Forthcoming in *Erkenntnis*. Penultimate version.

Demystifying Dilation

Arthur Paul Pedersen
Gregory Wheeler

Abstract

Dilation occurs when an interval probability estimate of some event E is properly included in the interval probability estimate of E conditional on every event F of some partition, which means that one's initial estimate of E becomes less precise no matter how an experiment turns out. Critics maintain that dilation is a pathological feature of imprecise probability models, while others have thought the problem is with Bayesian updating. However, two points are often overlooked: (i) knowing that E is stochastically independent of F (for all F in a partition of the underlying state space) is sufficient to avoid dilation, but (ii) stochastic independence is not the only independence concept at play within imprecise probability models. In this paper we give a simple characterization of dilation formulated in terms of deviation from stochastic independence, propose a measure of dilation, and distinguish between proper and improper dilation. Through this we revisit the most sensational examples of dilation, which play up independence between *dilator* and *dilatee*, and find the sensationalism undermined by either fallacious reasoning with imprecise probabilities or improperly constructed imprecise probability models.

1 Good Grief!

Unlike free advice, which can be a real bore to endure, accepting free information when it is available seems like a Good idea. In fact, it is: I. J. Good (1967) showed that under certain assumptions it pays you, in expectation, to acquire new information when it is free. This Good result reveals why it is rational, in the sense of maximizing expected utility, to use all freely available evidence when estimating a probability.

Another Good idea, but not merely a Good idea, is that probability estimates may be imprecise (Good 1952, p. 114).[1] Sometimes total evidence is insufficient to yield numerically determinate estimates of probability, or precise *credences* as

[1]Other notable pioneers of imprecise probability include B. O. Koopman (Koopman 1940), Alfreds Horn and Tarski (Horn and Tarski 1948), Paul Halmos (Halmos 1950), C. A. B. Smith (Smith 1961), Daniel Ellsberg (Ellsberg 1961), Henry Kyburg, Jr. (Kyburg 1961) and Isaac Levi . Notable contemporary advocates include Isaac Levi, Peter Walley, Teddy Seidenfeld, James Joyce, Fabio Cozman (Cozman 2000), Gert de Cooman and Enrique Miranda (de Cooman and Miranda 2007, 2009).

some may say, but instead only yield upper and lower constraints on probability estimates, or indeterminate *credal states*, as Isaac Levi likes to say (Levi 1974, 1980). The problem is that these two commitments can be set against one another by a phenomenon called *dilation*.[2] An interval probability estimate for a hypothesis is dilated by new evidence when the probability estimate for the hypothesis is strictly contained within the interval estimate of the hypothesis given some outcome from an experiment. It is no surprise that new information can lead one to waver. But there is more. Sometimes the interval probability estimate of a hypothesis dilates *no matter how the experiment turns out.* Here merely running the experiment, whatever the outcome, is enough to degrade your original estimate. Faced with such an experiment, should you refuse a free offer to learn the outcome? Is it rational for you to pay someone to *not* tell you?

Critics have found dilation beyond the pale but divide over why. For the rearguard, the prospect of increasing one's imprecision over a hypothesis no matter how an experiment turns out is tantamount to a *reductio* argument against the theory of imprecise probabilities. Conditioning on new information should reduce your ignorance, tradition tells us, unless the information is irrelevant, in which case we should expect there to be no change to your original estimate. However, dilation describes a case where the specific outcome of the experiment is irrelevant but imprecision increases by conditioning, come what may.

The conservatives lament that the proponents of imprecise probabilities trade established distinctions and time-honored methods for confusion and ruin. Simply observing the distinction between *objective* and *subjective* probabilities and sticking to Laplace's *principle of indifference* (White 2010), or the distinction between *known evidence* and *belief* (Williamson 2007, pp. 176-7) and calibrating belief by the *principle of maximum entropy* (Williamson 2010, Wheeler 2012), they argue, would avoid the dilation hullabaloo. The real debate for conservatives concerns which traditions to follow—not whether to abandon numerically determinate probabilities. Even some who think that belief states should be indeterminate to "match the character" of the evidence despair of imprecise probability theory ever being of service to epistemology (Sturgeon 2008, 2010, Wheeler 2013).

For the vanguard, imprecision is an unavoidable truth, and dilation is but another reason to reject Good's first idea in favor of selective but shrewd updating. Henry Kyburg, for example, long interested in the problem of selecting the appropriate reference class (Kyburg 1961, Kyburg and Teng 2001), avoids dilation by always selecting the most unambiguously precise estimate available.[3] There is no

[2]The first systematic study of dilation is (Seidenfeld and Wasserman 1993), which includes historical remarks that identify Levi and Seidenfeld's reaction to (Good 1967) as the earliest observation of dilation and Good's reply in (1974) as the first published record. Seidenfeld and Wasserman's study is further developed in (Herron et al. 1994) and (Herron et al. 1997). See our note 10, below, which discusses a variety of weaker dilation concepts that can be articulated and studied.

[3]Although evidential probability avoids strict dilation, there are cases where adding new information from conflicting but seemingly *ad hoc* reference classes yields a less precise estimate. See, for example, Seidenfeld's hollow cube example in (Seidenfeld 2007) and Kyburg's reply in the same volume (Kyburg 2007).

possibility for dilation to occur within his theory of evidential probability, but this policy is what places evidential probability in conflict with Bayesian conditionalization (Kyburg 1974, Levi 1977). Even so, the general idea of selective updating may not be incompatible with classical Bayesian methods (Harper 1982). Indeed, one recent proposal to avoid dilation replaces Good's first principle by a second-order principle purported to determine whether or not it pays you in expectation to update a particular hypothesis on a particular item of evidence (Grünwald and Halpern 2004). This approach faces the problem of how to interpret second-order probabilities (Savage 1972, p. 58), but that is another story.

A point often overlooked by dilation detractors—conservatives and progressives alike—is that dilation requires that your evidence about pairs of events in question to not rule out the possibility for some interaction between them (Seidenfeld and Wasserman 1993). This is a crucial point, for the most sensational alleged cases of dilation—recent examples include (Sturgeon 2010, White 2010, Joyce 2011), but also consider (Seidenfeld 1994)—appear to involve stochastically independent events which nevertheless admit some mysterious interaction to occur. Yet each of these recent examples rests on an equivocation concerning whether the events in question are indeed stochastically independent. If one event is completely stochastically independent of another, an implication of Seidenfeld and Wasserman's fundamental results on dilation tells us that there is no possibility for one event to mysteriously dilate the other. Claims to the contrary are instances of mishandled imprecise probabilities—not counterexamples to a theory of imprecise probabilities.

Another source of confusion over indeterminate probabilities is the failure to recognize that there are several distinct concepts of probabilistic independence and that they only become extensionally equivalent within a standard, numerically determinate probability model. This means that some sound principles of reasoning about probabilistic independence within determinate probability models are invalid within imprecise probability models. To take an example, within the class of imprecise probability models it does not follow that there must be zero correlation between two variables when the estimate of an event obtaining with respect to one of the variables is unchanged by conditioning on any outcome of the other: one event can be *epistemically irrelevant* to another without the two events enjoying complete stochastic independence.

The aim of this essay is to help demystify dilation by first giving necessary and sufficient conditions for dilation in terms of deviations from stochastic independence. Our simple characterization of dilation is new, improving on results of Seidenfeld, Wassermann and Heron, who have provided necessary but insufficient conditions, sufficient but unnecessary conditions, and characterization results which apply to some classes of models but not to all. We also propose a measure of dilation.

Second, we delineate three distinct but logically related independence concepts and explain how those notions behave within a very general family of imprecise probability models. This account is then used to explain what goes wrong in a recent line of attack against the theory of imprecise probabilities, and to explain how the theory of imprecise probabilities is more accommodating than some of its advocates have suggested. It should be stressed that this essay offers neither an exhaustive treatment of independence within imprecise probability models,[4] nor an exhaustive defense of imprecise probabilities.[5] Yet, as should become clear, our characterization results and the role these three independence concepts play are fundamental to understanding imprecise probabilities and their application.

The general class of imprecise probability theories considered in this essay cover several proper extensions of familiar, numerically determinate models: the underlying structure of numerically determinate probability models drop out as a special case of the more general indeterminate theory. This means that the plurality of independence concepts, and therefore the underlying mechanics which govern dilation, run deeper than the particular philosophical interpretations which normally lead discussions of theories of probability. As a result, we will make a judicious effort, to the extent we can, to place the mathematics driving dilation in the foreground and the interpretations advanced for different probability models in the background. Proceeding this way is not to devalue questions concerning interpretations of probability models. On the contrary, since numerically determinate probability models are simply a special case of this family of imprecise probability models, a large part of the philosophical discussion over proper interpretations can and indeed should be conducted on neutral grounds. This essay may be viewed as a guide to finding that neutral ground.

2 Preliminaries

When you are asked to consider a series of fair coin tosses, what you are being invited to think about, in one fashion or another, is an idealized mathematical model: a sequence of independent Bernoulli trials with probability $1/2$ for the outcome heads occurring on each toss.

In this section we explain each piece of this mathematical model. In the next we discuss variations of a coin toss experiment consisting of two tosses.

Probability. A *probability function* is a real-valued function p defined on an algebra $\mathscr{A}$ over a set of states Ω satisfying the following three conditions:

(P1) $p(E) \geq 0$ for every $E \in \mathscr{A}$;

(P2) $p(\Omega) = 1$;

(P3) $p(E \cup F) = p(E) + p(F)$ for all pairwise disjoint elements E and F in $\mathscr{A}$.

[4] See (Couso et al. 1999), (Cozman 2012), and (de Cooman and Miranda 2009).

[5] For instance, Adam Elga's (2010) alleged counterexample to imprecise decision models conflates extensive and normal form games, which is a valid reduction within standard, numerically determinate decision models, but invalid within an imprecise decision model (Seidenfeld 1994).

In plain terms, p is a single probability function assigning to each *event* in the algebra $\mathscr{A}$ a numerically determinate real number. The triple $(\Omega, \mathscr{A}, p)$ is called a *probability space*.

For arbitrary events E and F in $\mathscr{A}$, an immediate consequence of properties (P1) – (P3) is a generalization of (P3):

(P3′) $\quad p(E \cup F) \;=\; p(E) \;+\; p(F) \;-\; p(E \cap F).$

According to one way of understanding imprecision, even though p is a single, well-defined function, by strategically withholding information about $(\Omega, \mathscr{A}, p)$, it may be only possible to derive an interval constraint for a probability assignment rather than a numerically determinate value.

For example, suppose that numerically precise estimates have been evaluated for a subcollection $\mathscr{E} \subseteq \mathscr{A}$ of events including, say, E and F, but not their joint occurrence, $E \cap F$. In particular, suppose that $p(E) = 1/2$ and $p(F) = 1/2$, while a precise value for the binary meet of E and F, $E \cap F$, has not been specified: Solving for $\beta = p(E \cap F)$ admits any real number within $[0, 1/2]$ as a feasible value for β. This calculation for binary meets and binary joins when only the marginal probabilities of a pair of events have been specified conforms to the pair of rules from the following proposition.

Proposition 2.1 *Suppose that $p(E)$ and $p(F)$ are defined. Then*:

1. *If $p(E \cap F) \;=\; \beta$, then*:

$$\max\Big[\, 0,\; \big(p(E) \,+\, p(F)\big) \,-\, 1 \,\Big] \;\leq\; \beta \;\leq\; \min\Big[\, p(E),\; p(F) \,\Big];$$

2. *If $p(E \;\cup\; F) \;=\; \beta$, then*:

$$\max\Big[\, p(E),\; p(F) \,\Big] \;\leq\; \beta \;\leq\; \min\Big[\, p(E) \,+\, p(F),\; 1 \,\Big].$$

In view of this proposition, a first remark about imprecise probability assignments is that they may arise naturally when some information has not been specified. Nothing exotic or heterodox need obscure them.

Affirming a range of solutions β_E for each event E is to say that there is a set $\mathbb{P}$ of probability functions assigning real numbers β_E to each such event E. Each p in $\mathbb{P}$ is defined with respect to the same set of states Ω and algebra $\mathscr{A}$. Since the set of numbers β_E is bounded below, there is a *greatest* number $\underline{\beta_E}$ bounding these numbers from below. Similarly, since the set of numbers β_E is bounded above, there is a *least* number $\overline{\beta_E}$ bounding these numbers from above. These numbers correspond to the lower and upper probabilities for E: The *lower probability* of E in $\mathbb{P}$ is $\underline{\beta_E}$, and the *upper probability* of E in $\mathbb{P}$ is $\overline{\beta_E}$, denoted by $\underline{\mathrm{P}}(E)$ and $\overline{\mathrm{P}}(E)$, respectively. Putting this in formal terms, these numbers are defined by the following equations:

(P4.1) $\underline{\mathrm{P}}(E) = \inf\{p(E) : p \in \mathbb{P}\}.$

(P5.1) $\overline{\mathrm{P}}(E) = \sup\{p(E) : p \in \mathbb{P}\}.$[6]

If $\underline{\mathrm{P}}(F) > 0$, then conditional lower and conditional upper probabilities are defined, respectively, as follows:

(P4.2) $\underline{\mathrm{P}}(E \mid F) = \inf\{p(E \mid F) : p \in \mathbb{P}\}.$

(P5.2) $\overline{\mathrm{P}}(E \mid F) = \sup\{p(E \mid F) : p \in \mathbb{P}\}.$

Of course, if F is the sure event Ω, conditional lower probability and conditional upper probability reduce to unconditional lower probability and upper probability, respectively.

When lower probability and upper probability agree for all events in the algebra $\mathscr{A}$, then the set $\mathbb{P}$ is a singleton set consisting of a unique precise numerical probability function on $\mathscr{A}$ which realizes the upper and lower probability function:

(P6) If $\underline{\mathrm{P}} = \overline{\mathrm{P}}$, then $\{p\} = \mathbb{P}$ and $p = \underline{\mathrm{P}} = \overline{\mathrm{P}}$.

Lower probabilities may be defined in terms of upper probabilities, and *vice versa*, through the following conjugacy relation:

(P7) $\overline{\mathrm{P}}(E) = 1 - \underline{\mathrm{P}}(E^c).$[7]

Thus, we may restrict attention to lower probabilities without loss of generality.

