

ON THE IMPRECISION OF FULL CONDITIONAL PROBABILITIES

GREGORY WHEELER
FRANKFURT SCHOOL OF FINANCE & MANAGEMENT
g.wheeler@fs.de

FABIO G. COZMAN
UNIVERSITY OF SÃO PAULO
fgcozman@usp.br

ABSTRACT: The purpose of this paper is to show that if one adopts conditional probabilities as the primitive concept of probability, one must deal with the fact that even in very ordinary circumstances at least some probability values may be imprecise, and that some probability questions may fail to have numerically precise answers.

1 INTRODUCTION

De Finetti famously declared “every probability is conditional” (de Finetti 1974, §4.1) and he is not alone in maintaining that Kolmogorov’s axiomatization of probability, with its ratio definition of conditional probability, ought to be replaced by a *full conditional probability* theory that takes conditional probability as primitive.¹ Many proponents of full conditional probabilities also endorse *strict Bayesianism* as well, which is the thesis that imprecision in probability values may only arise from the incomplete elicitation of all quantities necessary to specify a distribution. Strict Bayesianism therefore operates on the assumption that no probabilities are indeterminate in value (Levi 1974), and some go further to maintain that no probabilities should remain imprecise.²

In this paper we point to a difficulty for those who embrace both full conditional probabilities and strict Bayesianism. In very ordinary circumstances, a full conditional probability that appears to be completely specified through the usual methods employed in probability theory will nevertheless leave some probability questions without a precise numerical answer. Even by fully specifying each and every layer of a full conditional probability model one may fail to fully specify a joint probability distribution.

After a brief review of full conditional probability theories and the introduction of our notation in Section 2, we illustrate in Section 3 some surprising ways in which commonplace attempts of specify full conditional probabilities lead to imprecision. In Section 4 we lay bare the problem this imprecision presents to strict Bayesianism by highlighting features of our two examples that weigh against treating them merely as models missing a few (albeit *new*) assessments. We then close in Section 5 with some remarks about two techniques from the theory of imprecise probabilities for dealing with conditioning on zero probability events, as there is a choice in imprecise probability between methods that mimic the standard definition of conditional probability and methods based on full conditional probabilities — reasons to favor full conditional probabilities, we argue, are reasons to favor sets of full conditional measures, too.

2 CONDITIONAL PROBABILITY AS PRIMITIVE: FULL CONDITIONAL PROBABILITIES

In this section we review some background material necessary for our argument. Readers acquainted with Dubins’s concept of full conditional probability may prefer to skim this section or skip to the next.

¹In particular, see the axiomatizations of Renyi (1955) and Popper (1959), and Dubins’s extension of de Finetti (1975). A sample of authors endorsing full conditional probabilities includes representatives from philosophy (van Fraassen 1976; Levi 1980; McGee 1994; Hájek 2003; Sprenger and Hartmann 2017), statistics (Kadane, Schervish, and Seidenfeld 1999), economics and game theory (Blume, Brandenburger, and Dekel 1991b; Myerson 1991; Hammond 1994; Battigalli and Veronesi 1996; Kohlberg and Reny 1997), logic (Adams 1966; Coletti and Scozzafava 2002; Makinson 2011), psychology (Pfeifer and Tulkki 2017), and computer science (Kraus, Lehmann, and Magidor 1990; Cowell, Dawid, Lauritzen, and Spiegelhalter 1999; Gilio 2012).

²A sample of recent objections to imprecise probabilities include (White 2010; Elga 2010; Mahtani 2017). For a survey of responses see (Walley 1991, §§5.6–5.9) and also (Joyce 2011; Pedersen and Wheeler 2014; Chandler 2014).

Kolmogorov’s theory of probability assumes that one has a sample space Ω , a σ -algebra of events \mathcal{A} , and a probability measure \mathbb{P} over those events. To define conditional distributions, Kolmogorov first uses Lebesgue integrals to define expectation, then uses an integral equation based on the Radon-Nikodym theorem (Pollard 2002). Naturally, if $\mathbb{P}(A) > 0$, then $\mathbb{P}(B | A) = \mathbb{P}(A \cap B) / \mathbb{P}(A)$; this is the *ratio definition* of conditional probability (Hájek 2003). Thus, in Kolmogorov’s theory conditional probability is merely defined from unconditional probability. But this account gives short shrift to the observation that every probability is conditional on some background set of facts, which suggests that probability ought to be instead a two-place function with unconditional probability as a special case.

Keynes (1921, Chapter XII) introduced a/h as a two-place relation of probability between a and h , governed, among other things, by the axiom $ab/h = a/bh \cdot b/h$. This approach is also adopted early by de Finetti (1937, 1949), who held that conditional events are primitive objects used to define coherent conditional probabilities. Renyi (1955), seeking to make Kolmogorov’s theory more flexible, published an elegant set of axioms that took conditional probability as primitive, and Popper engineered a system of axioms based on conditional probabilities that allowed *both* the probability of A given B and the probability of B given A to be defined—even if A or B have probability zero. Popper called this property “symmetry” (Popper 1959, Appendix *iv).

Arguments in favor of axiomatizations for full conditional probabilities can be summarized as follows. First, we ought to be able to specify a conditional probability given any possible event, regardless of the probability of this conditioning event, and without necessarily thinking of conditioning as a ratio of probabilities. In short, conditional probability should be bound by the normative constraints imposed over any probability and nothing else. This is the motivation for Popper’s symmetry property.

Second, the Kolmogorovian approach faces technical difficulties when spaces are uncountable, as it cannot directly handle probabilities conditioned on individual elements of the domain and must depend on the machinery of σ -algebras to work. Furthermore, solutions to Kolmogorov’s integral equation may fail to behave as probabilities (Seidenfeld, Schervish, and Kadane 2001; Hájek 2003). Next, Kolmogorov’s approach appears unable to handle some problems involving relative frequencies and proportions (van Fraassen 1976; Easwaran 2011), while economists have complained about difficulties in handling games involving some probability-zero states that one must nevertheless condition on and reason through (Hammond 1994; Battigalli and Veronesi 1996; Kohlberg and Reny 1997).

Finally, axiomatizations based on full conditional probabilities are very elegant. Sometimes one finds the notion that “conditional probabilities are primitive” slipping into textbooks as diverse as game theory (Myerson 1991) and Bayesian networks (Cowell, Dawid, Lauritzen, and Spiegelhalter 1999) simply for the sake of beauty.

There are several axiomatizations of conditional probability. We use here *full conditional probabilities* as defined by Dubins (1975); these are the *coherent probabilities* advocated by de Finetti (de Finetti 1974) once an algebra of events is fixed. The points we discuss in this paper could equally well be raised, after setting up appropriate translations, with Renyi’s *conditional probability spaces* (Renyi 1955) or with *Popper functions* (Popper 1959). However, full conditional probabilities are better behaved and easier to manipulate than those alternatives.

Consider a Boolean algebra \mathcal{B} and a two-place set-function $\mathbb{P} : \mathcal{B} \times (\mathcal{B} \setminus \emptyset) \rightarrow \mathfrak{R}$. Elements of \mathcal{B} are called events and let every $H \in \mathcal{B}$ be nonempty. Then, \mathbb{P} is a *full conditional probability* just in case the following conditions are satisfied (Dubins 1975):

1. $\mathbb{P}(H | H) = 1$;
2. $\mathbb{P}(G | H) \geq 0$, for every event $G \in \mathcal{B}$;
3. $\mathbb{P}(G_1 \cup G_2 | H) = \mathbb{P}(G_1 | H) + \mathbb{P}(G_2 | H)$, whenever $G_1 \cap G_2 = \emptyset$;
4. $\mathbb{P}(G_1 \cap G_2 | H) = \mathbb{P}(G_1 | G_2 \cap H) \mathbb{P}(G_2 | H)$, whenever $G_2 \cap H \neq \emptyset$.

To simplify notation, we use $\mathbb{P}(G)$ to denote $\mathbb{P}(G | \Omega)$ and refer to it as an unconditional probability. Furthermore, we assume hereafter that Ω is finite and \mathcal{B} is the powerset of Ω . Therefore, no idiosyncrasies of uncountable spaces are used in our arguments.

It is not immediately clear how one could specify a unique full conditional probability by assessing various conditional probability values. In the next section we discuss a well-known representation for full conditional probabilities in terms of “layers” and associated (standard) probability measures; we suspect that those interested in full conditional probabilities assume that by employing the assessment procedures routinely applied to standard probability measures to such layered probability measures one can fix a unique full conditional probability. Examples in the next section show that this is not the case in general.

