(A, \Delta))$, it follows that $P(D|pr(A, \Delta)) = {}^1\!/_2$.

The second point shows that the issue concerning direct inference discovered by Wallmann & Hawthorne is as problematic in HOPs as it is in classical probability spaces. Suppose we would like to model some agent's direct inference concerning proposition A given the information $pr(A, \Delta)$, that is, specifying the

[12] These are all inspired by Theorem 1 from Wallmann & Hawthorne (2018). On the one hand, as we have already pointed out, we believe that paper describes the issue too informally. On the other, the results are formulated in intricate notation which may obfuscate the message. We believe it will be beneficial to formulate the ideas in as simple a language as possible.

probabiity of A. Consider some proposition D the agent considers to be probabilistically independent from A given $pr(A, \Delta)$. Then if we do not chance upon the very specific case in which $P(D|pr(A, \Delta)) = {}^1\!/_2$, we have on our hands a defeater for A and $pr(A, \Delta)$ in the form of the biconditional $A \leftrightarrow D$: with no seeming rationale for this fact.

One could, of course, consider some alternative higher-order probabilistic approach. Note, however, that the above Fact and Corollaries will hold in any framework containing an algebra of events and a probability measure on it—regardless of any additional structure, useful as it may be for various purposes.

The upshot is the following: one could have hoped that the issues discovered by Wallmann & Hawthorne stemmed from nonrigorous talk about probabilities. We believe we have dispelled that hope here. "Innocent" biconditionals end up as defeaters for direct inference even when talk about higher-order probabilities is properly formally introduced. One could be, we think, excused for seriously considering the option that this points to the possibility that modelling direct inference via conditional probability—in classical probability spaces or not—is, in fact, misguided.

Consider the situation as abstractly as possible. An agent holds some two different prior credences $P(A)$ and $P(D)$. She acquires the information that $A \leftrightarrow D$. How should her credences in A and D change? No simple answer is forthcoming. If both A and D are propositions the agent considers to be somewhat random, but, initially, one to be more likely than the other, it might be rational for the agent to adopt the same credence in A and D which lies somewhere between the two prior credences. But if, as is the case in the examples of direct inferences under discussion, the agent's credence in A results from her obtaining information about the chance of A, while proposition D is about some other agent's choice, it is evident[13] that the agent's credence in A should be fixed, and it is the credence in D that should be modified. The Bayesian vision is, of course, that the probability space modelling the agent's credences will be such that Bayesian conditionalisation on $A \leftrightarrow D$ will deliver correct results in both cases. But some skepticism towards this claim is justifiable: we can see that even imbuing the space with structure specifically tailored to higher-order phenomena is not enough!

5 Conclusion

One could have reasonably suspected that the source of the "admissibility troubles" for Bayesian accounts of direct inference proposed by Wallmann & Hawthorne was the nonrigorous talk about higher-order probabilities employed in the paper (and most of the literature on the topic). The suspicion was rational because, as we have shown in Section 2, in classical probability spaces defeaters abound just due to cardinality-related reasons: if we do not make specific assumptions about the sizes of propositions A and B which "says that the chance of A is x", we should expect most propositions D to be defeaters for the direct inference which would have us put credence x in A given B. This discovery is, in our opinion, shows that discussion of direct inference should move to some formal framework which would rigorously defined what it was to be a proposition "that some proposition has a chance belonging to some interval". We have studied an example of such a framework—the Higher-Order Probability Spaces, introduced in Gaifman (1988). We have, however, eventually shown that the troubles proposed by Wallmann & Hawthorne persist after metaprobabilistic talk is made rigorous, and distilled the simple mathematical essence behind those troubles (Corollary 2). In our opinion this speaks against the possibility of modelling direct inference using conditional probability. That said, we should of course be on the lookout for some different framework, which might be better suited for the job than the one offered by Gaifman.

[13] At least, as evident as things can be in this context.

Appendix

Proof of Claim 1. By uniformity of P the condition $P(B) > 0$ amounts to $B \neq \emptyset$. We determine the number of non-defeaters for A and B. We have two main cases depending on whether or not A and B are disjoint. Recall that $|AB| = k$, $|A'B| = l$.

Case #1: Suppose $A \cap B = \emptyset$. D is a non-defeater (for A and B) if $|AB|/|B| = |ABD|/|BD|$. As the numerators are zero, in order to have this equality we need to make sure $|BD| > 0$ (otherwise the left-hand side is zero, the right-hand side is undefined). Therefore each D which overlaps B (i.e. $B \cap D \neq \emptyset$) is a non-defeater. The number of such events is $(2^{|B|} - 1)2^{|B'|}$, which we can write as $2^{N-l}(2^l - 1)$.