We remark that lower probabilities are *superadditive*:

(P8) $\underline{\mathrm{P}}(E \cup F) \geq \underline{\mathrm{P}}(E) + \underline{\mathrm{P}}(F) - \underline{\mathrm{P}}(E \cap F).$

By the conjugacy relation (P7), upper probabilities are therefore *subadditive*. Given a set of probabilities over an algebra $\mathscr{A}$ on a set of states Ω, let us as before call the triple $(\Omega, \mathscr{A}, \mathbb{P})$ a *probability space*, and call the quadruple $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ satisfying properties (P4) – (P5) a *lower probability space*.

Finally, we point out that the set of lower probabilities for an event E, $\underline{\mathbb{P}}(E) = \{p \in \mathbb{P} : p(E) = \underline{p}(E)\}$, need not be unique. This is true for the set of upper probabilities, $\overline{\mathbb{P}}(E) = \{p \in \mathbb{P} : p(E) = \overline{p}(E)\}$, and the corresponding sets of lower conditionals probabilities, $\underline{\mathbb{P}}(E|F) = \{p \in \mathbb{P} : p(E|F) = \overline{p}(E|F)\}$, and upper conditional probabilities, $\overline{\mathbb{P}}(E|F) = \{p \in \mathbb{P} : p(E|F) = \overline{p}(E|F)\}$, too.

Think of the conditions for lower probability this way. If we consider a single probability space, Proposition 2.1 tells us that a gap between lower and upper probability can open only by closing off some part of the algebra from view. Properties (P4) – (P8) accordingly furnish a barebones structure governing sets of probabilities to incorporate this game of peekabo directly within the model itself. These properties underlie a proper extension of the standard probability model: there is

[6] Alternative approaches which induce lower and upper probability are discussed in (Wheeler 2006) and (Haenni et al. 2011).

[7] E^c is the complement $\Omega \backslash E$ of E.

no reason to deviate from what the fully defined probability function says about events unless some information about the probability space has not been specified.

Although it can be useful to imagine the basic model as codifying the consequences of unknown parts of the algebra, do not assume that every lower probability model has precise probabilities kept out of sight in a game of peekabo. For example, imagine that a sample of eligible voters is asked whether they intend to vote for Mr. Smith or for his sole opponent in an upcoming election. The lower probability of voting for Smith is the proportion of respondents who pledge to vote for Smith, while the upper probability of voting for Smith is the proportion who have not pledged to vote for his opponent. Rare is the pre-election poll that finds these two groups to be one and the same, for some voters may be undecided, choosing neither to commit to Smith nor to commit to his opponent. The difference between lower probability and upper probability in this case is not due to the pollster's ignorance of the true strict preferences of the voters but to the presence of truly undecided voters in the sample.

To be sure, if voters must cast a ballot for one of the two candidates, then Smith and his opponent will split the votes on election day, so the proportion of votes cast for Smith will be precisely the proportion of votes not cast for his opponent. But the pre-election poll is designed to estimate voter support for Smith, not to predict the vote count for Smith on election day. The precision of the vote count is irrelevant to resolving the imprecision in a poll of pre-election attitudes. Indeed, often the very point of a pre-election poll is to identify undecided voters as part of an effort to influence how they will cast their ballots on election day.

Stochastic Independence The textbook definition of probabilistic or stochastic independence is formulated in terms of a single probability function. Thus, two events E and F in $\mathscr{A}$ are said to be *stochastically independent* just in case:

(SI) $p(E \cap F) = p(E)p(F)$.

In a standard probability space for a precise probability function p, events E and F are stochastically independent just in case conditioning on F is irrelevant to estimating E, and *vice versa*. Formally, an event F is said to be *epistemically irrelevant* to an event E precisely when:

(ER) $p(E \mid F) = p(E)$, when $p(F) > 0$.

Accordingly:

(EI) E is *epistemically independent* of F if and only if both E is epistemically irrelevant to F and F is epistemically irrelevant to E.

Although conditions (SI), (ER) and (EI) are logically equivalent with respect to a precise probability function, it turns out that these three conditions are logically *distinct* concepts with respects to imprecise probability models. We will return to this point in Section 6.

In addition, the degree to which two events diverge from stochastic independence, if they diverge at all, may be characterized by a simple measure of stochastic independence:

$$\mathrm{S}_p(E,F) \quad =_{df} \quad \frac{p(E \cap F)}{p(E)p(F)}.$$

This measure is just the covariance of E and F, $Cov(E,F) = p(E \cap F) - p(E)p(F)$, put in ratio form. Observe that $\mathrm{S}_p(E,F) = 1$ just in case E and F are stochastically independent; $\mathrm{S}_p(E,F) > 1$ when E and F are positively correlated; and $\mathrm{S}_p(E,F) < 1$ when E and F are negatively correlated.[8] The measure S naturally extends to a set of probability functions $\mathbb{P}$ as follows:

$$\begin{aligned}
\mathrm{S}^+_{\mathbb{P}}(E,F) \quad &=_{df} \quad \{p \in \mathbb{P} : \mathrm{S}_p(E,F) \; > \; 1\};\\
\mathrm{S}^-_{\mathbb{P}}(E,F) \quad &=_{df} \quad \{p \in \mathbb{P} : \mathrm{S}_p(E,F) \; < \; 1\};\\
\mathrm{I}_{\mathbb{P}}(E,F) \quad &=_{df} \quad \{p \in \mathbb{P} : \mathrm{S}_p(E,F) \; = \; 1\}.
\end{aligned}$$

The set of probability functions $\mathrm{I}_{\mathbb{P}}(E,F)$ from $\mathbb{P}$ with E and F stochastically independent is called *the surface of independence* for E and F with respect to $\mathbb{P}$. Subscripts shall be dropped when there is no danger of confusion.

Bernoulli Trials. A Bernoulli trial is an experiment designed to yield one of two possible outcomes, *success* or *failure*, which may be heuristically coded as 'heads' or 'tails,' respectively. Given a set of states Ω, the experiment C of interest will result either in *heads* or in *tails*, so either the event $(C = heads)$ obtains or the event $(C = tails)$ obtains. A fair coin toss is a Bernoulli trial C with probability $\frac{1}{2}$ for *heads*—that is, a Bernoulli trial such that the probability that $(C = heads)$ obtains is $1/2$, written $p(C = heads) = 1/2$. In general, a Bernoulli trial C with probability θ for *success* is such that the probability that $(C = success)$ obtains is θ, $p(C = success) = \theta$.

For a series of coin tosses, let C_i be the experimental outcome of the ith coin toss. To take an example, consider a sequence of fair tosses for which the probability that the outcome of the second toss is heads given that the outcome of the first toss is tails, a property which may be expressed as:

$$p(C_2 = heads \mid C_1 = tails) \; = \; 1/2. \tag{1}$$

Notation may be abbreviated by letting H_i denote the event that the outcome C_i is heads on toss i, and by letting T_i refer to the event that the outcome C_i is tails on toss i. With this shorthand notation, Equation (1), for example, becomes:

$$p(H_2 \mid T_1) \; = \; 1/2.$$

[8] The measure S has been given a variety of interpretations in philosophy of science and formal epistemology, including as a measure of coherence Shogenji (1999) and a measure of similarity Wayne (1995), and is a variation of ideas due to Yule (1911, Ch. 3). See Wheeler (2009a) for discussion and both Schlosshauer and Wheeler (2011) and (Wheeler and Scheines 2013) for a study of the systematic relationships between covariance, confirmation, and causal structure.

A sequence of fair coin tosses is a series of stochastically independent Bernoulli trials $C_1,\ldots,C_n$ with probability $\frac{1}{2}$—that is, a stochastically independent sequence of fair coin tosses $C_1,\ldots,C_n$ with $p(C_i = heads) = p(H_i) = 1/2$ for every $i = 1,\ldots,n$ and $p\big((C_1 = o_1)\cap(C_2 = o_2)\cap\cdots\cap(C_m = o_m)\big) = p(C_1 = o_1)p(C_2 = o_2)\cdots p(C_m = o_m)$ for every $m \leq n$ such that $o_i \in \{heads, tails\}$ for each $i = 1,\ldots,m$.

A useful piece of terminology comes from observing that the subset of events

$$\mathscr{B} = \Big\{(C = heads), (C = tails)\Big\}$$

partitions the outcome space Ω since, in the model under consideration, the outcome of any coin toss must be one element in $\mathscr{B}$. In general, given a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, call a collection of events $\mathscr{B}$ from $\mathscr{A}$ a *positive measurable partition* (of Ω) just in case $\mathscr{B}$ is a partition of Ω such that $\underline{P}(H) > 0$ for each partition cell H in $\mathscr{B}$. Note that elements of $\mathscr{B}$ are assumed to be events in the algebra $\mathscr{A}$ under consideration unless stated otherwise. In the coin example, the assumption that $\mathscr{B}$ is a positive measurable partition may be reasonable to maintain unless, for example, the coin is same-sided.

3 Dilation

Lower probabilities have been introduced in response to a mischievous riddler who blocks part of the algebra of events from your view in a game of peekaboo, but the question of how to interpret a set of probabilities has been intentionally set in the background. The reason for this is that the barebones model we have presented for a set of probabilities is sufficient to bring into focus the main components needed for dilation to occur, which number fewer than the properties needed to flesh out some natural interpretations for a set of probabilities.

In exchange for leaving the interpretation of a set of probabilities largely unspecified, it is hoped that readers will come to see that dilation does not hinge on whether the set of probabilities in question has been endowed with an interpretation as a model for studying sensitivity and robustness in classical Bayesian statistical inference, or as a model for aggregating a group of opinions, or as a model of indeterminate credal probabilities.

This said, we need to pick an interpretation to run our examples, so from here on we will interpret lower probability as a representation of some epistemic agent's credal states about events. To make this shift clear in the examples, 'You' will denote an arbitrary intentional system, and the set of probabilities $\mathbb{P}$ in question will denote that system's set of credal probabilities. We invite you to play along.[9]

With these preliminaries in place, we turn now to dilation.

[9] Alas, 'You', you will find, is also a Good idea. This convention has been followed by de Finetti (1974a,b) and Walley (1991), too, among others.

Dilation. Let $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ be a lower probability space, let $\mathscr{B}$ be a positive measurable partition of Ω, and let E be an event. Say that $\mathscr{B}$ *dilates* E just in case for each $H \in \mathscr{B}$:

$$\underline{\mathrm{P}}(E \mid H) < \underline{\mathrm{P}}(E) \leq \overline{\mathrm{P}}(E) < \overline{\mathrm{P}}(E \mid H).$$

In other words, $\mathscr{B}$ dilates E just in case the closed interval $\left[\underline{\mathrm{P}}(E), \overline{\mathrm{P}}(E)\right]$ is contained in the open interval $\left(\underline{\mathrm{P}}(E \mid H), \overline{\mathrm{P}}(E \mid H)\right)$ for each $H \in \mathscr{B}$. The remarkable thing about dilation is the specter of turning a *more precise* estimate of E into a *less precise* estimate, *no matter what event* from the partition occurs.[10]

Coin Example 1: The Fair Coin. Suppose that a fair coin is tossed twice. The first toss of the coin is a fair toss, but the second toss is performed in such a way that its outcome may depend on the outcome of the first toss. Nothing is known about the direction or degree of the possible dependence. Let H_1, T_1, H_2, T_2 denote the possible events corresponding to the outcomes of each coin toss.[11]

You know the coin is fair, so Your estimate of individual tosses is precise. Hence, from P6 it follows that Your upper and lower marginal probabilities are precisely $1/2$:

(a) $\quad \underline{\mathrm{P}}(H_1) = \overline{\mathrm{P}}(H_1) = 1/2 = \underline{\mathrm{P}}(H_2) = \overline{\mathrm{P}}(H_2).$

It is the interaction between the tosses that is unknown, and in the extreme the first toss may determine the outcome of the second: while the occurrence of H_1 may guarantee H_2, the occurrence of H_1 may instead guarantee T_2. Recalling Proposition 2.1, suppose Your ignorance is modeled by:

(b) $\quad \underline{\mathrm{P}}(H_1 \cap H_2) = 0 \quad$ and $\quad \overline{\mathrm{P}}(H_1 \cap H_2) = 1/2.$

Now suppose that You learn that the outcome of the second toss is heads. From (a) and (b), it follows by (P4) and (P5) that:

[10] We mention that while our terminology agrees with that of Herron et al. (1994, p. 252), it differs from that of Seidenfeld and Wasserman (1993, p. 1141) and Herron et al. (1997, p. 412), who call dilation in our sense *strict dilation*.

Indeed, weaker notions of dilation can be articulated and subject to investigation. Say that a positive measurable partition $\mathscr{B}$ *weakly dilates* E if $\underline{\mathrm{P}}(E \mid H) \leq \underline{\mathrm{P}}(E) \leq \overline{\mathrm{P}}(E) \leq \overline{\mathrm{P}}(E \mid H)$ for each $H \in \mathscr{B}$. If $\mathscr{B}$ weakly dilates E, say that (i) $\mathscr{B}$ *pseudo-dilates* E if in addition there is $H \in \mathscr{B}$ such that either $\underline{\mathrm{P}}(E \mid H) < \underline{\mathrm{P}}(E)$ or $\overline{\mathrm{P}}(E) < \overline{\mathrm{P}}(E \mid H)$ and that (ii) $\mathscr{B}$ *nearly dilates* E if in addition for each $H \in \mathscr{B}$, either $\underline{\mathrm{P}}(E \mid H) < \underline{\mathrm{P}}(E)$ or $\overline{\mathrm{P}}(E) < \overline{\mathrm{P}}(E \mid H)$. Thus, $\mathscr{B}$ pseudo-dilates E just in case the closed interval $\left[\underline{\mathrm{P}}(E), \overline{\mathrm{P}}(E)\right]$ is contained in the closed interval $\left[\underline{\mathrm{P}}(E \mid H), \overline{\mathrm{P}}(E \mid H)\right]$ for each $H \in \mathscr{B}$, with proper inclusion obtaining for some partition cell from $\mathscr{B}$, while $\mathscr{B}$ nearly dilates E just in case the closed interval $\left[\underline{\mathrm{P}}(E), \overline{\mathrm{P}}(E)\right]$ is properly contained in the closed interval $\left[\underline{\mathrm{P}}(E \mid H), \overline{\mathrm{P}}(E \mid H)\right]$ for each $F \in \mathscr{B}$. Seidenfeld and Wasserman (1993) and Herron et al. (1994, 1997) also investigate near dilation and pseudo-dilation.