3 FULL CONDITIONAL PROBABILITIES AND IMPRECISE PROBABILITY VALUES

Our main point is simple and can be conveyed by a simple example. We also use this first example to review the layer-based representation of full conditional probabilities. A second, more interesting example follows after.

EXAMPLE 1. Consider a fair coin with faces labeled heads and tails and the edge of the coin painted half in amber and half in blue. A normal flip of the coin may yield the outcome heads (H) or the outcome tails (T), each with probability one-half, or the exceedingly improbable outcome where the coin lands on its amber edge (A) or the equally improbably outcome where the coin lands on its blue edge (B). Events A and B will serve as unconditional probability-zero events. The features of this coin are depicted in Figure 1.

The coin is tossed. If the result is H, a second coin is flipped; this coin is identical to the first coin except that only one-third of its edge is painted amber and the remaining two-thirds is painted blue. If instead the first coin yields T, a third coin is tossed, similar to the other two, but with two-thirds of its edge painted amber and one-third painted blue. Finally, if the first coin lands on its edge, a fourth coin is thrown that is exactly like the first coin: half amber, half blue.

We may express full conditional probabilities using *layers* of regular probabilities; because our space Ω is finite, this representation is straightforward. To start, let L_0 denote the event consisting of all elements of Ω that have positive unconditional probability. For the first coin toss, the top layer L_0 assigns positive probability to the typical set of outcomes, heads and tails: $L_0 = \{H, T\}$.

Next denote by L_1 the set of elements of $\Omega \setminus L_0$ that have positive conditional probability given $\Omega \setminus L_0$. Here, $L_1 = \{A, B\}$. In general we say that L_i (for $i > 0$) is the set of elements of $\Omega \setminus \cup_{j=0}^{i-1} L_j$ that have positive probability conditional on $\Omega \setminus \cup_{j=0}^{i-1} L_j$.

Any full conditional probability \mathbb{P} can be represented by a sequence of probability measures, $\mathbb{P} = \langle \mathbb{P}_i \rangle$, where \mathbb{P}_i is strictly positive over layer L_i (Spohn 1986). For any nonempty event G , we then have $\mathbb{P}(G | H) = \mathbb{P}(G | H \cap L_H)$, where L_H is the first layer such that $\mathbb{P}(H | L_H) > 0$ (Battigalli and Veronesi 1996, Lemma 2.1a). The *layer number* of event H with respect to full conditional probability \mathbb{P} is the smallest index i such that H intersects layer L_i , in symbols $\circ(H) = i$, which is ∞ by

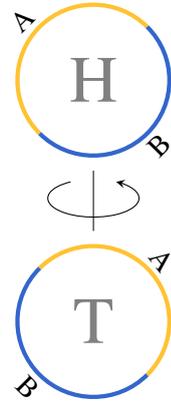


Figure 1: The first coin.

Table for L_0 →

	H	T	A	B
	1/2	1/2	0	0

Table for L_1 →

	H	T	A	B
	0	0	1/2	1/2

	H	T	A	B
	1/2	1/2	1/2	1/2

Figure 2: Two equivalent representations of the full conditional probability assignments for the first coin in Example 1. Top: two probability tables, one per layer. Bottom: one table, with layers distinguished by colors.

convention if H is empty (Coletti and Scozzafava 2002).

With this machinery on board, we can return to our discussion of coin tossing. How can we translate the assessments in the description of the problem to obtain a full conditional probability? For the first coin we have $\mathbb{P}(H) = 1/2 = \mathbb{P}(T)$ and $\mathbb{P}(A) = 0 = \mathbb{P}(B)$, but $\mathbb{P}(A | A \cup B) = 1/2 = \mathbb{P}(B | A \cup B)$ and $\mathbb{P}(H | A \cup B) = 0 = \mathbb{P}(T | A \cup B)$, which defines a (unique) full conditional probability over $\{H, T, A, B\}$. Figure 2 presents two equivalent ways to visualize this representation of the probability assessments for the first coin toss.

Because the same four possible outcomes may occur on the second toss, sixteen additional conditional probabilities are seemingly necessary to model the outcomes of the second toss given the first toss: in doing so we mimic the customary scheme where a joint distribution is built from its “prior” and its “conditional” pieces. We use the subscript 1 to specify the outcome of the first toss and the subscript 2 to indicate the result of the second. So, for example, $\mathbb{P}(H_2 | H_1) = 1/2$ expresses that the probability of the second coin toss landing heads given the first lands heads is one-half, and $\mathbb{P}(A_2 | A_2 \cup B_2, H_1) = 1/3$ expresses that the probability of the second coin toss landing on its amber edge given that the second toss lands on its edge and the first toss lands heads is one-third. With this convention, the following values specify all 16 possible outcomes of this experiment:

$$\mathbb{P}(H_1 | \Omega) = \mathbb{P}(T_1 | \Omega) = \mathbb{P}(A_1 | A_1 \cup B_1) = \mathbb{P}(B_1 | A_1 \cup B_1) = 1/2$$

$$\begin{aligned} \mathbb{P}(H_2 | H_1) &= \mathbb{P}(T_2 | H_1) = 1/2 & \mathbb{P}(A_2 | A_2 \cup B_2, H_1) &= 1/3 & \mathbb{P}(B_2 | A_2 \cup B_2, H_1) &= 2/3 \\ \mathbb{P}(H_2 | T_1) &= \mathbb{P}(T_2 | T_1) = 1/2 & \mathbb{P}(A_2 | A_2 \cup B_2, T_1) &= 2/3 & \mathbb{P}(B_2 | A_2 \cup B_2, T_1) &= 1/3 \\ \mathbb{P}(H_2 | A_1) &= \mathbb{P}(T_2 | A_1) = 1/2 & \mathbb{P}(A_2 | A_2 \cup B_2, A_1) &= 1/2 & \mathbb{P}(B_2 | A_2 \cup B_2, A_1) &= 1/2 \\ \mathbb{P}(H_2 | B_1) &= \mathbb{P}(T_2 | B_1) = 1/2 & \mathbb{P}(A_2 | A_2 \cup B_2, B_1) &= 1/2 & \mathbb{P}(B_2 | A_2 \cup B_2, B_1) &= 1/2. \end{aligned}$$

Now a question: *What is the probability that the first toss yields heads and the second toss yields blue, given that exactly one of the two tosses landed on its edge?* Because it appears all relevant marginal and conditional probabilities are specified, one might think that the answer to this question ought to be a precise number. Yet, $\mathbb{P}(H_1 \cap B_2 | C) \in [0, 2/3]$ is the answer, where C is the event that either the first coin toss landed on its edge or the second coin toss landed on its edge but not both. \square

To see that this result is correct, consider two possible full conditional probabilities that satisfy all assessments in Example 1, shown in Tables 3a and 3b. Table 3a specifies a full conditional probability that yields $\mathbb{P}(H_1 \cap B_2 | C) = 0$, whereas 3b specifies a full conditional probability which yields $\mathbb{P}(H_1 \cap B_2 | C) = 2/3$. In both tables the first layer such that $\mathbb{P}(C | L_H) > 0$ is L_1 , but each has a different sequence of possible outcomes for the pair of coin flips. To see that $\mathbb{P}(H_1 \cap B_2 | C) = 0$ according to Table 3a, observe that L_1 specifies probabilities only for the first toss landing on edge and the second

	H_2	T_2	A_2	B_2	Layers of (a)		
(a)	H_1	1/4	1/4	1/6	1/3	0	2
	T_1	1/4	1/4	1/3	1/6		
	A_1	1/4	1/4	1/4	1/4	1	3
	B_1	1/4	1/4	1/4	1/4		
	H_2	T_2	A_2	B_2	Layers of (b)		
(b)	H_1	1/4	1/4	1/3	2/3	0	1
	T_1	1/4	1/4	2/3	1/3		2
	A_1	1/4	1/4	1/4	1/4	3	4
	B_1	1/4	1/4	1/4	1/4		
	H_2	T_2	A_2	B_2			
(c)	H_1	1/4	1/4	$\alpha/6$	$\alpha/3$		
	T_1	1/4	1/4	$\alpha/3$	$\alpha/6$		
	A_1	$(1-\alpha)/4$	$(1-\alpha)/4$	1/4	1/4		
	B_1	$(1-\alpha)/4$	$(1-\alpha)/4$	1/4	1/4		

Table 3: Two different conditional probabilities, (a) and (b), that satisfy all assessments in Example 1, and a family of full conditional probabilities (c) that satisfy all assessments in Example 1, for $\alpha \in [0, 1)$. Cells that refer to layer 0 have a white background; layers 1, 2, 3, and 4 have gradually darker backgrounds.

landing on a face, and assigns zero to the remaining possible pairs, including the possibility that the first toss lands heads and the second coin toss lands on its blue edge. By contrast, $\mathbb{P}(H_1 \cap B_2 | C) = 2/3$ according to Table 3b, since L_1 in this probability table specifies probabilities for the first toss landing on a face and the second landing on an edge. However, notice that Table 3b does not merely permute the order of layers. For instance, according to Table 3b, $\mathbb{P}(T_1, A_2 \cup B_2 | C) = 0$. Finally, Table 3c shows that we do not even need four layers to satisfy all assessments: a three layer model will do, yielding $\mathbb{P}(H_1 \cap B_2 | C) = \frac{2\alpha}{3}$ for $\alpha \in [0, 1]$.