Case #2: Suppose $A \cap B \neq \emptyset$. Let D be an event and put $|ABD| = n$, $|BD| = n + c$. That D is a non-defeater means $|AB|/|B| = |ABD|/|BD|$, i.e. $k/k+l = n/n+c$. Suppose D is a non-defeater. Then $k > 0$ and the previous equality implies $n > 0$. We have two subcases.

Subcase #1: suppose $l = 0$ (that is, $B \subseteq A$). Then D is a non-defeater if and only if $ABD \neq \emptyset$. Thus, D should contain arbitrarily many but at least one element from AB, and any number of elements from $(AB)'$. The number of such D's is $(2^{|AB|} - 1)2^{|(AB)'|} = 2^{N-k}(2^k - 1)$.

Subcase #2: suppose $l > 0$. That D is a non-defeater, in particular the condition $k/k+l = n/n+c$, is equivalent to $0 < n \leq k$ and $c = nl/k \in \mathbb{N}$. Let us write

$$G(k, l) = \{0 < n \leq k : \; nl/k \in \mathbb{N}\}.$$

Now, D is a non-defeater for A and B if and only if for some $n \in G(k, l)$, D contains

- n elements from $A \cap B$;
- nl/k elements from $A' \cap B$;
- arbitrary many elements from B'.

Therefore the number of non-defeaters for A and B in this case is

$$2^{N-(k+l)} \sum_{n \in G(k,l)} \binom{k}{n} \binom{l}{\frac{nl}{k}}.$$

Summing up, given the uniform probability over the sample space, combining the cases above we obtain

$$\mathrm{NonDef}(N, k, l) = \begin{cases} 2^{N-l}(2^l - 1) & \text{if } k = 0, l > 0; \\ 2^{N-k}(2^k - 1) & \text{if } k > 0, l = 0; \\ 2^{N-(k+l)} \sum\limits_{n \in G(k,l)} \binom{k}{n}\binom{l}{\frac{nl}{k}} & \text{if } k > 0, l > 0. \end{cases} \tag{7}$$

∎

Proof of Fact 2.

Assuming $P(AD) > 0$ and $P(A) < 1$, the following list contains exclusively equivalences:

1. $P(A|A \leftrightarrow D) = P(A)$
2. $\frac{P(AD)}{P(AD)+P(\neg A \neg D)} = P(A)$
3. $P(AD) = P(A)P(AD) + P(A)P(\neg A \neg D)$

4. $1 = P(A) + P(A)\frac{P(\neg A \neg D)}{P(AD)}$

5. $P(\neg A) = P(A)\frac{P(\neg A \neg D)}{P(AD)}$

6. $P(\neg A)P(AD) = P(A)P(\neg A \neg D)$

7. $P(D|A) = P(\neg D|\neg A)$.

∎

References

Arló-Costa, H., Hendricks, V. F., Van Benthem, J., Boensvang, H., & Rendsvig, R. K. (2016). *Readings in Formal Epistemology: Sourcebook.* Springer.

Erdös, P. & Shapiro, H. N. (1951). On the changes of sign of a certain error function. *Canadian J. Math*, *3*, 373–385.

Gaifman, H. (1986). A theory of higher order probabilities. In J. Y. Halpern (Ed.), *Theoretical Aspects of Reasoning About Knowledge. Proceedings of the 1986 conference.* (pp. 275–292).

Gaifman, H. (1988). A theory of higher order probabilities. In B. Skyrms & W. Harper (Eds.), *Causation, chance and credence* (pp. 191–219). Springer.

Gyenis, B. & Wroński, L. (2017). Is it the Principal Principle that implies the Principle of Indifference? In G. Hofer-Szabó & L. Wroński (Eds.), *Making it Formally Explicit: Probability, Causality, and Determinism* (pp. 35–41). Springer.

Hall, N. (1994). Correcting The Guide to Objective Chance. *Mind*, *103*(412), 505–518.

Hardy, G. & Wright, E. (2008). *An Introduction to the Theory of Numbers* (6th ed.). Oxford University Press.

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

Högnäs, G. & Mukherjea, A. (2011). *Probability Measures on Semigroups, 2nd ed.* Springer.

Levi, I. (1977). Direct Inference. *The Journal of Philosophy*, *74*(1), 5–29.

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

Mertens, F. (1874). Ueber einige asymptotische Gesetze der Zahlentheorie. *J. Reine Angew. Math.*, *77*, 289–338.

Wallmann, C. & Hawthorne, J. (2018). Admissibility Troubles for Bayesian Direct Inference Principles. *Erkenntnis*. https://doi.org/10.1007/s10670-018-0070-0