[11] This is Walley's canonical dilation example (Walley 1991, pp. 298-299), except that here we are using lower probabilities instead of lower previsions.

$$\inf\left\{\frac{p(H_1 \cap H_2)}{p(H_2)} : p \in \mathbb{P}\right\} = \inf\left\{\frac{p(H_1 \cap H_2)}{1/2} : p \in \mathbb{P}\right\}$$
$$= 2 \cdot \inf\left\{p(H_1 \cap H_2) : p \in \mathbb{P}\right\}$$
$$= 0$$

and

$$\sup\left\{\frac{p(H_1 \cap H_2)}{p(H_2)} : p \in \mathbb{P}\right\} = \sup\left\{\frac{p(H_1 \cap H_2)}{1/2} : p \in \mathbb{P}\right\}$$
$$= 2 \cdot \sup\left\{p(H_1 \cap H_2) : p \in \mathbb{P}\right\}$$
$$= 1.$$

This yields:

(c) $\underline{\mathrm{P}}(H_1 \mid H_2) = 0$ and $\overline{\mathrm{P}}(H_1 \mid H_2) = 1$.

So although Your probability estimate for H_1 is precise, Your probability estimate for H_1 given that H_2 occurs is *much less* precise, with lower probability and upper probability straddling the entire closed interval $[0,1]$. An analogous argument holds if instead You learn that the outcome of the second toss is tails. Since $\mathscr{B} = \{H_2, T_2\}$ partitions the outcome space, these two cases exhaust the relevant possible observations, so the probability that the first toss lands heads *dilates* from $1/2$ to the vacuous unit interval upon learning the outcome of the second toss. Your precise probability about the first toss becomes vacuous *no matter how the first coin toss lands*. ◇

One way to understand the extremal points $\underline{\mathrm{P}}(H_1 \mid H_2) = 0$ and $\overline{\mathrm{P}}(H_1 \mid H_2) = 1$ is as two deterministic but opposing hypotheses about how the second toss is performed. One hypothesis asserts that the outcome of the second toss is certain to match the outcome of the first, whereas the second hypothesis asserts that the second toss is certain to land opposite the outcome of the first. However, after observing the second toss, Your estimate of the first toss becomes maximally imprecise because the two hypotheses predict different outcomes.

4 Dilation Explained

As made clear by Seidenfeld and Wasserman (1993), deviation from stochastic independence is essential for dilation to occur. Here are two of their observations. To give the reader a sense of the argumentation involved in demonstrating each result, we also supply proofs.

Theorem 4.1 (Seidenfeld and Wasserman, 1993, Theorem 2.2) *Let* $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ *be a lower probability space, let* $\mathscr{B}$ *be a positive measurable partition of* Ω*, and let* $E \in \mathscr{A}$*. Suppose that* $\mathscr{B}$ *dilates* E*. Then for every* $H \in \mathscr{B}$ *:*

$$\underline{\mathbb{P}}(E|H) \subseteq \mathrm{S}^{-}_{\mathbb{P}}(E,H) \quad \textit{and} \quad \overline{\mathbb{P}}(E|H) \subseteq \mathrm{S}^{+}_{\mathbb{P}}(E,H).$$

Proof. Let $H \in \mathscr{B}$, and suppose that $p \in \underline{\mathbb{P}}(E \mid H)$. Then:

$$\begin{aligned} \frac{p(E \cap H)}{p(H)} &= \underline{\mathrm{P}}(E \mid H) \\ &< \underline{\mathrm{P}}(E) \\ &\leq \overline{\mathrm{P}}(E) \\ &\leq p(E). \end{aligned}$$

Hence, $\mathrm{S}_p(E,H) < 1$, whereby $p \in \mathbb{S}^{-}(E,H)$, establishing that $\underline{\mathbb{P}}(E|H) \subseteq \mathrm{S}^{-}_{\mathbb{P}}(E,H)$. An analogous argument for a representative $p \in \overline{\mathbb{P}}(E \mid H)$ shows that $\mathrm{S}_p(E,H) > 1$, establishing $\overline{\mathbb{P}}(E|H) \subseteq \mathrm{S}^{+}_{\mathbb{P}}(E,H)$. □

Theorem 4.2 (Seidenfeld and Wasserman, 1993, Theorem 2.3) *Let* $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ *be a lower probability space, let* $\mathscr{B}$ *be a positive measurable partition of* Ω*, and let* $E \in \mathscr{A}$*. Suppose that for every* $H \in \mathscr{B}$ *:*

$$\underline{\mathbb{P}}(E) \cap \mathrm{S}^{-}_{\mathbb{P}}(E,H) \neq \emptyset \quad \textit{and} \quad \overline{\mathbb{P}}(E) \cap \mathrm{S}^{+}_{\mathbb{P}}(E,H) \neq \emptyset.$$

Then $\mathscr{B}$ *dilates* E*.*

Proof. Let $H \in \mathscr{B}$, and let $p \in \underline{\mathbb{P}}(E) \cap \mathrm{S}^{-}_{\mathbb{P}}(E,H)$. Then $p(E) = \underline{\mathrm{P}}(E)$, and since $\mathrm{S}^{-}_{p}(E,H) < 1$, it follows that $p(E \cap H) < p(E)p(H)$, whence:

$$\begin{aligned} \underline{\mathrm{P}}(E \mid H) &\leq p(E \mid H) \\ &= p(E) \\ &= \underline{\mathrm{P}}(E). \end{aligned}$$

Therefore, $\underline{\mathrm{P}}(E|H) < \underline{\mathrm{P}}(E)$ for each $H \in \mathscr{B}$. A similar argument establishes that $\overline{\mathrm{P}}(E) < \overline{\mathrm{P}}(E|H)$ for each $H \in \mathscr{B}$. □

In plain terms, Theorem 4.1 states that when an event E is dilated by a positive measurable partition $\mathscr{B}$, any probability function realizing the infimum $\underline{\mathrm{P}}(E|H)$ must be such that E and H are negatively correlated, and each probability function from the supremum $\overline{\mathrm{P}}(E|H)$ must be such that E and H are positively correlated. While Theorem 4.1 gives a necessary condition for strict dilation, Theorem 4.2 gives a sufficient condition: If for every partition cell H from a positive measurable partition $\mathscr{B}$, the infimum $\underline{\mathrm{P}}(E)$ is achieved by some probability function for which E and H are negatively correlated, and the supremum $\overline{\mathrm{P}}(E)$ is achieved by some probability function for which E and H are positively correlated, then $\mathscr{B}$ dilates E.[12]

We mention that the conclusion of Theorem 4.1 is trivially satisfied for each partition cell H for which $\underline{\mathbb{P}}(E|H)$ or $\overline{\mathbb{P}}(E|H)$ is empty. Similarly, the hypothesis of Theorem 4.2 is vacuously satisfied whenever $\underline{\mathbb{P}}(E)$ or $\overline{\mathbb{P}}(E)$ is empty. Indeed, Seidenfeld and Wasserman (1993) frame their results with respect to regularity conditions on the set of probability functions $\mathbb{P}$, requiring that it be at once a *closed* and *convex* set.[13] Together, these requirements entail that $\underline{\mathbb{P}}(E|H)$ and $\overline{\mathbb{P}}(E|H)$ are nonempty for any event H with positive lower probability.

Authors working on imprecise probabilities often require that the set of probabilities $\mathbb{P}$ under consideration be a *closed convex* set. Accordingly, we make some brief remarks about the elementary mechanics of imprecise probabilities with respect to the important, interdependent roles of closure and convexity. These remarks are a technical digression of sorts, but they place our later discussion in sharper focus. Readers may wish to skim or skip the next few paragraphs and refer back to them later.

On the one hand, to say that a set of probabilities $\mathbb{P}$ is *closed* is to assert that the set enjoys a topological property with respect to a topology called the *weak*-topology* of the topological dual of a particular collection of real-valued functions equipped with the sup-norm $\| f \| = \sup_{\omega \in \Omega} |f(\omega)|$: the set of probabilities in question includes all its limit points, so it is identical to its (topological) closure. Accordingly, in this context, a closed set is called *weak*-closed*. In the finite setting, where only finitely many events live in the algebra, a set of probabilities $\mathbb{P}$ is weak*-closed just in case it is closed with respect to the total variation norm $\| p \|_{tv} = \sup_{A \in \mathscr{A}} |p(A)|$.[14]

Intuitively, a set of probabilities is closed if the only probability functions "close" to the set are elements of the set: the set of probabilities includes its "endpoints," or *boundary*. Put differently, any probability function falling *outside* the set of probabilities can be jiggled around a small amount and remain outside the set. To illustrate, consider a probability space for a coin toss for which the set of probabilities includes all probability functions assigning probability greater than $1/4$ to the event that the coin will land heads. This set of probabilities is not closed because it is missing an endpoint, the probability function assigning $1/4$ to the event that the coin will land heads and $3/4$ to the event that the coin will land tails. By adding this probability function to the set of probabilities, however, the resulting set of probabilities becomes closed.

[12] Seidenfeld and Wasserman's results are about dependence of particular events, not about dependence of variables. Independence of variables implies independence of all their respective values, but not conversely.

[13] Specifically, $\mathbb{P}$ is assumed to be closed with respect to the total variation norm Seidenfeld and Wasserman (1993, p. 1141).

[14] To maintain brevity, we shall not go into details. See, for example, (Walley 1991, Chapter 3, Chapter 6 , Appendix D) for further details. (Rao and Rao 1983, Chapter 5) contains useful background.

On the other hand, to say that a set of probabilities $\mathbb{P}$ is *convex* is to assert that the set enjoys a vectorial property with respect to pointwise arithmetic operations: the set of probabilities in question is closed under convex combinations—that is, for all $p_1, p_2 \in \mathbb{P}$ and $\lambda \in [0,1]$, $\lambda p_1 + (1-\lambda)p_2 \in \mathbb{P}$—so it is identical to its *convex hull*:

$$\text{co}(\mathbb{P}) \quad =_{df} \quad \left\{ \sum_{i=1}^{n} \lambda_i p_i : \lambda_i \geq 0,\ p_i \in \mathbb{P},\ \text{and} \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$

That is, a set of probabilities $\mathbb{P}$ is convex just in case $\mathbb{P} = \text{co}(\mathbb{P})$.

Informally, a set of probabilities is convex if each probability function on a line segment formed from a convex combination of probability functions from the set is also an element of the set. For example, the set of probabilities from the coin toss discussed above is convex since each convex combination of two probability functions from the set of probabilities functions assigning probability greater than 1/4 to the event that the coin will lands heads is also an element of the set. However, the set of probabilities consisting of two probability functions—one probability function assigning 1/4 to the event that the coin will land heads and 3/4 to the event that the coin will land tails, the second assigning 3/4 to the event that the coin will land heads and 1/4 to the event that the coin will land tails—is not convex: A 50-50 convex combination of the two probability functions is not a member of the set of probabilities.

In the finite setting, closure with respect to the total variation norm is equivalent to closure with respect to any favored norm of Euclidean space (as they are all equivalent), so a (sequentially) closed set of probabilities is (sequentially) compact. Importantly, when the set of probabilities is in addition convex and lower probability is defined to be the lower envelope $\underline{P}$ of the set $\mathbb{P}$, as we have done above in (P4) and (P5), then the lower probability of an event is in fact the *minimum* number assigned to the event by all probability functions in the set (and not merely the *infimum*). Indeed, a probability function from the collection of *extreme points* of the set witnesses the lower probability of the event, and in this setting any compact convex set of probabilities is the convex hull of its extreme points (Walley 1991, pp. 145 ff.).[15] Thus, in the finite setting, only some conventional machinery must be employed to get things up and running.

In the general setting, where infinitely many events live in the algebra, somewhat fancier machinery must be employed. We discuss the general case in more detail in the Appendix. The upshot is that in both the finite setting and the general setting, given a nonempty weak*-closed convex set of probabilities $\mathbb{P}$ and an event E, the lower probability of E, $\underline{P}(E)$, is given by $\underline{P}(E) = \min\{p(E) : p \in \mathbb{P}\}$. More generally, when the set of probabilities $\mathbb{P}$ is no longer required to be weak*-closed and convex, then where $\overline{\text{co}}(\mathbb{P})$ denotes the weak*-closed convex hull of $\mathbb{P}$, we have $\underline{P}(E) = \min\{p(E) : p \in \overline{\text{co}}(\mathbb{P})\}$. The following proposition records this

[15] An *extreme point* of a set of probabilities is a probability function from the set that cannot be written as a nontrivial convex combination of elements from the set.

property of lower probabilities. We include the proof in the Appendix to illustrate the mechanics discussed above in action.

Proposition 4.3 *Let* $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ *be a lower probability space. Then for every* $E, H \in \mathscr{A}$ *such that* $\underline{\mathrm{P}}(H) > 0$ *:*

$$\underline{\mathrm{P}}(E) = \min\{\, p(E) : p \in \overline{\mathrm{co}}(\mathbb{P}) \,\};$$

$$\underline{\mathrm{P}}(E|H) = \min\{\, p(E|H) : p \in \overline{\mathrm{co}}(\mathbb{P}) \,\}.$$

We thus see that specifying $\underline{\mathrm{P}}$ in the quadruple $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ does not render $\mathbb{P}$ superfluous. To be sure, two different sets of probability functions may yield the same lower probability function. To take a simple example, consider a coin that with lower probability $1/4$ will land heads and with lower probability $1/4$ will land tails. Such a lower probability function may be realized by, for example, the two point set $\mathbb{P}_1 =_{df} \{p, q\}$, where p assigns probability $1/4$ to heads and q assigns probability $1/4$ to tails, or by the convex closure $\mathbb{P}_2 =_{df} \mathrm{co}(\mathbb{P}_1) = \overline{\mathrm{co}}(\mathbb{P}_1)$ (also a weak*-closed set).

Returning to our discussion of dilation, although the conclusion of Theorem 4.1 is necessary for dilation, it is easily seen that the conclusion is not sufficient, even if the set of probabilities $\mathbb{P}$ is weak*-closed and convex, as implied by Proposition 4.3. Assuming that $\mathbb{P}$ is weak*-closed and convex, perhaps a second necessary condition will be enough for sufficiency.