Example 1 reveals there are many full conditional probabilities that satisfy the apparently complete set of assessments for Example 1. Put differently, although assigning regular, non-zero values to these sixteen outcomes would yield a complete model in which $\mathbb{P}(T_1, A_2 \cup B_2 | C)$ has a precise numerical answer, numerical determinacy is lost in Example 1 by the introduction of zero-probability events and switch to full conditional probabilities. Although it is possible to fix a determinate probability value for $\mathbb{P}(T_1, A_2 \cup B_2 | C)$ by requiring additional assessments, maintaining that one can and should always do so comes at a cost. First, requiring additional assessments sacrifices the commonplace notion of what a complete probability model specification amounts to. For these new assessments concern quantities that are calculable consequences of a standard probability model, when defined, not premises we are expected to assign a value to. Second, these new assessments can be saddled with surprisingly strong constraints—a point that will come out in our next example.

So far we have worked only with “marginal” and conditional probabilities, bearing in mind our remarks at the end of Section 2 about our use of this terminology since here all probabilities are conditional. In addition to probability assessments one may also have structural independence relations, which raises two questions. First, would the inclusion of explicit independence conditions ward off imprecision in full conditional probabilities? The reason to ask is that it is known that a similar phenomenon, dilating sets of probabilities (Seidenfeld and Wasserman 1993), is due to possible deviations from independence that imprecision does not explicitly rule out (Pedersen and Wheeler 2015; Pedersen and Wheeler 2019) and Example 1 involves dependent outcomes. The short answer is, *no*, but this is a complicated issue. For there are several different concepts of independence for full conditional probabilities. Even so, it may come as a surprise that no known independence concept can ensure a unique joint conditional probability.³ This raises a second question: Does adding independence conditions to the mix expose a new problem for strict Bayesianism that we haven’t seen already? The answer is, *yes*. Demanding additional assessments to secure precision is not always reasonable. Let’s see why.

For our next example we consider a proposal from the philosophy of science to model the successful reduction of one scientific theory to another, such as the reduction of chemistry to atomic physics.

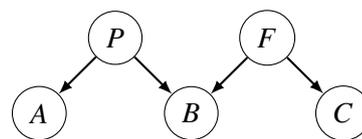


Figure 4: The Bayesian network for the GNS theory reduction model.

$\mathbb{P}(P)$	$\mathbb{P}(F)$
$\mathbb{P}(A P)$	$\mathbb{P}(C F)$
$\mathbb{P}(A \neg P)$	$\mathbb{P}(C \neg F)$
$\mathbb{P}(B P, F)$	$\mathbb{P}(B P, \neg F)$
$\mathbb{P}(B \neg P, F)$	$\mathbb{P}(B \neg P, \neg F)$

Table 5: The conditional probability table for the Bayesian network in Figure 4.

$\mathbb{P}(P) = \frac{1}{2}$	$\mathbb{P}(F) = \frac{1}{2}$
$\mathbb{P}(A P) = 1$	$\mathbb{P}(C F) = 1$
$\mathbb{P}(A \neg P) = 0$	$\mathbb{P}(C \neg F) = 0$
$\mathbb{P}(B P, F) = 1$	$\mathbb{P}(B P, \neg F) = 0$
$\mathbb{P}(B \neg P, F) = 0$	$\mathbb{P}(B \neg P, \neg F) = 0$

Table 6: Values for the supremely confident experts in Example 2.

³This type of difficulty has been noted earlier by Kohlberg and Reny in connection with their definition of *strong independence* (Kohlberg and Reny 1997, Remark 7) and by Cozman with respect to the specification of Bayesian networks (Cozman 2013). Our aim here is to show that imprecision in probability values is a fundamental feature of full conditional probability, not a narrow technical issue that appears in some accounts of full conditional probability but not others, nor a property that appears only under some notions of independence but not others.

The Generalized Nagel-Schaffner (GNS) model of theory reduction (Dizadji-Bahmani, Frigg, and Hartmann 2011) aims to connect theory reduction to theory confirmation through a particular Bayesian network, displayed in Figure 4. The details of the GNS model will not concern us except for one assumption, which is that all probabilities in the model “lie in the open interval $(0, 1)$ ” in order to “avoid technical complications” (2011, p. 326). Our next example can be viewed as an explanation of the technical complications that arise when probability values are extreme and couched within full conditional probabilities. This is a particularly interesting point since elsewhere Hartmann has forcefully argued for conditional probability as a primitive concept in such settings (Sprengrer and Hartmann 2017).

EXAMPLE 2. The GNS model (Dizadji-Bahmani, Frigg, and Hartmann 2011) has five propositional variables, P, F, A, B, C . A variable V can take two values, written V and $\neg V$, expressing that the variable V is true or false, respectively. Proposition P is an empirical proposition of a phenomenological theory \mathbf{T}_P and proposition F is an empirical proposition of a fundamental theory \mathbf{T}_F . Suppose A confirms P and C confirms F , while B confirms both P and F . This situation is modeled by the Bayesian network in Figure 4,⁴ and Table 5 gives a complete list of the probability values required to parameterize the graphical model in Figure 4.

Dizadji-Bahmani, Frigg, and Hartmann tinker with the probability values in Table 5, exploring their effect on the reduction of \mathbf{T}_P to \mathbf{T}_F . As noted already, they only consider non-extreme probability values. But now imagine we consult a group of supremely confident experts on the probabilities attached to A, B , and C . One expert, responsible for A , tells us that A reproduces P , $\mathbb{P}(A | P) = 1$ and $\mathbb{P}(A | \neg P) = 0$, and another expert, responsible for C , reports that $\mathbb{P}(C | F) = 1$ and $\mathbb{P}(C | \neg F) = 0$. Finally, a third expert responsible for B tells us that B reproduces F and P together, that is $\mathbb{P}(B | F, P) = 1$ and, $\mathbb{P}(B | F, \neg P) = \mathbb{P}(B | \neg F, P) = \mathbb{P}(B | \neg F, \neg P) = 0$. To complete the example, suppose for simplicity that $\mathbb{P}(F) = \mathbb{P}(P) = 1/2$. Table 6 summarizes this assessment.

⁴Briefly, a Bayesian network is a pair consisting of a directed acyclic graph and a probability distribution (Pearl 1988). The graph consists of nodes and edges, and each node is a random variable. The graph and the distribution are related by the following Markov condition: a random variable V is independent of its nondescendants given its parents (a parent of V is a node U such that there is an edge from U to V ; a descendant of V is a node U such that there is a directed path from U to V). Consequently, the joint distribution over all random variables factorizes into *local conditional distributions*: each node/variable V is associated with the probability values $\mathbb{P}(V = v | \text{pa}(V) = \pi) = \pi$, for each value v of V and each value π of $\text{pa}(V)$, the parents of V .

$(P \perp\!\!\!\perp F, C)$
$(F \perp\!\!\!\perp P, A)$
$(A \perp\!\!\!\perp B, C, F P)$
$(B \perp\!\!\!\perp A, C F, P)$
$(C \perp\!\!\!\perp A, B, P F)$

Table 7: Independence conditions encoded by the Bayesian network in Figure 4.

	\bar{P}, \bar{F}	\bar{P}, F	P, \bar{F}	P, F
$\bar{A}, \bar{B}, \bar{C}$	0	1	4	7
\bar{A}, \bar{B}, C	1	0	5	6
\bar{A}, B, \bar{C}	2	3	6	5
\bar{A}, B, C	3	2	7	4
A, \bar{B}, \bar{C}	4	5	0	3
A, \bar{B}, C	5	4	1	2
A, B, \bar{C}	6	7	2	1
A, B, C	7	6	3	0

Table 8: A full conditional probability that satisfies all assessments and assertions in Example 2. Each row refers to a configuration of truth assignments of (A, B, C) , and the cell number refers to the layer number of each configuration of (P, F, A, B, C) . For each variable V , \bar{V} codes $\neg V$.

	\bar{F}	F	\bar{B}	B	
\bar{P}, \bar{A}	0	0	\bar{A}, \bar{C}	0	2
\bar{P}, A	1	1	\bar{A}, C	1	3
P, \bar{A}	1	1	A, \bar{C}	4	6
P, A	0	0	A, C	5	7

Table 9: Left: The full conditional probability for (P, F, A) , obtained by marginalization of the full conditional probability in Table 8. Right: The full conditional probability for (A, B, C) given $\{\neg P, \neg F\}$, obtained by conditioning of the full conditional probability specified in Table 8. For each table, cell numbers refer to the layer number of the configuration of that table’s variables.

Now suppose we observe A and $\neg B$ and C , and we are interested in the probability of P given this observation. If all we have is the Bayesian network in Figure 4 and the probability values of Table 6, Kolmogorovian-style conditioning leaves $\mathbb{P}(P \mid A, \neg B, C)$ undefined.

However, suppose we are committed to full conditional probabilities and are interested in the “joint” full conditional probability that complies with our supremely confident experts and satisfies the independence conditions of the GNS model—that is, we are interested in a full conditional probability that satisfies the values in Table 6 and the independence conditions in Table 7, where $(X \perp\!\!\!\perp Y \mid Z)$ expresses that X and Y are independent given Z .

Turn again to $\mathbb{P}(P \mid A, \neg B, C)$. Is there a *unique* full conditional probability that satisfies these assessments and independence conditions? For all existing concepts of independence in the literature, the answer is *no*. In fact, $\mathbb{P}(P \mid A, \neg B, C) \in [0, 1]$. Thus, the “technical complication” that is avoided by the GNS model restriction to non-extreme probabilities is that otherwise the value of $\mathbb{P}(P \mid A, \neg B, C)$ is maximally imprecise. \square

There are many different independence concepts on offer, but there are four basic types of independence, we maintain, that canvas the options available in the literature. As we will see over the following subsections, for all four types, full conditional probabilities plus independence do not fix a unique conditional probability in Example 2.

3.1 EPISTEMIC INDEPENDENCE

Suppose that we take $(X \perp\!\!\!\perp Y \mid Z)$ to mean that, for any possible values x, y and z , both

$$\mathbb{P}(X = x \mid Y = y, Z = z) = \mathbb{P}(X = x \mid Z = z)$$

whenever $\{Y = y, Z = z\} \neq \emptyset$;

and

$$\mathbb{P}(Y = y \mid X = x, Z = z) = \mathbb{P}(Y = y \mid Z = z)$$

whenever $\{X = x, Z = z\} \neq \emptyset$.

We then say that X and Y are *epistemically independent* given Z (Walley 1991). Informally, X is epistemically independent of Y just when, conditional on Z , X is irrelevant to Y and Y is irrelevant to X .

Now consider the full conditional probability depicted in Table 8. In contrast to previous tables, the cells of this table contain only the layer numbers for all possible configurations of (P, F, A, B, C) . For instance, the layer number of the truth-value assignment $\{P, F, \neg A, \neg B, C\}$ is 6. Each layer receives a uniform distribution; for instance, we have $\mathbb{P}(\neg P, \neg F, \neg A, \neg B, \neg C) = 1/4$. It can be verified that this full conditional probability satisfies all assessments and assertions in Example 2. To illustrate the necessary calculations, consider showing that P and (F, C) are epistemically independent. Table 9 (left) shows the full conditional probability obtained by marginalizing the joint full conditional probability. Again, only layer numbers are presented and each layer is given a uniform distribution.

	\bar{P}, \bar{F}	\bar{P}, F	P, \bar{F}	P, F
$\bar{A}, \bar{B}, \bar{C}$	0	2	1	7
\bar{A}, \bar{B}, C	2	0	3	5
\bar{A}, B, \bar{C}	4	6	5	3
\bar{A}, B, C	6	4	7	1
A, \bar{B}, \bar{C}	1	3	0	6
A, \bar{B}, C	3	1	2	4
A, B, \bar{C}	5	7	4	2
A, B, C	7	5	6	0

	\bar{P}, \bar{F}	\bar{P}, F	P, \bar{F}	P, F
$\bar{A}, \bar{B}, \bar{C}$	0	1	1	4
\bar{A}, \bar{B}, C	1	0	2	3
\bar{A}, B, \bar{C}	2	3	3	2
\bar{A}, B, C	3	2	4	1
A, \bar{B}, \bar{C}	1	2	0	3
A, \bar{B}, C	2	1	1	2
A, B, \bar{C}	3	4	2	1
A, B, C	4	3	3	0

Table 10: Two additional full conditional probabilities that satisfies all assessments and assertions in Example 2. Both tables present only layer numbers. For example, the first layer in the top table that has a positive probability distribution for the event $\neg A, B, C, P, \neg F$ is layer 7, whereas in the bottom table it is layer 4.

The epistemic independence of P and (F, C) can be easily verified. Similarly, Table 9 (right) can be used to verify that B and (A, C) are epistemically independent given $\{\neg P, \neg F\}$.

Table 10 shows two additional full conditional probabilities that also satisfy the probability assessments in Table 6 and the epistemic independence assertions in Table 7. In both cases assume that the probability measure associated with each layer is a uniform distribution. Although many other full conditional probabilities could be constructed, the point to notice is that $\mathbb{P}(P | A, \neg B, C) = 1$ with respect to the full conditional probability in Table 8, and $\mathbb{P}(P | A, \neg B, C) = 0$ with respect to the full conditional probability in Table 10 (top). These two full conditional probabilities are maximally disagreeing on the probability of interest. The full conditional probability on Table 10 (bottom), on the other hand, does not produce an extreme value for $\mathbb{P}(P | A, \neg B, C)$; rather, we obtain $\mathbb{P}(P | A, \neg B, C) = 1/2$.

Therefore, if we have assessments as in Table 6 and assertions of *epistemic* independence as in Table 7, the “technical complication” the GNS model faces is that $\mathbb{P}(P | A, \neg B, C)$ is maximally imprecise.

3.2 HAMMOND INDEPENDENCE

Perhaps the problem is that epistemic independence is too weak. To see why, return to the first coin in Example 1 and imagine two epistemically independent tosses of that coin. A possible full conditional probability is shown in Table 11: even though the first three layers seem perfectly appropriate, the last layer does not impose any sort of independence between the tosses.

	H ₂	T ₂	A ₂	B ₂
H ₁	1/4	1/4	1/4	1/4
T ₁	1/4	1/4	1/4	1/4
A ₁	1/4	1/4	1/9	2/9
B ₁	1/4	1/4	5/9	1/9

Table 11: A full conditional probability for two epistemically independent tosses of the first coin from Example 1.

To address pathological cases like the one depicted in Table 11, Hammond proposed a stronger concept of independence (1994). His concept has reappeared in other studies (Battigalli and Veronesi 1996; Blume, Brandenburger, and Dekel 1991a; Swinkels 1993), and the corresponding concept of conditional independence has been studied by Cozman and Seidenfeld (2009). Hammond’s concept of independence amounts to requiring equality amongst full conditional probabilities rather than amongst probability values. We say that X and Y are *h-independent* given Z iff

- for any possible x , any possible y , any possible z , and any possible event \mathcal{X} consisting of values of X ,

$$\mathbb{P}(X = x | \mathcal{X}, Y = y, Z = z) = \mathbb{P}(X = x | \mathcal{X}, Z = z) \text{ whenever } \{\mathcal{X}, Y = y, Z = z\} \neq \emptyset,$$

- for any possible x , any possible y , any possible z , and any possible event \mathcal{Y} consisting of values of Y ,

$$\mathbb{P}(Y = y | \mathcal{Y}, X = x, Z = z) = \mathbb{P}(Y = y | \mathcal{Y}, Z = z) \text{ whenever } \{\mathcal{Y}, X = x, Z = z\} \neq \emptyset.$$

Unlike epistemic independence, the full conditional probability of the two coin tosses in Table 11 does not satisfy h-independence. Note the effect of the additional conditioning events in the definition, as compared to epistemic independence: \mathcal{X} lets us “look into” each layer of the distribution of X , and \mathcal{Y} lets us “look into” each layer of the distribution of Y . Thus in Table 11 we see that h-independence fails when we choose \mathcal{X} to refer to layer $\{A_1 \cup B_1\}$.