Theorem 4.4 (Seidenfeld and Wasserman, 1993, Theorem 2.1) *Let* $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ *be a lower probability space such that* $\mathbb{P}$ *is weak*-closed and convex, let* $\mathscr{B}$ *be a positive measurable partition of* Ω*, and let* $E \in \mathscr{A}$*. Suppose that* $\mathscr{B}$ *dilates* E*. Then for every* $H \in \mathscr{B}$ *:*

$$\mathbb{P} \cap \mathrm{I}_{\mathbb{P}}(E, H) \neq \emptyset.$$

In other words, a necessary condition for dilation is that the surface of independence cuts through $\mathbb{P}$. As Seidenfeld and Wasserman indicate, however, the conditions comprising the conclusions of Theorem 4.1 and Theorem 4.4 are again easily seen to be insufficient for the occurrence of dilation.[16] Likewise, although the hypothesis of Theorem 4.2 is sufficient for dilation, the hypothesis is not necessary, even if $\mathbb{P}$ is weak*-closed and convex.

Seidenfeld and Wasserman (1993) claim that the aforementioned theorems show that "the independence surface plays a crucial role in dilation" (p. 1142).

[16] We note that in their article, Seidenfeld and Wasserman (1993) assume that the set of probabilities under consideration is convex and closed with respect to the total variation norm. In the special case they consider, Seidenfeld and Wasserman (1993, p. 1142) correctly point out that their Theorem 2.1 goes through without the assumption of closure and that their Theorem 2.2, Theorem 2.3, and Theorem 2.4 (below) go through without the assumption of convexity (p. 1143). However, we have shown that their Theorem 2.2 and Theorem 2.3 trivially go through even without the assumptions of convexity and closure.

Observing that Theorem 4.1 and Theorem 4.4 offer necessary but jointly insufficient conditions for dilation while Theorem 4.2 provides a sufficient but unnecessary condition for dilation, they assert that "a variety of cases exist in between" (Seidenfeld and Wasserman 1993, p. 1142). This remark may leave the impression that the "variety of cases" in between are somehow irreconcilable, resisting a uniform classification, an impression which may perhaps be supported by Seidenfeld and Wasserman's hodgepodge of results in a series of articles. While Seidenfeld and Wasserman may have maintained that the "variety of cases" in between resist uniform classification as a result of an analysis tied to viewing dilation through the properties of supporting hyperplanes rather than *neighborhoods*, which we introduce in the next section, it may very well be that the purpose of pointing to the "variety of cases" in between is simply to frame their research program.

Whatever the case, Seidenfeld and Wasserman (1993), along with Herron (1994, 1997), have explored a number of different cases, sometimes offering necessary and sufficient conditions for dilation with respect to certain regularity assumptions consistent with paradigmatic models.[17] While focusing on special classes of probability spaces to explore their status with respect to the phenomenon of dilation may be a valuable exercise, the presence of dilation can be straightforwardly shown to admit a complete characterization in terms of rather simple conditions of deviation from stochastic independence. We articulate this characterization in the next section. Such a characterization may facilitate a smoother, integrated investigation of the existence and extent of dilation as well as questions concerning the preservation of dilation under coarsenings, questions Seidenfeld, Wasserman, and Herron have addressed in their articles. Importantly, investigations and explanations of dilation need not be tied to the particular way in which Seidenfeld and Wasserman have framed their own research program.

5 A Simple Characterization of Dilation

In this section, we offer simple necessary and sufficient conditions for dilation formulated in terms of deviation from stochastic independence, much like the conditions from Theorem 4.1. We illustrate an immediate application of such a characterization with measures of dilation. To begin, we introduce some notation.

Given a nonempty set of probabilities $\mathbb{P}$, events $E, H \in \mathscr{A}$ with $\underline{\mathrm{P}}(H) > 0$, and $\varepsilon > 0$, define:

$$\begin{aligned}\underline{\mathbb{P}}(E|H,\varepsilon) &=_{df} \{p \in \mathbb{P} : |p(E|H) - \underline{\mathrm{P}}(E|H)| < \varepsilon\};\\ \overline{\mathbb{P}}(E|H,\varepsilon) &=_{df} \{p \in \mathbb{P} : |p(E|H) - \overline{\mathrm{P}}(E|H)| < \varepsilon\}.\end{aligned}$$

We call the sets $\underline{\mathbb{P}}(E|H,\varepsilon)$ and $\overline{\mathbb{P}}(E|H,\varepsilon)$ lower and upper *neighborhoods* of E conditional on H, respectively, with radius ε. Thus, a probability function is an

[17] They also discuss total variation neighborhoods and ε-contamination neighborhoods, but these neighborhoods are distinct from the neighborhoods we discuss.

element of a lower neighborhood of E conditional on H with radius ε if $p(E|H)$ is within ε of $\underline{\mathrm{P}}(E|H)$, and similarly for an upper neighborhood.

Given a nonempty set I, we let $\mathbb{R}^I_+$ denote the set of elements $(r_i)_{i\in I}$ of $\mathbb{R}^I$ such that $r_i > 0$ for each $i \in I$. We now state a proposition characterizing (strict) dilation and then deduce a few immediate corollaries. The proof can be found in the Appendix.

Proposition 5.1 *Let $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ be a lower probability space such that $\mathbb{P}$ is weak*-closed and convex, let $\mathscr{B} = \{H_i : i \in I\}$ be a positive measurable partition, and let $E \in \mathscr{A}$. Then the following are equivalent*:

(i) *$\mathscr{B}$ dilates E;*

(ii) *There is $(\varepsilon_i)_{i\in I} \in \mathbb{R}^I_+$ such that for every $i \in I$:*

$$\underline{\mathbb{P}}(E|H_i, \varepsilon_i) \subseteq \mathrm{S}^-(E, H_i) \quad \textit{and} \quad \overline{\mathbb{P}}(E|H_i, \varepsilon_i) \subseteq \mathrm{S}^+(E, H_i);$$

(iii) *There are $(\underline{\varepsilon}_i)_{i\in I} \in \mathbb{R}^I_+$ and $(\overline{\varepsilon}_i)_{i\in I} \in \mathbb{R}^I_+$ such that for every $i \in I$:*

$$\underline{\mathbb{P}}(E|H_i, \underline{\varepsilon}_i) \subseteq \mathrm{S}^-(E, H_i) \quad \textit{and} \quad \overline{\mathbb{P}}(E|H_i, \overline{\varepsilon}_i) \subseteq \mathrm{S}^+(E, H_i).$$

Furthermore, each radius $\underline{\varepsilon}_i$ may be chosen to be the unique positive minimum of $|p(E|H_i) - \underline{\mathrm{P}}(E|H_i)|$ attained on $C_i^+ =_{df} \{p \in \mathbb{P} : \mathrm{S}_p(E, H_i) \geq 1\}$, and similarly each radius $\overline{\varepsilon}_i$ may be chosen to be the unique positive minimum of $|p(E|H_i) - \overline{\mathrm{P}}(E|H_i)|$ attained on $C_i^- =_{df} \{p \in \mathbb{P} : \mathrm{S}_p(E, H_i) \leq 1\}$.

Accordingly, Proposition 5.1 implies that a positive measurable partition $\mathscr{B}$ dilates an event E just in case for each partition cell H, there are upper and lower neighborhoods of E conditional on H such that the lower neighborhood of E conditional on H lies entirely within the subset of the set of probabilities in question for which E and H are negatively correlated, while the upper neighborhood of E given H lies entirely within the subset of the set of probabilities in question for which E and H are positively correlated.

Proposition 5.1 immediately yields a corollary for *arbitrary* nonempty sets of probabilities. For the sake for readability in what follows, given a nonempty set of probabilities $\mathbb{P}$, let $\mathbb{P}_* =_{df} \overline{\mathrm{co}}(\mathbb{P})$ (i.e., the weak*-closed convex hull of $\mathbb{P}$). Thus, $\underline{\mathbb{P}}_*(E|F, \varepsilon) = \underline{\overline{\mathrm{co}}(\mathbb{P})}(E|F, \varepsilon)$ and $\overline{\mathbb{P}}_*(E|F, \varepsilon) = \overline{\overline{\mathrm{co}}(\mathbb{P})}(E|F, \varepsilon)$, the written expressions themselves justifying introducing abbreviations. Similarly, let $\mathrm{S}^+_*(E, F)$ and $\mathrm{S}^-_*(E, F)$ be defined by:

$$\begin{aligned} \mathrm{S}^+_*(E, F) &=_{df} \{p \in \overline{\mathrm{co}}(\mathbb{P}) : S_p(E, F) > 1\} \\ \mathrm{S}^-_*(E, F) &=_{df} \{p \in \overline{\mathrm{co}}(\mathbb{P}) : S_p(E, F) < 1\}. \end{aligned}$$

The next corollary of Proposition 5.1 follows immediately from Proposition 4.3. The proof can also be found in the Appendix.

Corollary 5.2 *Let* $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ *be a lower probability space, let* $\mathscr{B} = \{H_i : i \in I\}$ *be a positive measurable partition, and let* $E \in \mathscr{A}$. *Then the following are equivalent*:

(i) $\mathscr{B}$ *dilates* E;

(ii) *There is* $(\varepsilon_i)_{i\in I} \in \mathbb{R}_+^I$ *such that for every* $i \in I$:

$$\underline{\mathbb{P}}_*(E|H_i, \varepsilon_i) \subseteq \mathrm{S}_*^-(E, H_i) \quad and \quad \overline{\mathbb{P}}_*(E|H_i, \varepsilon_i) \subseteq \mathrm{S}_*^+(E, H_i);$$

(iii) *There are* $(\underline{\varepsilon}_i)_{i\in I} \in \mathbb{R}_+^I$ *and* $(\overline{\varepsilon}_i)_{i\in I} \in \mathbb{R}_+^I$ *such that for every* $i \in I$:

$$\underline{\mathbb{P}}_*(E|H_i, \underline{\varepsilon}_i) \subseteq \mathrm{S}_*^-(E, H_i) \quad and \quad \overline{\mathbb{P}}_*(E|H_i, \overline{\varepsilon}_i) \subseteq \mathrm{S}_*^+(E, H_i).$$

Furthermore, each radius $\underline{\varepsilon}_i$ *may be chosen to be the unique positive minimum of* $|p(E|F) - \underline{\mathrm{P}}(E|H_i)|$ *attained on* $\mathrm{C}_i^+ =_{df} \{p \in \mathbb{P}_* : \mathrm{S}_p(E, H_i) \geq 1\}$, *and similarly for each radius* $\overline{\varepsilon}_i$.

The above corollary simplifies in the particularly relevant case in which the positive measurable partition $\mathscr{B}$ is finite.

Corollary 5.3 *Let* $(\Omega, \mathscr{A}, \mathbb{P}, \underline{\mathrm{P}})$ *be a lower probability space, let* $\mathscr{B} = (H_i)_{i=1}^n$ *be a finite positive measurable partition, and let* $E \in \mathscr{A}$. *Then the following are equivalent:*

(i) $\mathscr{B}$ *dilates* E;

(ii) *There is* $\varepsilon > 0$ *such that for every* $i = 1, \ldots, n$:

$$\underline{\mathbb{P}}_*(E|H_i, \varepsilon) \subseteq \mathrm{S}_*^-(E, H_i) \quad and \quad \overline{\mathbb{P}}_*(E|H_i, \varepsilon) \subseteq \mathrm{S}_*^+(E, H_i).$$

Thus, whereas Proposition 5.1 and Corollary 5.2 can ensure a positive real number ε_i for each $i \in I$, Corollary 5.3 can ensure a unique positive ε playing the role of each ε_i. Like the proof of Corollary 5.2, the proof of Corollary 5.3 is straightforward, and it may be found in the Appendix.

Proposition 5.1 and its corollaries should hardly be surprising. The correlation properties that dilation entail are rather straightforward consequences of the definition of dilation. Moreover, these correlation properties entail that each dilating partition cell and dilated event live on the surface of independence under some probability function from the closed convex hull of the set of probabilities in question. Albeit straightforward results, Proposition 5.1 and its corollaries show that by looking *beyond* the upper and lower supporting hyperplanes $\overline{\mathbb{P}}_*(E|H)$ and $\underline{\mathbb{P}}_*(E|H)$ to the upper and lower supporting *neighborhoods* $\overline{\mathbb{P}}_*(E|H, \varepsilon)$ and $\underline{\mathbb{P}}_*(E|H, \varepsilon)$, it becomes possible to characterize dilation completely in terms of positive and negative correlation. The results also show that dilation, properly understood, is a property of the *convex closure* of a set of probabilities.

Seidenfeld, Wasserman, and Herron define a function intended to measure the extent of pseudo-dilation with respect to several statistical models (see footnote 10). Thus, in the present notation and terminology, given a nonempty convex set of probabilities $\mathbb{P}$, a finite positive measurable partition $\mathscr{B} = (H_i)_{i=1}^n$ and an event $E \in \mathscr{A}$, Herron et al. (1994, 1997) define what they call the *extent of dilation*, $\Delta(E,\mathscr{B})$, by setting:

$$\Delta(E,\mathscr{B}) \quad =_{df} \quad \min_{i=1,\ldots,n}\left[\left(\underline{\mathrm{P}}(E) - \underline{\mathrm{P}}(E|H_i)\right) + \left(\overline{\mathrm{P}}(E|H_i) - \overline{\mathrm{P}}(E)\right)\right].$$

Seidenfeld, Wasserman, and Herron study how the proposed function measuring the extent to which a finite positive measurable partition $\mathscr{B}$ pseudo-dilates E may be related to a model-specific index. Of course, for dilation proper, such a function is useful insofar as dilation is known to exist with respect to a model. To be sure, if $\Delta(E,\mathscr{B}) > 0$, it does not generally follow that $\mathscr{B}$ dilates E. Although Seidenfeld, Wasserman, and Herron obtain a number of results reducing Δ to model-specific indices, while sometimes even offering necessary and sufficient conditions for dilation stated in terms other than Δ, their results in connection with Δ cannot generally be translated to results for dilation proper, as the measure they offer does not reliably measure the extent of *bona fide* dilation. This limitation, however, can be overcome by exploiting the above results.