However, the three full conditional probabilities in Tables 8 and 10 *do* satisfy all probability assertions of our supremely confident experts in Table 6 and assertions of h-independence in Table 7. So, h-independence is not strong enough to guarantee a unique full conditional probability. Still we obtain $\mathbb{P}(P | A, \neg B, C) \in [0, 1]$.

3.3 LAYER INDEPENDENCE

Coletti and Scozzafava have argued that concepts of independence must pay attention to the “depth” of various layers in the underlying full conditional probability (Coletti and Scozzafava 2002). Vantaggi has examined this proposal in detail (Vantaggi 2001), and has introduced concepts of independence that take into account the layer number of each event.

Recall that $\circ(V) = i$ denotes the smallest index i such that V intersects layer L_i . One condition introduced by Vantaggi is that

$$\circ(X = x | Y = y, Z = z) = \circ(X = x | Y \neq y, Z = z) \text{ whenever } \{Y = y, Z = z\} \neq \emptyset \neq \{Y \neq y, Z = z\}$$

and

$$\circ(X \neq x | Y = y, Z = z) = \circ(X \neq x | Y \neq y, Z = z) \text{ whenever } \{Y = y, Z = z\} \neq \emptyset \neq \{Y \neq y, Z = z\},$$

where $\circ(H | G)$ is defined by $\circ(G \cap H) - \circ(G)$, for non-empty event G (Coletti and Scozzafava 2002). Vantaggi takes this condition to be necessary for X to be independent of Y given Z (Vantaggi 2001, Definition 9.3). Note that there is an implicit universal quantification over possible x , y , and z in these equalities, which run over all possible values of the random variables. Obviously, each equality requires the computation of various layer numbers with respect to the same full conditional probability, even though the notation does not insist on indicating this full conditional probability explicitly. A similar condition, referred to as the *conditional layer condition* by Cozman and Seidenfeld (2009), is that, again with implicit universal quantification over possible x , y , z : $\circ(X = x, Y = y | Z = z) = \circ(X = x | Z = z) + \circ(Y = y | Z = z)$, whenever $\{Z = z\} \neq \emptyset$. Table 12 gives an example of a full conditional probability for two binary variables, X and Y , that satisfies the conditional layer condition.

However, neither Vantaggi’s condition nor the conditional layer condition helps to rule out any of the three full conditional probabilities in Tables 8 and 10 demonstrating that $\mathbb{P}(P | A, \neg B, C) \in [0, 1]$. In fact, for any independence assertion of the form $(X \perp\!\!\!\perp Y | Z)$ in Table 7, these full conditional probabilities satisfy Vantaggi’s condition, as expressed above, and additionally in reverse form:

$$\circ(Y = y | X = x, Z = z) = \circ(Y = y | X \neq x, Z = z) \text{ whenever } \{X = x, Z = z\} \neq \emptyset \neq \{X \neq x, Z = z\}$$

and

	\bar{Y}	Y
\bar{X}	0	1
X	1	2

$$\begin{aligned} \circ(X = 0, Y = 0) &= 0 = 0 + 0 = \circ(X = 0) + \circ(Y = 0) \\ \circ(X = 0, Y = 1) &= 1 = 0 + 1 = \circ(X = 0) + \circ(Y = 1) \\ \circ(X = 1, Y = 0) &= 1 = 1 + 0 = \circ(X = 1) + \circ(Y = 0) \\ \circ(X = 1, Y = 1) &= 2 = 1 + 1 = \circ(X = 1) + \circ(Y = 1) \end{aligned}$$

Table 12: A full conditional probability distribution for two variables, X and Y (with $Z = \Omega$), satisfying the conditional layer condition.

$\mathbb{P}(P) = \frac{1}{2}$	$\mathbb{P}(F) = \frac{1}{2}$
$\mathbb{P}(A P) = 1 - \varepsilon^4$	$\mathbb{P}(C F) = 1 - \varepsilon$
$\mathbb{P}(A \neg P) = \varepsilon^4$	$\mathbb{P}(C \neg F) = \varepsilon$
$\mathbb{P}(B P, F) = 1 - \varepsilon^2$	$\mathbb{P}(B P, \neg F) = \varepsilon^2$
$\mathbb{P}(B \neg P, F) = \varepsilon^2$	$\mathbb{P}(B \neg P, \neg F) = \varepsilon^2$
$\mathbb{P}(P) = \frac{1}{2}$	$\mathbb{P}(F) = \frac{1}{2}$
$\mathbb{P}(A P) = 1 - \varepsilon$	$\mathbb{P}(C F) = 1 - \varepsilon^2$
$\mathbb{P}(A \neg P) = \varepsilon$	$\mathbb{P}(C \neg F) = \varepsilon^2$
$\mathbb{P}(B P, F) = 1 - \varepsilon^4$	$\mathbb{P}(B P, \neg F) = \varepsilon^4$
$\mathbb{P}(B \neg P, F) = \varepsilon^4$	$\mathbb{P}(B \neg P, \neg F) = \varepsilon^4$
$\mathbb{P}(P) = \frac{1}{2}$	$\mathbb{P}(F) = \frac{1}{2}$
$\mathbb{P}(A P) = 1 - \varepsilon$	$\mathbb{P}(C F) = 1 - \varepsilon$
$\mathbb{P}(A \neg P) = \varepsilon$	$\mathbb{P}(C \neg F) = \varepsilon$
$\mathbb{P}(B P, F) = 1 - \varepsilon^2$	$\mathbb{P}(B P, \neg F) = \varepsilon^2$
$\mathbb{P}(B \neg P, F) = \varepsilon^2$	$\mathbb{P}(B \neg P, \neg F) = \varepsilon^2$

Table 13: Alternative probability tables that induce imprecision under kr-independence.

$\circ(Y \neq y | X = x, Z = z) = \circ(Y \neq y | X \neq x, Z = z)$ whenever $\{X = x, Z = z\} \neq \emptyset \neq \{X \neq x, Z = z\}$.

Another concept of independence that employs layer numbers is *layer independence* (Cozman 2013): X and Y are layer independent given Z when the conditional layer condition is satisfied and moreover, for each layer L_i ,

$$\mathbb{P}(X = x, Y = y | L_i^z) = \mathbb{P}(X = x | L_i^z) \mathbb{P}(Y = y | L_i^z)$$

whenever $L_i^z \neq \emptyset$, where L_i^z denotes the event $\{Z = z\} \cap L_i$. The full conditional probability in Table 8 satisfies all independence assertions in Table 7 when these assertions are interpreted as layer independence. Similarly, the full conditional probability in Table 10 (top) satisfies the same layer independence assertions. But we are finally able to eliminate the full conditional probability in Table 10 (bottom), as this full conditional probability fails several layer independence assertions. For instance, $\mathbb{P}(A, \neg B, C | \{\neg P, \neg F\} \cap L_2) = 1/2$, but $\mathbb{P}(A, C | \{\neg P, \neg F\} \cap L_2) \mathbb{P}(\neg B | \{\neg P, \neg F\} \cap L_2) = 1/4$.

In any case, even if layer independence may succeed in eliminating this full conditional probability, it still leaves $\mathbb{P}(P | A, \neg B, C)$ without a unique value.