To illustrate, let $\mathbb{P}$ be a nonempty set of probabilities. Given a positive measurable partition $\mathscr{B} = \{H_i : i \in I\}$, an event $E \in \mathscr{A}$, and $i \in I$, let $\underline{\varepsilon}_{E,i}$ and $\overline{\varepsilon}_{E,i}$ be real-valued functions defined by setting:

$$\underline{\varepsilon}_{E,i}(p) \quad =_{df} \quad |p(E|H_i) - \underline{\mathrm{P}}(E|H_i)|;$$

$$\overline{\varepsilon}_{E,i}(p) \quad =_{df} \quad |p(E|H_i) - \overline{\mathrm{P}}(E|H_i)|.$$

As above, let $C^-_{E,i} =_{df} \{p \in \mathbb{P}_* : S_p(E,H_i) \leq 1\}$ and $C^+_{E,i} =_{df} \{p \in \mathbb{P}_* : S_p(E,H_i) \geq 1\}$.

We may define the ρ^*-*maximum dilation*, ρ^*, by setting for each event $E \in \mathscr{A}$ and positive measurable partition $\mathscr{B} = \{H_i : i \in I\}$:

$$\rho^*(E,\mathscr{B}) \quad =_{df} \quad \eta^*(E,\mathscr{B}) \cdot \varepsilon(E,\mathscr{B})$$

where:

$$\eta^*(E,\mathscr{B}) \quad =_{df} \quad \sup_{i\in I}\left(\min_{p\in C^+_{E,i}} \underline{\varepsilon}_{E,i}(p) + \min_{p\in C^-_{E,i}} \overline{\varepsilon}_{E,i}(p)\right)$$

measures the η^*-*maximum extent of dilation*, while:

$$\varepsilon(E,\mathscr{B}) \quad =_{df} \quad \min_{i\in I} \max_{p\in\mathbb{P}_*} 1_{\mathrm{I}_{\mathbb{P}}(E,H_i)}(p)\left[\min_{p\in C^+_{E,i}} \underline{\varepsilon}_{E,i}(p) \cdot \min_{p\in C^-_{E,i}} \overline{\varepsilon}_{E,i}(p)\right]$$

measures the *existence of dilation*. The pair of brackets $\lceil \cdot \rceil$ denotes the ceiling function. Note the role of the surface of independence in the indicator functions $1_{\mathrm{I}(E,H_i)}$ (where $1_{\mathrm{I}(E,H_i)}(p) = 1$ if $p \in \mathrm{I}(E,H_i)$, and 0 otherwise). In addition, observe that $\rho^*(E,\mathscr{B}) = 0$ if and only if $\mathscr{B}$ does not dilate E, and $\rho^*(E,\mathscr{B}) > 0$ if and only if $\varepsilon(E,\mathscr{B}) = 1$, so positive values of ρ^* properly report the maximum extent of dilation. We may say that $\mathbb{P}$ *admits dilation* if:

$$\min_{(E,\mathscr{B}) \in \mathscr{A} \times \Pi(\mathbb{P})} \varepsilon(E,\pi) \quad > \quad 0$$

where $\Pi(\mathbb{P})$ is the collection of all positive measurable partitions of Ω with respect to $\mathbb{P}$.

Similarly, we may define the ρ_**-minimum dilation*, ρ_*, by setting for each event $E \in \mathscr{A}$ and positive measurable partition $\mathscr{B} = \{H_i : i \in I\}$:

$$\rho_*(E,\mathscr{B}) \quad =_{df} \quad \varepsilon(E,\mathscr{B}) \cdot \max\Big(\eta_*(E,\mathscr{B}), 1 + \eta_*(E,\mathscr{B})\Big)$$

where ε is defined as before and:

$$\eta_*(E,\mathscr{B}) \quad =_{df} \quad \inf_{i \in I} \left(\min_{p \in C^+_{E,i}} \underline{\varepsilon}_{E,i}(p) + \min_{p \in C^-_{E,i}} \overline{\varepsilon}_{E,i}(p) \right)$$

measures the η_**-minimum extent of dilation*. Observe that $\rho_*(E,\mathscr{B}) = 0$ if and only if $\mathscr{B}$ does not dilate E, and $\rho_*(E,\mathscr{B}) \geq 1$ if and only if $\varepsilon(E,\mathscr{B}) = 1$.

The measure Δ may also be used. Replacing the 'min'-operator in the definition Δ with the 'inf'-operator to ensure generality for partitions of infinite cardinality, we may define the Δ*-minimum dilation*, ρ_Δ, by setting for each event $E \in \mathscr{A}$ and positive measurable partition $\mathscr{B} = \{H_i : i \in I\}$:

$$\rho_\Delta(E,\mathscr{B}) \quad =_{df} \quad \varepsilon(E,\mathscr{B}) \cdot \max\Big(\Delta(E,\mathscr{B}), 1 + \Delta(E,\mathscr{B})\Big)$$

where ε is defined as before. Other measures of dilation may prove to be more useful in some respects, admitting, for example, more or less straightforward reductions to model-specific indices.

6 A Plurality of Independence Concepts

Another important point which will be made clearer in the next two examples is that the familiar logical equivalence between independence of a joint distribution as the product of its marginal distributions (SI) and independence as irrelevant information (ER) does not hold in imprecise probability models. For one thing, in the imprecise probability setting, irrelevance is not symmetric: F may be epistemically irrelevant to E without E being epistemically irrelevant to F. For another, even when E is epistemically independent of F, that itself does not guarantee that the set of probabilities factorize. In other words, within imprecise probability models, even if each event has no effect on the estimate of the other, it still may be that they fail to be stochastically independent events.

The existence of a plurality of independence concepts is a crucial difference between imprecise probability models and precise probability models, for within precise probability models we may reckon that two events are stochastically independent from observing that the probability of one event is unchanged when conditioning on the other. However, this step, from observed irrelevance of one event to the probability estimate of another to concluding that the one event is stochastically independent of the other, is fallacious within imprecise probability models. What this means is that the straightforward path to providing a behavioral justification for judgments of stochastic independence is unavailable when at least one of the outcomes has an imprecise value.[18] Even so, when the decision modeler knows that one event is irrelevant to another, there are ways to parameterize a set of probabilities to respect this knowledge which, in some cases, suffices to defuse dilation.[19]

The upshot from these two points is that there are two kinds of dilation phenomena, what we call *proper* dilation and *improper* dilation. Theorem 4.1, Theorem 4.2, and Theorem 4.4, as well as Proposition 5.1 and its corollaries, hold for both notions. Dilation occurs only if stochastic independence does not hold. However, whereas proper dilation occurs within a model which correctly parameterizes the set of distributions to reflect what is known about how the events are interrelated, if anything is known at all, improper dilation occurs within a model whose parameterization does not correctly represent what is known about how the events interact. If a decision modeler knows that one event is epistemically independent of another, for example, he knows that observing the outcome of one event does not change the estimate in another. That knowledge should constrain how a set of probabilities is parameterized, and that knowledge should override the diluting effects of dilation when updating. Rephrased in imprecise probability parlance, our results and Seidenfeld and Wasserman's results hold for a variety of natural extensions—including unknown interaction, irrelevant natural extensions, and independent natural extensions (Couso et al. 1999)—but do not discriminate between models which correctly and incorrectly encode knowledge of either epistemic irrelevance or epistemic independence. Our proposal is that irrelevant natural extensions, which correspond to a parameterized set of probabilities satisfying epistemic irrelevance, and independent natural extensions, which correspond to a parameterized set of probabilities satisfying epistemic independence, can provide principled grounds for avoiding the loss of precision by dilation that would otherwise come from updating. The upshot is that, even if the conditions for Proposition 5.1 hold, there are cases where enough is known about the relationship between the events in

[18] However, Seidenfeld et al. (2010) give an axiomatization of choice functions which can separate epistemic independence from complete stochastic independence. See (Wheeler 2009b) and (Cozman 2012) for a discussion of these results.

[19] This strategy is outlined in (Haenni et al. 2011, §9.3).

question to support a parameterization that defuses the diluting effect that dilation has upon updating.

An extension in this context is simply a parameterization of a set of probabilities, and there are several discussions of the properties of different kinds of extensions (Walley 1991, Cozman 2012, de Cooman et al. 2011, Haenni et al. 2011). Since our focus here is on independence, we will discuss three different classes of parameterizations that correspond to the analogues within imprecise probabilities of our three familiar independence concepts:

($\mathbb{ER}$) Given $\underline{\mathrm{P}}(F), \underline{\mathrm{P}}(F^c) > 0$, event F is *epistemically irrelevant* to E if and only if:

1. $\underline{\mathrm{P}}(E \mid F) \ = \ \underline{\mathrm{P}}(E \mid F^c) \ = \ \underline{\mathrm{P}}(E)$;
2. $\overline{\mathrm{P}}(E \mid F) \ = \ \overline{\mathrm{P}}(E \mid F^c) \ = \ \overline{\mathrm{P}}(E)$.

For imprecise probabilities, epistemic irrelevance is not symmetric: F may be epistemically irrelevant to E without E being epistemically irrelevant to F.

($\mathbb{EI}$) E is *epistemically independent* of F just when both F is epistemically irrelevant to E and E is epistemically irrelevant to F.

($\mathbb{SI}$) E and F are *completely stochastically independent* if and only if for all $p \in \mathbb{P}$, $p(E \cap F) = p(E)p(F)$.

We observe that ($\mathbb{SI}$) $\Rightarrow$ ($\mathbb{EI}$) $\Rightarrow$ ($\mathbb{ER}$). For suppose that ($\mathbb{SI}$) holds and that both $\underline{\mathrm{P}}(F), \underline{\mathrm{P}}(F^c) > 0$ and $\underline{\mathrm{P}}(E), \underline{\mathrm{P}}(E^c) > 0$. Then for every $p \in \mathbb{P}$, since $p(E \cap F) = p(E)p(F)$, we have $p(E|F) = p(E)$, so $\{p(E|F) : p \in \mathbb{P}\} = \{p(E) : p \in \mathbb{P}\}$ and therefore $\underline{\mathrm{P}}(E|F) = \underline{\mathrm{P}}(E)$ and $\overline{\mathrm{P}}(E|F) = \overline{\mathrm{P}}(E)$. Similarly, $\underline{\mathrm{P}}(E|F^c) = \underline{\mathrm{P}}(E)$, $\overline{\mathrm{P}}(E|F^c) = \overline{\mathrm{P}}(E)$, $\underline{\mathrm{P}}(F|E) = \underline{\mathrm{P}}(F)$, $\overline{\mathrm{P}}(F|E) = \overline{\mathrm{P}}(F)$, $\underline{\mathrm{P}}(F|E^c) = \underline{\mathrm{P}}(F)$, and $\overline{\mathrm{P}}(F|E^c) = \overline{\mathrm{P}}(F)$. Thus, ($\mathbb{SI}$) $\Rightarrow$ ($\mathbb{EI}$), and ($\mathbb{EI}$) $\Rightarrow$ ($\mathbb{ER}$) follows immediately.

However, ($\mathbb{ER}$) $\not\Rightarrow$ ($\mathbb{EI}$): although epistemic independence is symmetric, by definition, epistemic irrelevance is not symmetric. Moreover, ($\mathbb{EI}$) $\not\Rightarrow$ ($\mathbb{SI}$): two experiments may be epistemically independent even when their underlying uncertainty mechanisms are not stochastically independent. A joint set of distributions may satisfy epistemic independence without being factorizable (Cozman 2012). However, with some mild conditions, when $\mathbb{P}$ is parameterized to satisfy epistemic independence, some attractive properties hold, including associativity, marginalization and external additivity (de Cooman et al. 2011).

7 A Declaration on Independence

> We hold these truths to be self-evident, that *not* all independence concepts are equal, but are endowed with increasing logical strength, that among these are epistemic irrelevance, epistemic independence, and complete stochastic independence. To secure equivalence, a set of probabilities must be governed

> by regularity conditions instituted by its Creator, deriving just power from the consent of the epistemic agent—that whenever any governing regularity conditions become destructive of its epistemic ends, it is the right, it is the duty, of the epistemic agent to alter or to abolish them, and to institute new governing conditions, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect the agent's epistemic aims.

Probabilistic independence only appears to be a singular notion when looking through the familiar lens of a numerically precise probability function. Thus, constructing an imprecise probability model requires one to make explicit what independence concepts (if any) are invoked in a problem. Example 1, recall, states that the interaction between the tosses is unknown. Compare this example to Example 2, below, which states that the tosses are independent but does not specify *which* notion of independence is operative.

Coin Example 2: The Mystery Coin. Suppose that there are two coins rather than one.[20] Both are tossed normally, but only the first is a fair coin toss. The second coin is a mystery coin of unknown bias.

(d) $p(H_1) = 1/2$ for every $p \in \mathbb{P}$;

(e) $0 < \underline{\mathrm{P}}(H_2) \leq \overline{\mathrm{P}}(H_2) < 1$, written $p(H_2) \in (0,1)$.[21]

Since both coins are tossed normally, the tosses are performed independently. So, the lower probability of the joint event of heads is approximately zero, and the upper probability of heads is approximately one-half:

(f) $\underline{\mathrm{P}}(H_1 \cap H_2) = \underline{\mathrm{P}}(H_1)\underline{\mathrm{P}}(H_2) \approx 0$;

$\overline{\mathrm{P}}(H_1 \cap H_2) = \overline{\mathrm{P}}(H_1)\overline{\mathrm{P}}(H_2) \approx 1/2$.

Now suppose both coins are tossed and the outcomes are known to *us* but *not* to You. We then announce to You *either* that the outcomes "match," $C_1 = C_2$, *or* that they are "split," $C_1 \neq C_2$. In effect, either:

(g) We announce that "H_1 iff H_2" (M), or we announce that "H_1 iff T_2" ($\neg M$).

Since You are told that the first and the second tosses are performed independently, and initially Your estimate that the outcomes are matched is 1/2, then since for each $p \in \mathbb{P}$:

$$p(H_1) = p(H_1 \mid M)p(M) + p(H_1 \mid \neg M)p(\neg M) = 1/2, \tag{2}$$

[20] Variations of this example have been discussed by (Seidenfeld 1994, 2007), (Sturgeon 2010), (White 2010), and (Joyce 2011).

[21] The open interval (0,1) includes all real numbers in the unit interval except for 0 and 1. This means that we are excluding the possibility that the second coin is either double-headed or double-tailed. Conveniently, this also allows us to avoid complications arising from conditioning on measure zero events, although readers interested in how to condition on zero-measure events within an imprecise probability setting should see (Walley 1991, §6.10) for details.

it follows:

$$p(H_1 \mid M) = 1 - p(H_1 \mid \neg M). \tag{3}$$

Also, since for each $p \in \mathbb{P}$:

$$p(H_2) = p(H_2 \mid M)p(M) + p(H_2 \mid \neg M)p(\neg M) \in (0,1), \tag{4}$$

it follows from Equation 3:

$$\begin{aligned} p(H_2) &= p(H_2|M)p(M) + (1 - p(H_1|\neg M))p(\neg M) \\ &= 2p(M)p(H_2|M) \end{aligned}$$

Hence, for each $p \in \mathbb{P}$:

$$p(H_2 \mid M) = p(H_2). \tag{5}$$

Equation 5 implies that our announcing that the two tosses are matched is epistemically irrelevant to Your estimate of the second toss landing heads. Analogously, one might also think that announcing that the two tosses are matched is epistemically irrelevant to estimating the first toss. It may seem strange to at once maintain that our announcement is irrelevant to changing Your estimate of a coin that You know nothing about while holding that our announcement is nevertheless relevant to changing Your view about a fairly tossed coin, so one might think

$$p(H_1 \mid M) = p(H_1) \tag{6}$$

must hold as well.

However, learning the outcome of the first toss dilates Your estimate that the pair of outcomes match. After all, for all You know, the second coin could be strongly biased heads or strongly biased tails:

$$\underline{\mathrm{P}}(M|H_1) < 1/2 < \overline{\mathrm{P}}(M|H_1).$$

Yet since $p(M) = 1/2$ for each $p \in \mathbb{P}$, it follows:

$$\underline{\mathrm{P}}(M|H_1) < \underline{\mathrm{P}}(M) = 1/2 = \overline{\mathrm{P}}(M) < \overline{\mathrm{P}}(M|H_1). \tag{7}$$

and a symmetric argument holds if instead the first coin lands tails, so the first coin toss dilates your estimate that toss match.

So, although the second toss is independent of our announcement (Equation 5), and the first toss appears to be independent of our announcement (Equation 6), our announcing to You that the outcomes match is *not* independent of the first toss (Equation 7) generates a contradiction. So, what gives? ◇

The short answer is that the contradiction between Equation 6 and Equation 7 stems from equivocating over different concepts of independence. Equation 6 entails that our announcement that the tosses match is epistemically irrelevant to Your estimate that the first toss lands heads, and Equation 7 expresses that Your learning that the first toss lands heads is epistemically relevant to Your estimate of whether the tosses match. Yet Equation 6, if understood to apply to all probability functions from the set of probabilities in question, expresses complete stochastic independence, which is stronger than mere irrelevance.

In any case, it should be clear that (f) does not entail that our announcement is completely stochastically independent of the first toss. Equation (f) merely says that the coin-tossing mechanism is independent: coins are flipped the same no matter their bias. But suppose that β is the unknown bias of the second toss landing heads. Then, condition (f) may be understood to say:

$$\begin{aligned} p(H_1 \cap H_2) &= {}^1\!/\!_2\beta \\ p(H_1 \cap T_2) &= {}^1\!/\!_2(1-\beta) \\ p(T_1 \cap H_2) &= {}^1\!/\!_2\beta \\ p(T_1 \cap T_2) &= {}^1\!/\!_2(1-\beta), \end{aligned}$$

which entails $p(H_1 \mid M) = \beta$. Therefore, if $0 \leq \beta \leq 1$ is not $1/2$, our announcement that the outcomes match cannot be stochastically independent of Your estimate for the first coin landing heads.

This observation is what is behind Jim Joyce's (2011) response to Example 2, which is to reject $p(H_1 \mid M) = p(H_1)$ in Equation 6.

> There are two ways in which one event can be "uninformative about" another: the two might be stochastically independent or it might be in an "unknown interaction" situation. Regarding M and H_1 as independent in Coin Game means having a credal state $p(H_1 \mid M) = p(H_1) = p(M) = 1/2$ holds everywhere. While proponents of [The principle of Indifference] will find this congenial, friends of indeterminate probabilities will rightly protest that there is no justification for treating the events as independent (Joyce 2011, *our notation*).

According to both Joyce (2011) and Seidenfeld (1994, 2007), announcing "match" or "split" dilates your estimate of the first toss from $1/2$, and it *should* dilate Your estimate because either announcement is "highly evidentially relevant to H_1 even when you are entirely ignorant of H_2" (Joyce 2011).

Yet suppose you start with the idea that Your known chances about the first toss should not be modified by an epistemically irrelevant announcement. After all, how can You learn anything about the first toss by learning that it matches the outcome of a second toss about which You know nothing at all? Yet this commitment combined with (f) appears to restrict β to $1/2$ and rule out giving the second toss an imprecise estimate altogether. This observation is what drives Roger White (2010) to view the conflict in Example 2 to be a decisive counterexample to imprecise credal probabilities.

Joyce and White each have it half right, for there is a mathematically consistent imprecise probability model for Example 2 interpreted thus:

∘ *Idem quod* Joyce, *pace* White:

The first toss and the announcement "match" are dependent; however,

∘ *Idem quod* White, *pace* Joyce:

The announcement "match" is irrelevant to Your estimate of the first toss.

Indeed, in presenting our blended set of conditions for the mystery coin example, we are less interested in defending our proposal as *the* correct model for the mystery coin than we are in demonstrating that imprecise probability models are flexible enough to consistently encode seemingly clashing intuitions underpinning the example. Put differently, both Joyce and White make a category mistake in staking out their positions on the mystery coin case. The argument is over model fitting—not the coherence of imprecise probability models.

White commits a fallacy in reasoning by falsely concluding that epistemic irrelevance is symmetric within imprecise probability models, and by falsely supposing one event as epistemically irrelevant to another only if the two are stochastically independent. Joyce, failing to distinguish *proper* from *improper* dilation, concludes, invalidly, that he is compelled by the internal logic of imprecise probability models to maintain that the announcement dilates Your estimate of the first toss.

Blocking Dilation Through Proper Parameterization. Let us explore how to consistently represent the two claims that set Joyce and White apart in Example 2. Start with the observation that the pair of coin tosses yields four possible outcomes. A joint probability distribution may be defined in terms of those four states, namely:

$$\begin{aligned} p(H_1 \cap H_2) = \alpha_1 \quad & p(T_1 \cap T_2) = \alpha_4 \\ p(H_1 \cap T_2) = \alpha_2 \quad & p(T_1 \cap H_2) = \alpha_3. \end{aligned}$$

Given this parameterization, a set $\mathbb{P}$ of all probability measures compatible with what we know about the tosses can be represented by a unit tetrahedron (3-simplex), Figure 1 illustrates this parameterization, where maximal probabilities for each of the four possible outcomes corresponds to the four vertices,

$$\begin{aligned} \alpha_1 = 1 \quad & (1,0,0,0), \\ \alpha_2 = 1 \quad & (0,1,0,0), \\ \alpha_3 = 1 \quad & (0,0,1,0), \\ \alpha_4 = 1 \quad & (0,0,0,1). \end{aligned}$$

In a fully specified precise probability model for independent tosses, the α_i's are identical and their value is a single point within the unit tetrahedron. (If both flips are fair coin tosses, that value would be the point corresponding to $1/4$.) At the other extreme, in a completely unconstrained imprecise probability model for the set $\mathbb{P}$ of measures, the entire tetrahedron would represent the admissible values.[22] So, what You know initially about the coin tosses will translate to conditions that constrain the space of admissible probabilities, which we can visualize geometrically as regions within the unit tetrahedron, and what You learn by updating will translate to some other region within this tetrahedron. Different independence concepts translate to different ways of rendering one event irrelevant to another, but not every way of interpreting independence is consistent with the information provided by Example 2.

Now consider how the key constraints in Example 2 are represented in Figure 1.

(d) Within the unit tetrahedron there are four edges on which the constraint $1/2$ on outcome H_1 appears: the points x on the edge $\alpha_1\alpha_3$, y on the edge $\alpha_2\alpha_4$, z on the edge $\alpha_2\alpha_3$, and w on the edge $\alpha_1\alpha_4$. The omitted two edges specify either that H_1 is certain to occur or that T_1 is certain to occur, respectively. So, the hyperplane $xwyz$ represents the constraint $p(H_1) = 1/2$.

(e) The entire tetrahedron represents $p(H_2) \in [0,1]$.

(f) *i*. The upper and lower probabilities of both tosses landing heads is depicted by the shaded pentahedron, where the base of this polytope defined by the coordinates α_2, α_3, α_4 represents the lower probability $\underline{\mathrm{P}}(H_1 \cap H_2) = 0$, and the top of the polytope defined by x, w and the corresponding sharp value $1/2$ marked on the $\alpha_1\alpha_2$ edge represents the upper probability $\overline{\mathrm{P}}(H_1 \cap H_2) = 1/2$.

ii. Toss C_1 is independent of C_2,[23] so $\mathrm{I}(C_1,C_2) \neq \emptyset$ but $\mathrm{S}^+(C_1,C_2) = \mathbb{S}^-(C_1,C_2) = \emptyset$. At minimum, the outcome of the first toss is epistemically independent of the outcome of the second. This independence condition is represented by the saddle-shaped surface of independence connecting the orthogonal axes $\alpha_1\alpha_2$ and $\alpha_2\alpha_4$ in Figure 1, representing the $p \in \mathbb{P}$ such that

$$p(H_2 \mid H_1) = p(H_2) = p(H_2 \mid T_1). \tag{8}$$

Symmetrically, this is precisely the $p \in \mathbb{P}$ that connect the orthogonal axes $\alpha_1\alpha_3$ and $\alpha_2\alpha_4$ and satisfy

$$p(H_1 \mid H_2) = p(H_1) = p(H_1 \mid T_2). \tag{9}$$

Equation 8 says that the outcome of the first toss is epistemically irrelevant to Your estimate of heads occurring on the second toss, and Equation 9 says that the outcome of the second toss is epistemically irrelevant to Your estimate of heads occurring on the first toss. Taken together we have that the first toss is epistemically independent of the second toss.

[22] If $\mathbb{P}$ is closed and convex, then every point in the tetrahedron is admissible if the constraint is the closed unit interval $[0,1]$.

[23] Recall that the random variables C_1 and C_2 were introduced in Equation 1.

iii. The only region satisfying independence (*ii*), the interval constraint on the joint outcome of heads on both tosses (*i*), the sharp constraint on the first toss landing heads (d), and the interval constraint on the second toss landing heads (e), is the line segment *xy*, which rests on the surface of independence determined by Equations 8 and 9.

(g) Suppose that we announce that the outcomes match, $C_1 = C_2$. M is the edge $\alpha_1\alpha_4$. Suppose instead that we announce that the outcomes split, $C_1 \neq C_2$. $\neg M$ is the edge $\alpha_2\alpha_3$.

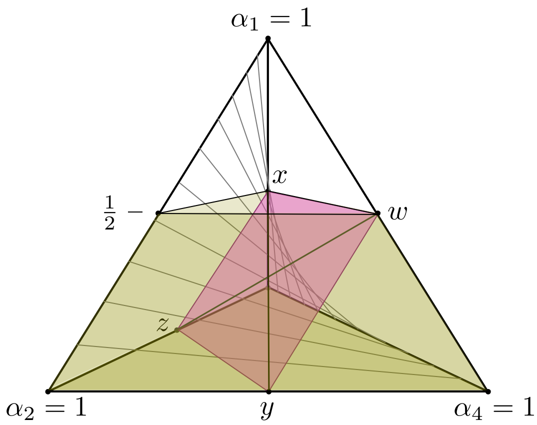

Figure 1: The constraints on Coin Example 2.

On this parameterization, the pair of coin tosses are epistemically independent and this relationship is represented by the saddle-shaped surface of independence satisfying Equations 8 and 9. Furthermore, Your initial estimate that we will announce "match" is 1/2, which is the point at the dead center of the polytope: the intersection of the lines *zw* and *xy*. This point also sits on the surface of independence.

However, the announcement "match" is the edge $\alpha_1\alpha_4$, which is entirely off of the surface of independence for the two coin tosses. The same is true if instead we announce "split", which is the edge $\alpha_2\alpha_3$. So, announcing whether the two tosses are the same or they differ is not independent of the outcome of the first toss. Here we are in agreement with the first half of Joyce's analysis.

The question now is whether this failure of independence between announcing "match" and the first toss is sufficient to ensure that this announcement is epistemically relevant to You. There is a sense in which learning whether the outcomes match or split is epistemically irrelevant to Your estimate of heads. This raises two questions. First, is there room within an imprecise probability model to accommodate this view? Second, if so, is it a rational view to maintain? Let's address the first question here and return to the second in the next section.

The *irrelevant natural extension* (Couso et al. 1999) of the marginal distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ is the the set $\mathbb{P}$ of all joint distributions p which have the form

$$p(H_1 \cap M) = p_1(M)p_2(H_1 \mid M),$$

for *some* $p_1 \in \mathbb{P}_1$ and $p_2(\cdot \mid M) \in \mathbb{P}_2$. Here the point w in Figure 1 denotes $p_2(H_1 \mid M) = 1/2$. Thus, the joint set of distributions constructed by any $p_1(M) \in \mathbb{P}_1$ will do, since all are $1/2$, but the set $\mathbb{P}_2$ is restricted to $p_2(H_1 \mid M) = 1/2$. With these selections for the marginal set of distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ and the judgment that announcing "match" is epistemically irrelevant to Your probability estimate of the first coin landing heads, the set of joint distributions $\mathbb{P}(H_1 \cap M)$ encodes that learning M is an irrelevant extension of Your estimate of $1/2$ that the first toss lands heads. Thus, we may agree with Joyce that the first toss and our announcement are probabilistically dependent but still maintain Equation 6 on the grounds that the announcement is irrelevant to the first toss. Likewise, a dual argument holds for point z where $p_2(H_1 \mid \neg M) = 1/2$.