3.4 KOHLBERG-RENY INDEPENDENCE

We now consider a concept of independence that is rather different from the previous notions. This independence concept was proposed by Kohlberg and Reny in the context of relative probabilities, but it can be adapted to our context as follows. First, consider a full conditional probability \mathbb{P} over sample space Ω , and say that a sequence $\langle \mathbb{P}_n \rangle$ of positive standard probability measures over Ω *approaches* \mathbb{P} iff whenever $\mathbb{P}(H | G \cup H) > 0$, then $\mathbb{P}(G | G \cup H) / \mathbb{P}(H | G \cup H) = \lim_{n \rightarrow \infty} \mathbb{P}_n(G) / \mathbb{P}_n(H)$, and whenever $\mathbb{P}(H | G \cup H) = 0$ and H is nonempty, then $\lim_{n \rightarrow \infty} \mathbb{P}_n(G) / \mathbb{P}_n(H) = \infty$. Then Kohlberg and Reny define X and Y to be (strongly) independent iff there is $\langle \mathbb{P}_n \rangle$ approaching \mathbb{P} such that X and Y are stochastically independent with respect to each \mathbb{P}_n . The extension of this definition to conditional independence is natural (Cozman 2013): Say that X and Y are *kr-independent* given Z iff there is $\langle \mathbb{P}_n \rangle$ approaching \mathbb{P} such that X and Y are stochastically independent given Z with respect to each \mathbb{P}_n . Note that this definition is a little loose when it comes to dealing with several independence assertions simultaneously, because we must only find an approaching sequence separately for each assertion, and this may seem rather weak. In any case, in our present example we can find approaching sequences for each one of the three full conditional probabilities we have been contemplating, such that each one of these sequences *simultaneously* satisfies *all* independence assertions in Table 7.

Consider first the full conditional probability in Table 9. We build an approaching sequence for it by taking the graph in Figure 4 and the assessments in Table 13 (top). Note that these assessments are parameterized by some $\varepsilon \in (0, 1)$. The joint distribution is presented in Table 14. Now take $\varepsilon_n = 1/(2n)$; we obtain a sequence of Bayesian networks whose associated joint probability distributions approach the full conditional probability in Table 8. Similarly, we create an approaching sequence for the full conditional probability in Table 10 (top) by adopting the values listed in Table 13 (middle). Finally, we create an approaching sequence for the full conditional probability in Table 10 (bottom) by adopting the values in Table 13 (bottom).

Consequently, all three full conditional probabilities in Tables 8 and 10 satisfy kr-independence assertions in Table 7. In fact it should be clear that by changing exponents of ε in these approaching sequences, we can build many other full conditional probabilities that satisfy all assertions. In particular, by choosing every exponent of ε to be just 1, we can produce a full conditional probability with just four layers.

	\bar{P}, \bar{F}	\bar{P}, F	P, \bar{F}	P, F
$\bar{A}, \bar{B}, \bar{C}$	$\frac{(1-\varepsilon)(1-\varepsilon^2)(1-\varepsilon^4)}{4}$	$\frac{\varepsilon(1-\varepsilon^2)(1-\varepsilon^4)}{4}$	$\frac{(1-\varepsilon)(1-\varepsilon^2)\varepsilon^4}{4}$	$\frac{\varepsilon^7}{4}$
\bar{A}, \bar{B}, C	$\frac{\varepsilon(1-\varepsilon^2)(1-\varepsilon^4)}{4}$	$\frac{(1-\varepsilon)(1-\varepsilon^2)(1-\varepsilon^4)}{4}$	$\frac{\varepsilon^5(1-\varepsilon^2)}{4}$	$\frac{\varepsilon^6(1-\varepsilon)}{4}$
\bar{A}, B, \bar{C}	$\frac{(1-\varepsilon)\varepsilon^2(1-\varepsilon^4)}{4}$	$\frac{\varepsilon^3(1-\varepsilon^4)}{4}$	$\frac{\varepsilon^6(1-\varepsilon)}{4}$	$\frac{\varepsilon^5(1-\varepsilon^2)}{4}$
\bar{A}, B, C	$\frac{\varepsilon^3(1-\varepsilon^4)}{4}$	$\frac{(1-\varepsilon)\varepsilon^2(1-\varepsilon^4)}{4}$	$\frac{\varepsilon^7}{4}$	$\frac{(1-\varepsilon)(1-\varepsilon^2)\varepsilon^4}{4}$
A, \bar{B}, \bar{C}	$\frac{(1-\varepsilon)(1-\varepsilon^2)\varepsilon^4}{4}$	$\frac{\varepsilon^5(1-\varepsilon^2)}{4}$	$\frac{(1-\varepsilon)(1-\varepsilon^2)(1-\varepsilon^4)}{4}$	$\frac{\varepsilon^3(1-\varepsilon^4)}{4}$
A, \bar{B}, C	$\frac{\varepsilon^5(1-\varepsilon^2)}{4}$	$\frac{(1-\varepsilon)(1-\varepsilon^2)\varepsilon^4}{4}$	$\frac{\varepsilon(1-\varepsilon^2)(1-\varepsilon^4)}{4}$	$\frac{(1-\varepsilon)\varepsilon^2(1-\varepsilon^4)}{4}$
A, B, \bar{C}	$\frac{\varepsilon^6(1-\varepsilon)}{4}$	$\frac{\varepsilon^7}{4}$	$\frac{(1-\varepsilon)\varepsilon^2(1-\varepsilon^4)}{4}$	$\frac{\varepsilon(1-\varepsilon^2)(1-\varepsilon^4)}{4}$
A, B, C	$\frac{\varepsilon^7}{4}$	$\frac{\varepsilon^6(1-\varepsilon)}{4}$	$\frac{\varepsilon^3(1-\varepsilon^4)}{4}$	$\frac{(1-\varepsilon)(1-\varepsilon^2)(1-\varepsilon^4)}{4}$

Table 14: Positive joint conditional distribution used to approach the full conditional probability in Table 8. For each configuration of (P, F, A, B, C) , the corresponding cell contains its probability value.

4 INCOMPLETENESS OR INDETERMINACY?

Let's review. A full conditional probability can always be represented by a sequence of standard probability distributions. One would think that it would be enough to apply the well-known specification schemes to each layer in the sequence. Yet, we presented two examples demonstrating that this strategy is insufficient to fully specify a conditional probability distribution. In fact, one can face maximal imprecision. We considered whether augmenting probability assessments with explicit assertions of independence can avoid this imprecision, but found that none of the available concepts of independence for full conditional probabilities guarantees that a parameterized Bayesian network will factorize a unique probability. The upshot is that, even in very ordinary circumstances, situations that would seem to fix a unique full conditional probability measure fail to do so.

Before drawing some consequences for strict Bayesianism, a brief aside on terminology is due. Unfortunately, the use of the terms 'imprecision' and 'indeterminacy' is not yet standardized. Following (Walley 1991, §5.1), 'precision' refers to particular mathematical properties of subjective probability and its application to decision making, namely uniqueness in probability value, completeness of a preference ordering, additivity of the measure, or linearity of previsions (expectations). The term 'imprecision' then refers to non-uniqueness in value, incomplete preference orderings, non-additivity of the lower/upper probabilities, or non-linearity of lower/upper previsions. The terms 'determinacy' and 'indeterminacy', by contrast, refer to two different types of uncertainty (Knight 1921), thus are properties of beliefs, not mathematics. Indeterminacy therefore is a reason, but not the only reason, to use imprecise probabilities to represent your commitments. To illustrate, suppose you do not strictly prefer one of a pair of options over the other yet refuse to judge the two equivalent in value because you are uncertain how to compare their value. Your indeterminate uncertainty in your judgment about the two options will have different behavioral consequences for you than if you were to instead abide by the strict Bayesian mandate to render all uncertainty as determinate uncertainty. You might then turn to an imprecise probability model to draw out those consequences. In such cases the phrase 'indeterminate probabilities' is sometimes used as shorthand to refer to an imprecise probability model that stems from indeterminate uncertainty.

Another use of 'imprecision' and 'indeterminacy', due to Isaac Levi (1974, 2002), describes two different ways to respond, rationally, to a probability interval. An imprecise probability interval for Levi describes a decision maker's uncertainty about an event along with a mandate to resolve this uncertainty with a coherent, precise probability before the agent makes a choice. An indeterminate probability interval, by contrast, describes a decision maker who is not committed to resolving her

uncertainty with a precise probability before choice (Vicig and Seidenfeld 2012).

These two different conventions can cause confusion. For example, on Walley’s terms, a strict Bayesian can have imprecise probability values but cannot endorse imprecise (i.e., non-additive) probabilities, whereas according to Levi’s nomenclature a strict Bayesian may endorse imprecise probabilities but rejects indeterminate probabilities. Levi’s notion of imprecision describes a situation in which the decision maker believes there is a determinate probability but does not know it, a situation that Walley describes as invoking a sensitivity analysis interpretation of imprecise probability values. Both Walley and Levi refer to Knightian uncertainty as indeterminacy, although Levi’s focus on the probability interval representation of Knightian uncertainty in the context of decision making prompts him to add that there is no mandate to compel a decision maker with indeterminate probabilities to select a precise probability before choice.