However, learning that the first toss is heads is epistemically relevant to Your estimate of whether we will announce "match" or announce "split." After the first coin is tossed, Your estimate of M dilates from $1/2$ to $(0,1)$ because You remain completely ignorant of the bias of the second coin. This means that our method for constructing the set of joint distributions, $\mathbb{P}(M \cap H_1)$, cannot be the same method we used above to construct the irrelevant natural extension of $\mathbb{P}_1(H_1)$ and $\mathbb{P}_2(H_1 \mid M)$. Here the points w, x, y, z in Figure 1 represent $p_1(H_1) = 1/2$, but conditioning reduces to the line segment zw, since $p(M \mid H_1) \in (0,1) \neq p(M)$.

For this example we have an imprecise probability model which accommodates the fact that our announcement is not independent of the first toss. Indeed, on this model the two are neither stochastically independent nor epistemically independent. Even so, there is room to accommodate the view that our announcement is irrelevant information to Your estimate about whether the first coin toss lands heads. Moreover, due to the asymmetry of epistemic irrelevance, You can maintain that the announcement is irrelevant to the first toss even though the first toss dilates Your estimate about the announcement!

Accommodating the conflicting intuitions about Example 2 involves exploiting the distinction between epistemic independence, which is symmetric, and epistemic irrelevance, which is not. Even so, there is also a difference between stochastic independence and epistemic independence. Equations 8 and 9 together express that the coin tosses are epistemically independent, but they do not specify that the tosses are completely stochastically independent. To get at the difference between these two independence concepts, consider another example which illustrates a set of distributions that satisfies epistemic independence but fails to satisfy complete stochastic independence.

Coin Example 3: The Fat Coin. Imagine there is a fat coin that functions as a three-sided dice. The heads side is painted black, the tails side white, and the remaining fat edge is unpainted. The probability of heads is 0.3, tails 0.3, and edge 0.4. All this is known to You.

Now consider the following procedure for constructing a joint distribution for a pair of coin tosses with this coin, where the variable of interest is whether the outcome of a toss is black or white. We toss a coin, but rather than show You the outcome, we announce to You, truthfully, the outcome if the coin lands black or lands white. However, if the coin lands edge, we announce white or we announce black for reasons entirely unknown to You.

From this description You know that the outcomes of two tosses are not necessarily stochastically independent: the procedure for deciding how edges are assigned a color is unknown to You and, similar to Example 1, the bias of the second coin toss might depend on the outcome of the first toss. Nevertheless, the conditional probability of the second coin landing black is between 0.3 and 0.7, no matter the color announced for the first toss, and the conditional probability of the first coin landing black is between 0.3 and 0.7, no matter the color announced for the second toss. In other words, despite knowing that the pair of tosses may be stochastically dependent, the tosses are epistemically independent for You.

To walk through the example, given that the color outcome of the first toss is epistemically independent of the color outcome of the second, we have the following constraints

$$\begin{aligned} 0.3 \leq p_1(B_1) \leq 0.7 &\qquad 0.3 \leq p_2(B_2) \leq 0.7 \\ 0.3 \leq p_1(B_1 \mid B_2) \leq 0.7 &\qquad 0.3 \leq p_2(B_2 \mid B_1) \leq 0.7 \\ 0.3 \leq p_1(B_1 \mid W_2) \leq 0.7 &\qquad 0.3 \leq p_2(B_2 \mid W_1) \leq 0.7. \end{aligned}$$

To construct the set $\mathbb{P}$ of Your joint distributions for the two tosses $\{C_1 \in \{B,W\}, C_2 \in \{B,W\}\}$ according to the principle of epistemic independence, $\mathbb{P}$ is the largest set of joint probability distributions that are symmetrically epistemically irrelevant:

$$\begin{aligned} \mathbb{P}(C_1,C_2) = &\{p_1(C_1)p_2(C_2 \mid C_1)\} \cap \{p_2(C_1)p_1(C_2 \mid C_1)\}, \\ &\text{for } p_1 \in \mathbb{P}_1, p_2 \in \mathbb{P}_2. \end{aligned} \tag{10}$$

Your set of distributions over the possible outcomes, $\{B_1B_2, B_1W_2, W_1B_2, W_1W_2\}$, is just the closed convex set of distributions satisfying $0.3 \leq p(B_i) \leq 0.7$, for $i = 1,2$. However, there are $p' \in \mathbb{P}$ for which $p'(C_1 \in \{B,W\}, C_2 \in \{B,W\}) \neq p'(C_1 \in \{B,W\})p'(C_2 \in \{B,W\})$. For example, the constraints on the construction of $\mathbb{P}$ do not rule out $p' \in \mathbb{P}$ such that $p'(B_1 \mid B_2) \neq p'(B_1)$. This illustrates that the closed convex set of distributions constructed under epistemic independence ($\mathbb{EI}$) is a proper superset of the closed convex set of distributions constructed under stochastic independence ($\mathbb{SI}$). ◇

The fat coin example shows that we cannot ensure that two variables are stochastically independent merely from knowing that two partitioned events are epistemically independent. But, by observing that two events can be epistemically independent without being completely stochastically independent, have we merely traded a smaller problem for a much larger one? What, one may wonder, is gained by our declaration on independence?

To put this worry to rest, notice that sometimes the reason for treating outcomes as epistemically independent is because we know they result from stochastically independent mechanisms. What is left is to show how to encode this information into a model of Your epistemic state.

Consider again Example 2 but with this twist. Suppose that our announcement of either M or $\neg M$ is completely stochastically independent of the first toss and completely stochastically independent of the second toss as well. In statistical parlance, the tosses are completely stochastically independent and so are the pivotal variables, "match" and "split". Notice that on these assumptions our announcement is no longer a reliable report about the outcome of the pair of tosses. Like before, whether we announce "match" or announce "split" is irrelevant to Your estimate that the first coin toss lands heads. But unlike before, now it is true that learning how the first toss lands is irrelevant to your estimate of whether we will announce "match" or "split". Surely, You should be able to handle our irrelevant announcements *in either direction* without resorting to introducing phony precision into Your model to maintain stochastic independence.

In this revised version of Example 2 the epistemic independence between the tosses and our announcement is because we understand the underlying uncertainty mechanisms governing each toss to be completely stochastically independent of one another. To represent complete stochastic independence we consider another parameterization of the second coin toss example in Figure 2 that explicitly represents that the pair of coin tosses described in Example 2 are epistemically independent *because* they are stochastically independent, and similarly that each coin toss is stochastically independent of random announcements. Suppose that

$$\begin{aligned} \gamma &= p(H_1) = 1 - p(T_1), \\ \delta_1 &= p(T_1 \mid H_2) = 1 - p(H_1 \mid H_2), \\ \delta_2 &= p(T_1 \mid T_2) = 1 - p(H_1 \mid T_2). \end{aligned}$$

The two coin tosses are stochastically independent precisely when $\delta_1 = \delta_2$ (Haenni et al. 2011, §8.1). Visually, we see within Figure 2 that there are two surfaces of independence which are symmetric to one another, since each constraint can be mapped in two different but symmetric ways into the unit cube.[24]

$$\begin{aligned} p(H_2 \mid H_1) = \{a, d\} \quad & p(T_2 \mid T_1) = \{r, s\} \\ p(H_2 \mid T_1) = \{b, c\} \quad & p(T_2 \mid H_1) = \{q, t\}. \end{aligned}$$

One surface of independence is marked out explicitly by the hyperplane *abcd* bisecting the unit cube. Its dual, *qrst*, is the surface of independence for the second toss landing tails rather than heads.

[24]The geometric representation of these constraints in Figure 2, similar to that presented in Haenni et al. (2011), is due to Jan-Willem Romeijn. See also (Romeijn 2006).

As for the numerical constraints provided in Example 2, the first toss is a fair coin toss. So $\gamma = 1/2$, which is represented by the points x and y on the edges ab and cd, respectively. Also, by stochastic independence, we know that δ_1 and δ_2 are $1/2$, which is the line segment xy.

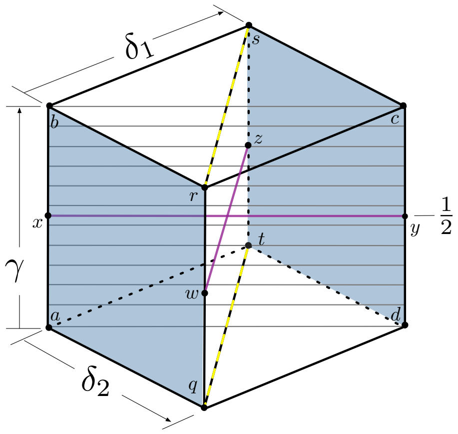

Figure 2: Independence constraints in Coin example 2.

Now suppose we announce to You either (M) or ($\neg M$). Announcing "match" corresponds to telling You that either the line segment ad (both heads) or the line segment rs (both tails). Symmetrically, announcing "split" corresponds to telling You that either the line segment bc constrains Your estimate of H_1 or the line segment qt does.

Clearly the announcements "match" and "split" form a measurable partition, $\mathscr{B}$. Since our focus here is the effect on Your estimate of H_1 from our announcement, the issue is whether our announcement dilates Your estimate of H_1.

Because this model explicitly ensures that the tosses are completely stochastically independent of our announcements, we see that conditioning on our announcement only yields probabilities within the set $\mathbb{P}$ that are on the surface of independence. As a result, You know that the tosses are epistemically independent: the outcome from one toss is irrelevant to estimating the outcome of the other. Similarly, You know that there is no effect on Your estimate of H_1 by us announcing "match" or announcing"split." In addition, You also know that Your estimate of M is not affected by knowing that the first toss is heads.

8 Is Dilation Reasonable?

The motor driving Example 1 and Example 2 is uncertainty that arises from constructing a joint probability estimate for two coin flips, but the examples differ over the source of that uncertainty. In the first example, the origin is an unknown interaction between the pair of tosses. In this case You know that the coin is fair but

You do not know how the second toss is performed and You cannot rule out that the outcome of the second toss depends on the outcome of the first. It is therefore reasonable for You to dilate in this case.

The origin of imprecision in the second example is an unknown bias of the second coin. In this case You know the tosses are performed independently, but the conditions under which we announce the outcome of the fair toss partitions the space and strictly dilates Your estimate of whether the outcomes of *both* tosses match or split. This additional condition appears to make the tosses epistemically relevant to one another after all, thereby creating the illusion that You do not know how to estimate the probability of heads occurring on the first toss after we announce to You whether the outcomes match. This illusion is helped along by mistakenly assuming that probabilistic dependence ensures epistemic relevance. But, epistemic irrelevance is not symmetric within imprecise probability models, and in this case our announcement is epistemically irrelevant to Your estimate of the first toss but the outcome of the first toss is epistemically relevant to Your estimate of whether we will announce "match." This error is compounded by mistakenly assuming that epistemic independence is sufficient for stochastic independence, creating the illusion of a mysterious interaction between events which are mistakenly thought to be stochastically independent.

The third example illustrates how knowing that two events are dependent is nevertheless insufficient to ensure that there will be information for You to glean about either event from observing the other. Furthermore, if one drops convexity, it can become impossible to distinguish between this case and one where the two events are completely stochastically independent.

Putting this together, we see that the first example is a clear and uncontroversial case of dilation, but the second example is a mixed bag. In Example 2 our announcement of how the first toss turned out is relevant information to Your estimate of whether both outcomes match or differ, so this announcement would and should dilate Your estimate of whether both outcomes match. However, the announcement of whether the outcomes match is irrelevant to Your original estimate *even though our announcement and the first toss are correlated*. My announcement about how both coins landed is epistemically irrelevant to Your estimate of heads on the first toss. Hence, our announcement *should not* dilate Your original estimate.

Finally, if one takes seriously the claim that the two events are completely stochastically independent and models this correctly, then there is zero correlation between the outcome of the tosses and our announcement. But then there is also no dilation and no controversy, since Your initial estimate will not be affected by our independent announcement.

The last section provided an explanation for how to give a coherent imprecise probability model of Example 2, which reconciles the thought that the first toss and our announcement that the outcomes match are dependent but that our announcement that they match is epistemically irrelevant to Your estimate of the first toss. It is one thing to find a way to model these conflicting intuitions, quite another to determine whether it is rational to adopt such a model. We will focus our remaining remarks on the most controversial assumption of this model, at least from an imprecise probability point of view, which is the endorsement of Equation 6.

Imagine that You maintain that $p(H_1 \mid M) = p(H_1) = p(M) = 1/2$ and are willing to maintain that the probability of the first toss landing heads is $1/2$ even after we truthfully report to you that the outcomes of the first and second toss match. Suppose that You are willing to use this estimate to post betting odds at a €1.00 stake for an unlimited number of trials. The question is this: Can You lose money from refusing to change Your degree of belief when You learn that the coins match?

Here is the setup.[25] You have a fair coin, we have a coin of unknown bias, and You have some strategy for guessing heads or tails for Your coin on the nth flip of an unbounded sequence of trials. We pay You €1.00 if Your guess is right, You pay us €1.00 if Your guess is wrong. Payoffs are only made if the two coins are both heads or both tails. The procedure in Example 2 is followed, but in addition we make the following bets: On each trial, You flip Your coin but do not announce Your guess and do not see the outcome. We flip our coin, which has a fixed bias β for all trials. We observe the outcome of Your toss and our toss, then we announce to You whether they match. If we announce that the two tosses match, You flip another fair coin to decide whether to announce that the matched outcomes are both heads or both tails. Note that You perform this secondary fair coin toss because, by hypothesis, Your degree of belief that Your coin lands heads is $1/2$ and is unchanged by our announcement that the coins match. Without Your knowledge, suppose we pick $\beta < 1/2$. Then our expected profit on each trial is zero: $1/2(1/2 - \beta) + 1/2(\beta - 1/2) = 0$.

If, before we pick the bias β of the coin, we know Your betting strategy, and it is other than flipping Your secondary coin, then we can bankrupt You by strategically choosing β. If you know our choice for β, then You can alter your betting strategy to bankrupt us. But if You choose heads with a long-run frequency equal to Your undilated degree of belief of $1/2$ in the first coin landing heads, then every strategy has zero expected loss, regardless of the bias of the second coin. So the additional information about equivalent outcomes, while probabilistically associated with the first toss, is nevertheless both epistemically irrelevant and practically irrelevant.