For our purposes, it is enough to observe that a strict Bayesian will not accept indeterminate probabilities in either guise for the simple reason that rejecting Knightian uncertainty is a *sine qua non* of strict Bayesianism. As de Finetti and Savage put it,

“the vagueness seemingly intrinsic in certain probability assessments should not be regarded as something qualitatively different from uncertainty in any quantities, numbers and data one works with in applied mathematics” (de Finetti and Savage 1962, p. 95).

A strict Bayesian thus may acknowledge that additional probability assessments are necessary to completely specify a full conditional probability model and may even concede that these new assessments will make specifying a complete model far more demanding than previously believed. The strict Bayesian nevertheless will not grant that our examples license indeterminacy in probability values, since doing so would be to renounce strict Bayesianism.

That said, there are two considerations that, together, weigh in favor of indeterminate full conditional probabilities. The first concerns the surprising complications that arise in the models for our simple examples, complications that thwart efforts to construct rules for selecting a unique value that apply to the measures in the model. Even in our first and simplest example, Table 3 shows variation in the admissible probability models along two dimensions, namely in the number of layers of a model and the order of those layers, rather than a single dimension, thwarting attempts to “objectively” pick determinate probability values through a symmetry principle devised to operate on layers of the model. Indeed, as Table 3c demonstrates, permuting the order of layers is not even necessary for the argument in Example 1 to go through. For this reason, we are doubtful there is a suitable mechanical method operating on the elements of a full conditional probability model to pin down a single, complete specification of probability values in practical settings. Instead, numerical precision will be secured, if it is to be secured at all, by requiring additional assessments of probability values.

This brings us to our next consideration. Sometimes the additional assessments required to secure numerically precise full conditional probabilities will impose unreasonable constraints on admissible values for those assessments. Return once more to Example 2. Although no notion of independence fixes a unique probability, layer independence eliminates one of our three admissible probability models. Specifically, layer independence rules out assigning the value $1/2$ to $\mathbb{P}(P | A, \neg B, C)$. Thus, finishing the assessment of $\mathbb{P}(P | A, \neg B, C)$ under layer independence comes at the price of prohibiting the strict Bayesian from expressing complete uncertainty about $\mathbb{P}(P | A, \neg B, C)$. Instead, the strict Bayesianism is forced to be opinionated about P given $A, \neg B, C$, assigning some probability $\mathbb{P}(P | A, \neg B, C)$ that is strictly greater than or strictly less than even-odds. By contrast, we might instead view the elicited probabilities in Table 6 as a complete model of our (partially) indeterminate judgments, including a maximally indeterminate judgment about $\mathbb{P}(P | A, \neg B, C)$. Finally, as a final remark, probability values are sometimes surprisingly under-constrained, too. For as Table 3b shows, an admissible model that pins down a non-zero probability for the first toss landing heads given the event C may still fail to pin down a non-zero probability for the first toss landing tails given C .

In closing, we turn to briefly examine the opposite route; that is, whether a desire to model indeterminacy in probability values may prompt one to favor sets of full conditional probabilities.

The path from indeterminate probabilities to full conditional probabilities is complicated by choice, for there are many different options available to formalize indeterminate probability assessments. One very popular representation, and in many respects the simplest, conceptually, is through a (closed and convex) set of probability measures, \mathcal{K} , called a *credal set* (Levi 1980); for any event A , an agent entertaining an indeterminate probability estimate of A can compute its *lower probability* $\underline{\mathbb{P}}(A) = \inf_{\mathbb{P} \in \mathcal{K}} \mathbb{P}(A)$. Even though there are alternatives (Walley 2000), some of which offer more expressive power than credal sets (Williams 1975; Quaeghebeur 2014; Wheeler 2021), credal sets remain both a powerful and widespread framework for working with imprecise probabilities, particularly in philosophy, computer science, and statistics. It is natural then to consider whether a credal set ought to consist of Kolmogorovian probabilities or full conditional probabilities.

Suppose an agent is interested in computing the lower probability of some event A given some other event B . Can conditioning on events of zero lower probability be handled in the same manner as the Kolmogorovian theory treats conditioning on events of zero probability? Recall that in the Kolmogorovian setting, events of zero probability are ignored for all practical purposes. However, when some probabilities are indeterminate, we may find ourselves tasked with conditioning on an event that *may* have probability zero, yet may also have a non-zero probability that is, possibly, very large. The upshot is that conditioning on probability-zero events is a commonplace issue for indeterminate probabilities. Yet we cannot simply leave conditional lower probability undefined when the conditioning event has zero lower probability,⁵ as the next example illustrates.

EXAMPLE 3. A fair coin is tossed; if heads, then a ball is taken from urn U_1 ; if tails, then a ball is taken from urn U_2 . Each urn contains red, green, and blue balls, and the possible proportions of balls leave us with the three possible probability distributions in Table 15. Suppose the coin was flipped, a ball was taken, yet we are only informed of the color of the ball. Now we are asked, *What is the probability of heads given that announced color?* Notice that the lower probability of every color is zero. So, if we refuse to define lower conditional probability when the conditioning event has zero lower probability, we are left with no lower conditional probability at all! \square

There are two ways to define lower conditional probabilities with respect to a credal set, and these two options resemble the choice between Kolmogorovian probabilities and full conditional probabilities.

A strategy akin to Kolmogorov's ratio definition of conditional probability is to define lower conditional probability as $\underline{\mathbb{P}}(A | B) = \inf_{\mathbb{P} \in \mathcal{K}: \mathbb{P}(B) > 0} \mathbb{P}(A | B)$ whenever there is at least one element of \mathcal{K} such that $\mathbb{P}(B) > 0$; otherwise the lower conditional probability is left undefined. Similarly, the *upper conditional probability* is defined as $\overline{\mathbb{P}}(A | B) = \sup_{\mathbb{P} \in \mathcal{K}: \mathbb{P}(B) > 0} \mathbb{P}(A | B)$ whenever there is at least an element of \mathcal{K} such that $\mathbb{P}(B) > 0$. In short, according to this strategy, elements of \mathcal{K} that assign probability zero to the conditioning event are discarded. Yet, owing to \mathcal{K} being a closed convex set, \mathcal{K} without these measures is itself a closed-open set whose infimum is nevertheless zero. This is sometimes referred to as *regular conditioning*, as it resembles Walley's concept of *regular extension* (Walley 1991,

	red	green	blue
U_1	0	1/6	1/3
U_2	0	1/3	1/6

	red	green	blue
U_1	1/6	0	1/3
U_2	1/3	0	1/6

	red	green	blue
U_1	1/6	1/3	0
U_2	1/3	1/6	0

Table 15: The possible probability distributions for Example 3. Note that $\underline{\mathbb{P}}(\text{red}) = \underline{\mathbb{P}}(\text{green}) = \underline{\mathbb{P}}(\text{blue}) = 0$.

⁵As Giron and Rios do, for example, as their preferences are undefined when conditioned on events of zero probabilities (Giron and Rios 1980).

Appendix J). Regular conditioning is adopted for instance in Weichselberger’s theory of interval probabilities (Weichselberger 2000). Regular conditioning eliminates exactly those probability measures that are denied conditioning in the Kolmogorovian approach.

The arguments against Kolmogorovian-style conditioning reviewed in Section 2 can be mounted against regular conditioning. In essence, regular conditioning presents lower conditional probability as merely derived from unconditional probabilities; this is questionable as conditional probabilities do make sense in many circumstances regardless of the probability of the conditioning event. Moreover, one motivation for credal set representations is that the set of measures provide the basis for the indeterminate judgment, and of particular importance are the specific measures in \mathcal{K} that determine the span of values that events can take. The information destroyed by eliminating those measures is not necessarily recoverable from lower and upper conditional probabilities.

Moving away from regular conditioning, the other possible strategy is to assume that a credal set is always a set of full conditional probabilities. In this case we can define the lower conditional probability $\underline{\mathbb{P}}(A | B)$ directly as $\inf_{\mathbb{P} \in \mathcal{K}} \mathbb{P}(A | B)$ whenever $B \neq \emptyset$. If one adopts this *set of full conditional probabilities* strategy, then the circle is closed: natural assessment strategies applied to full conditional probabilities lead to indeterminacy in probability values, and conditioning in the presence of indeterminacy in probability values resorts to element-wise manipulation of full conditional probabilities.

ACKNOWLEDGMENTS

Gregory Wheeler’s research was supported in part by the joint *Agence Nationale de la Recherche* (ANR) & *Deutsche Forschungsgemeinschaft* (DFG) project “Collective Attitudes Formation” Co-IAForm, award RO 4548/8-1, and Fabio Cozman’s by awards from the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) and *Fundação de Amparo à Pesquisa do São Paulo* (FAPESP).

REFERENCES

- Adams, E. W. (1966). Probability and the logic of conditionals. In J. Hintikka and P. Suppes (Eds.), *Aspects of Inductive Logic*. North Holland.
- Battigalli, P. and P. Veronesi (1996). A note on stochastic independence without Savage-null events. *Journal of Economic Theory* 70(1), 235–248.
- Blume, L., A. Brandenburger, and E. Dekel (1991a, January). Lexicographic probabilities and choice under uncertainty. *Econometrica* 58(1), 61–79.
- Blume, L., A. Brandenburger, and E. Dekel (1991b, January). Lexicographic probabilities and equilibrium refinements. *Econometrica* 58(1), 81–98.
- Chandler, J. (2014). Subjective probabilities need not be sharp. *Erkenntnis* 7(6), 1273–1286.
- Coletti, G. and R. Scozzafava (2002). *Probabilistic Logic in a Coherent Setting*. Trends in logic, 15. Dordrecht: Kluwer.
- Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter (1999). *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag.
- Cozman, F. G. (2013). Independence for full conditional probabilities: Structure, factorization, non-uniqueness, and Bayesian networks. *International Journal of Approximate Reasoning* 54, 1261–1278.
- Cozman, F. G. and T. Seidenfeld (2009). Independence for full conditional measures and their graphoid properties. In B. Lowe, E. Pacuit, and J.-W. Romeijn (Eds.), *Reasoning about Probabilities and Probabilistic Reasoning*, Volume 16 of *Foundations of the Formal Sciences VI*, pp. 1–29. London: College Publications.
- de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l’Institut Henri Poincaré* 7(1), 1–68. Trans. Henry E. Kyburg, Jr., Foresight: Its logical laws, its subjective sources. In H. E. Kyburg, Jr. and H. Smokler (Eds.), *Studies in Subjective Probability*. New York: John Wiley and Sons, 1964.

- de Finetti, B. (1949). Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Triestini dell'Università* 19, 19–29. Trans. Giandomenico Machi and Adrian Smith, On the Axiomatization of Probability Theory. In Bruno de Finetti, *Probability, Induction and Statistics: The Art of Guessing*. London: John Wiley and Sons, 1972.
- de Finetti, B. (1974). *Theory of Probability*, vol. 1-2. New York: Wiley.
- de Finetti, B. and L. J. Savage (1962). Sul modo di scegliere le probabilità iniziali. *Biblioteca del Metron, Serie C 1*, 81–154.
- Dizadji-Bahmani, F., R. Frigg, and S. Hartmann (2011). Confirmation and reduction: a Bayesian account. *Synthese* 179, 321–338.
- Dubins, L. E. (1975). Finitely additive conditional probability, conglomerability and disintegrations. *Annals of Statistics* 3(1), 89–99.
- Easwaran, K. (2011). Varieties of conditional probability. In P. Bandyopadhyay and M. Forster (Eds.), *Philosophy of Statistics*, Volume 7 of *Handbook of the Philosophy of Science*. Amsterdam: North Holland.
- Elga, A. (2010, May). Subjective probabilities should be sharp. *Philosophers Imprint* 10(5).
- Gilio, A. (2012). Generalizing inference rules in a coherence-based probabilistic default reasoning. *International Journal of Approximate Reasoning* 53(3), 413–434.
- Giron, F. J. and S. Rios (1980). Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics*, pp. 17–38. Valencia, Spain: University Press.
- Hájek, A. (2003). What conditional probability could not be. *Synthese* 137, 273–323.
- Hammond, P. J. (1994). Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys (Ed.), *Patrick Suppes: Scientific Philosopher; Volume 1*, pp. 25–59. Dordrecht, The Netherlands: Kluwer.
- Joyce, J. (2011, December). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives* 24(1), 281–323.
- Kadane, J. B., M. J. Schervish, and T. Seidenfeld (1999). *Rethinking the Foundations of Statistics*. Cambridge Series in Probability, Induction and Decision Theory. Cambridge University Press.
- Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan and Co.
- Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Boston: Hart, Schaffner & Marx; Houghton Mifflin Company.
- Kohlberg, E. and P. J. Reny (1997). Independence on relative probability spaces and consistent assessments in game trees. *Journal of Economic Theory* 75, 280–313.
- Kraus, S., D. Lehmann, and M. Magidor (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 14(1), 167–207.
- Levi, I. (1974). On indeterminate probabilities. *Journal of Philosophy* 71, 391–418.
- Levi, I. (1980). *The Enterprise of Knowledge*. Cambridge, Massachusetts: MIT Press.
- Levi, I. (2002). Indeterminate probability and change of view. In S. Haller and G. Simmons (Eds.), *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2002)*, pp. 503–507. AAAI.
- Mahtani, A. (2017). Imprecise probabilities and unstable betting behavior. *Noûs*.
- Makinson, D. (2011). Conditional probability in the light of qualitative belief change. *Journal of Philosophical Logic* 40(2), 121–153.
- McGee, V. (1994). Learning the impossible. In E. Bells and B. Skyrms (Eds.), *Probability and Conditionals*, pp. 179–199. Cambridge University Press.
- Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann.
- Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis* 79(6), 1305–1342.
- Pedersen, A. P. and G. Wheeler (2015). Dilation, disintegrations, and delayed decisions. In *Proceedings of the 9th Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Pescara, Italy, pp. 227–236.

- Pedersen, A. P. and G. Wheeler (2019). Dilation and asymmetric relevance. In J. De Bock, C. P. Campos, G. de Cooman, E. Quaeghebeur, and G. Wheeler (Eds.), *Proceedings of Machine Learning Research*, Volume 103 of *Proceedings of the 11th Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, pp. 324–326.
- Pfeifer, N. and L. Tulkki (2017). Conditionals, counterfactuals, and rational reasoning: An experimental study on basic principles. *Minds and Machines 1*, 119–165.
- Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. London: Routledge.
- Quaeghebeur, E. (2014). Desirability. In T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes (Eds.), *Introduction to Imprecise Probabilities*, pp. 1–27. Wiley.
- Renyi, A. (1955). On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hungarica 6*, 285–335.
- Seidenfeld, T., M. J. Schervish, and J. B. Kadane (2001). Improper regular conditional distributions. *The Annals of Probability 29*(4), 1612–1624.
- Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *Annals of Statistics 21*(9), 1139–1154.
- Spohn, W. (1986). The representation of Popper measures. *Topoi 5*(1), 69–74.
- Sprengr, J. and S. Hartmann (2017, April). Bayesian philosophy of science. Unpublished Manuscript.
- Swinkels, J. M. (1993). Independence for conditional probability systems. Technical Report 1076, Northwestern University, Center for Mathematical Studies in Economics and Management Science.
- van Fraassen, B. C. (1976). Representation of conditional probabilities. *Journal of Philosophical Logic 5*, 417–430.
- Vantaggi, B. (2001). Conditional independence in a coherent finite setting. *Annals of Mathematics and Artificial Intelligence 32*(1-4), 287–313.
- Vicig, P. and T. Seidenfeld (2012). Bruno de Finetti and imprecision: Imprecise probability does not exist! *International Journal of Approximate Reasoning 53*, 1115–1123.
- Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.
- Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning 24*, 125–148.
- Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning 24*(2-3), 149–170.
- Wheeler, G. (2021). A gentle approach to imprecise probability. In T. Augustin, F. Cozman, , and G. Wheeler (Eds.), *Essays in Honor of Teddy Seidenfeld*, Theory and Decision Library A, Dordrecht. Springer.
- White, R. (2010). Evidential symmetry and mushy credence. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Volume 3, pp. 161–186. Oxford University Press.
- Williams, P. M. (1975). Notes on conditional previsions. School of Mathematical and Physical Sciences, University of Sussex. Republished in *International Journal of Approximate Reasoning*, 44(3): 366–383, 2007.