Compare this to Example 1, where β is not fixed through each trial. You maintain that $p(H_2 \mid H_1) \neq p(H_2)$ and are unwilling to maintain that the probability of the second toss landing heads is $1/2$ after we faithfully report to you that the first toss is heads.

In this case, You have a fair coin and we have a fair coin, but only Yours is tossed fairly. Just as before, we pay You €1.00 if Your guess is right, You pay us €1.00 if Your guess is wrong. Payoffs are only made if Your coin lands heads. The procedure in Example 1 is followed, but in addition we make the following bets: On each trial, You flip Your coin but do not announce Your guess and do not see the outcome. After viewing the outcome of Your toss, we either arrange our coin

[25] Thanks to Clark Glymour for putting this argument to us.

to match the outcome of Your toss, or we arrange our coin to ensure the pair are different. You do not know which. Before each toss, Your degree of belief that the outcome of Your toss is heads is $1/2$, the same as Your degree of belief that the outcome of our "toss" is heads. However, after we announce that Your coin landed heads, Your probability that our coin is heads dilates to $[0, 1]$. Given this evidence, Your expectation on each trial is *either* $-€1.00$ *or* $€1.00$. Within an imprecise probability model this translates to Your willingness to post odds of $1/2$ on the second coin landing heads prior to observing the first coin, but an unwillingness to bet on the outcome of the second toss at any odds after learning the first toss land heads.

Like Example 2, You might nevertheless choose to treat the outcome of the first toss as practically irrelevant. Then your expected loss would be zero. Unlike Example 2, there is not a shred of evidence for doing so, since the first toss is epistemically relevant to Your estimate of the second toss. Here practical relevance and epistemic relevance come apart.

9 Conclusion

The key to sorting genuine dilation examples from bogus examples is to focus on the interaction between the *dilator* and the *dilatee*. The most sensational examples of dilation are engineered to try to show that events which have nothing to do with each other can nevertheless have mysterious effects on one or the other's probability estimates. But the sizzle in these examples invariably fizzles for one of three reasons. One type of error is to equivocate over whether *dilator* and *dilatee* are completely stochastically independent. If the events are completely stochastically independent, then the examples cannot be true examples of dilation. Conversely, where there is dilation, there is the possibility of an interaction between the events of some kind or another which may or may not appear to be epistemically relevant to Your probability estimates. A second error concerns whether the agent's credal state—the parameterized set of probabilities—correctly encodes what the agent knows about how the events are related to one another. Proper dilation occurs within a correctly parameterized model, whereas improper dilation occurs within a model that fails to correctly encode what is known about the problem. A third error concerns a distinction between practical relevance and epistemic relevance. In cases where interactions are epistemically relevant, they may or may not be practically relevant to rational decision making. This can depend on whether it is more valuable to investigate the matter further before taking an action, or whether the agent must take a decision on the best evidence available. Although dilation degrades estimates, there is nevertheless information from proper dilation about a possible dependency in the decision maker's model. Having this information and knowing when this possibility is "activated" can be useful to the decision maker, which counts against a blanket policy that either expunges or embraces dilation wholesale.

So, to sum up, cases of proper dilation are far less mysterious than critics contend because they only arise when your probability model allows for the possibility for an unknown interaction between events and the effect of that interaction can be a source of information. So, the message to conservative critics is that they should not shoot the messenger: proper dilation merely reports to you a warning about the live consequences of a possible, problematic interaction within your model. The message to progressive critics is that there is more that goes into strategic updating than simply knowing Your probabilities and Your preferences: knowing what the agent does not know about the interaction between events is also an important consideration in deciding what his probability estimates should be in light of new evidence. And to friends of indeterminate probabilities, much has been learned about the properties of natural extensions in the last few years, especially independent natural extensions. The nature of the controversy about dilation is not existential; it is a run-of-the-mill model selection problem.

In any event, reckoning with dilation, while not a Good idea, is not a bad idea, either.

Acknowledgements Thanks to Horacio Arló-Costa, Jim Joyce, Isaac Levi, Teddy Seidenfeld, and Jon Williamson for their comments on early drafts, and to Clark Glymour and Choh Man Teng for a long discussion on dilation that began one sunny afternoon in a Lisbon café. This research was supported in part by award (LogICCC/0001/2007) from the European Science Foundation.

Appendix

We now return to our discussion of the technical machinery for the general case from Section 5. In the general setting, where infinitely many events inhabit the algebra $\mathscr{A}$, closure with respect to the total variation norm topology (i.e., *strong topology*) or with respect to the *weak topology* (distinct from the weak*-topology) introduces more closed sets than the weak*-topology. The former topologies are stronger and indeed *too* strong for the purposes of imprecise probabilities, which demand a very weak topology, the weak*-topology. Hence, the discussion in (Seidenfeld and Wasserman 1993, pp. 1141-1443), which employs the total variation norm, suits sets of probabilities over an algebra consisting of finitely many events.

A set of probabilities in this general setting is always norm-bounded (with respect to the norm-dual), so a weak*-closed set of probabilities is weak*-compact. Accordingly, every weak*-continuous functional on the dual space achieves its minimum at the extreme points of a closed convex (i.e., a weak*-closed and convex) set of probabilities, and since all and only evaluation functionals are weak*-continuous, the lower probability of an event is the minimum number assigned to the event by all probability functions in the set, where as in the finite setting a probability function from the collection of extreme points of the set witnesses the lower probability of the event.

Much as in the finite case, any compact convex set of probabilities is the closed convex hull of its extreme points (i.e, any convex weak*-compact set of probabilities is the weak*-closed convex hull of its extreme points). However, while in the finite setting a compact convex set of probabilities is identified with the convex hull of its extreme points, in the general setting a compact convex set of probabilities is identified with the *closed* convex hull of its extreme points. In both the finite and general setting, the convex hull of the set of extreme points is dense in the compact convex set in question, but in the general setting, a compact convex set may properly contain the convex hull of its extreme points and so the hull must in addition be closed.

Proof of **Proposition** 4.3. Let $\mathbb{P}(\mathscr{A})$ be the set of all probability functions on $\mathscr{A}$, and let $\mathbb{D} =_{df} \{p \in \mathbb{P}(\mathscr{A}) : p(E) \geq \underline{\mathrm{P}}(E) \text{ for all } E \in \mathscr{A}\}$. Observe that $\mathbb{P} \subseteq \mathbb{D}$ and that $\mathbb{D}$ is convex and weak*-closed and so weak*-compact. In addition, since $\mathrm{co}(\mathbb{P}) \subseteq \mathbb{D}$, it follows that the $\overline{\mathrm{co}}(\mathbb{P}) \subseteq \mathbb{D}$, a weak*-compact set, whence $\underline{\mathrm{P}}(E) = \inf\{p(E) : p \in \mathbb{P}\} = \min\{p(E) : p \in \overline{\mathrm{co}}(\mathbb{P})\}$, the minimum being achieved at an extreme point of $\overline{\mathrm{co}}(\mathbb{P})$.

Now given $\underline{\mathrm{P}}(H) > 0$, let $\mathbb{D}[H]$ be defined by:

$$\mathbb{D}[H] \quad =_{df} \quad \{p \in \mathbb{P}(\mathscr{A}) : p(E \cap H) \geq \underline{\mathrm{P}}(E|H)p(H) \text{ for all } E \in \mathscr{A}\}.$$

As before, observe that $\mathbb{P} \subseteq \mathbb{D}[H]$ and that $\mathbb{D}[H]$ is convex and weak*-closed and so weak*-compact. Hence, $\overline{\mathrm{co}}(\mathbb{P}) \subseteq \mathbb{D}[H]$. Since $\underline{\mathrm{P}}(H) > 0$, by the first part it follows that $\overline{\mathrm{co}}(\mathbb{P}) \subseteq \{p \in \mathbb{P}(\mathscr{A}) : p(H) > 0\}$. Since with respect to the weak* topology on the dual space every evaluation functional f^* is a real-valued continuous linear functional on the dual space with $f^*(p) = p(f)$ for each p from the dual space, it follows that $\frac{(E\cap H)^*}{H^*}$ is a continuous explicitly quasiconcave function of p on $\overline{\mathrm{co}}(\mathbb{P})$, so it attains a minimum at an extreme point of $\overline{\mathrm{co}}(\mathbb{P})$ (thus, $\min\{\frac{(E\cap H)^*(p)}{H^*(p)} : p \in \overline{\mathrm{co}}(\mathbb{P})\} = \min\{\frac{p(E\cap H)}{p(H)} : p \in \overline{\mathrm{co}}(\mathbb{P})\}$ exists and is an extreme point of $\overline{\mathrm{co}}(\mathbb{P})$), whence $\underline{\mathrm{P}}(E|H) = \inf\{p(E|H) : p \in \mathbb{P}\} = \min\{p(E|H) : p \in \overline{\mathrm{co}}(\mathbb{P})\}$, as desired. □

Proof of **Proposition** 5.1. We first show that (i)$\Longleftrightarrow$(iii), and we then show that (i)$\Longleftrightarrow$(ii).

(i)$\Rightarrow$(iii) Suppose that $\mathscr{B}$ dilates E. Then for each $i \in I$, $\underline{\mathrm{P}}(E|H_i) < \underline{\mathrm{P}}(E) \leq \overline{\mathrm{P}}(E) < \overline{\mathrm{P}}(E|H_i)$. For each $i \in I$, consider the real-valued function $\underline{\varepsilon}_i(p) =_{df} |p(E|H_i) - \underline{\mathrm{P}}(E|H_i)|$ and the real-valued function $\mathrm{S}_{p,i}(E,H_i)$. We recall that the weak* topology on the dual space is a locally convex topological vector space with respect to which every evaluation functional f^* is a real-valued continuous linear functional on the dual space. It follows that $\underline{\varepsilon}_i(p)$ and $\mathrm{S}_{p,i}(E,H_i)$ are continuous functions of p on $\mathbb{P}$ for each $i \in I$.

Now let $i \in I$. By hypothesis, there is $p_1 \in \mathbb{P}$ such that $\mathrm{S}_{p_1,i}(E,H_i) > 1$, so importantly, C_i^+ is nonempty. Then since $C_i^+ =_{df} \{p \in \mathbb{P} : \mathrm{S}_p(E,H_i) \geq 1\}$ is a weak*-closed and so weak*-compact set, it follows that $\underline{\varepsilon}_i$ achieves a minimum value on C_i^+ (and the set of minimizers of $\underline{\varepsilon}_i$ is also compact). Choosing a minimizer $p_i \in \mathbb{P}$ of $\underline{\varepsilon}_i$, we see that for every $p \in \mathbb{P}$, if $|p(E|H_i) - \underline{\mathrm{P}}(E|H_i)| < \underline{\varepsilon}_i(p_i) = |p_i(E|H_i) - \underline{\mathrm{P}}(E|H_i)|$, then $\mathrm{S}_p(E,H_i) < 1$. We have accordingly shown that $\underline{\mathbb{P}}(E|H_i,\underline{\varepsilon}_i(p_i)) \subseteq \mathrm{S}^-_{\mathbb{P}}(E,H_i)$. Of course, we may suppress reference to the minimizer p_i in $\underline{\varepsilon}_i(p_i)$. The other inclusion $\overline{\mathbb{P}}(E|H_i,\overline{\varepsilon}_i) \subseteq \mathrm{S}^+_{\mathbb{P}}(E,H_i)$ is established by a similar argument.

(iii) $\Leftarrow$ (i) Suppose that $\mathscr{B}$ does not dilate E. Then there is $i \in I$ such that $\underline{\mathrm{P}}(E|H_i) \geq \underline{\mathrm{P}}(E)$ or $\overline{\mathrm{P}}(E) \geq \overline{\mathrm{P}}(E|H_i)$. We may assume without loss of generality that $\underline{\mathrm{P}}(E|H_i) \geq \underline{\mathrm{P}}(E)$ for some $i \in I$. First, if $\underline{\mathrm{P}}(E) \leq \overline{\mathrm{P}}(E) \leq \underline{\mathrm{P}}(E|H_i)$, then choosing a minimizer $p \in \mathbb{P}$ of $\underline{\mathrm{P}}(E|H_i)$, we see that $\mathrm{S}_p(E,H_i) \geq 1$. Second, if $\underline{\mathrm{P}}(E) < \underline{\mathrm{P}}(E|H_i) < \overline{\mathrm{P}}(E)$, then for every $\varepsilon > 0$ we can find a convex combination $p \in \mathbb{P}$ of $p_0, p_1 \in \mathbb{P}$ assigning a probability to E within ε-distance below $\underline{\mathrm{P}}(E|H_i)$, where $\underline{\mathrm{P}}(E) \leq p_0(E) < \underline{\mathrm{P}}(E|H_i) < p_1(E) \leq \overline{\mathrm{P}}(E)$, so $\mathrm{S}_p(E,H_i) > 1$. Third, if $\underline{\mathrm{P}}(E) = \underline{\mathrm{P}}(E|H_i) < \overline{\mathrm{P}}(E)$, then choosing a minimizer $p \in \mathbb{P}$ of $\underline{\mathrm{P}}(E)$, we see that $\mathrm{S}_p(E,H_i) \geq 1$. Evidently, the conditions of the main claim cannot be jointly satisfied.

(i) $\Leftrightarrow$ (ii) On the one hand, suppose that (i) obtains. Then since (iii) accordingly obtains, define $(\varepsilon_i)_{i\in I}$ by setting $\varepsilon_i =_{df} \min(\underline{\varepsilon}_i, \overline{\varepsilon}_i)$ for each $i \in I$. Clearly the inclusions still obtain for the ε_i. On the other hand, if (ii) obtains, obviously by setting $\underline{\varepsilon}_i =_{df} \varepsilon_i$ and $\overline{\varepsilon}_i =_{df} \varepsilon_i$ for each $i \in I$, condition (iii) obtains and so (i) obtains. □

Proof of **Corollary** 5.2. Only (i)$\Longleftrightarrow$(iii) requires proof. On the one hand, suppose that $\mathscr{B}$ dilates E. Then for each $i \in I$, $\underline{\mathrm{P}}(E|H_i) < \underline{\mathrm{P}}(E) \leq \overline{\mathrm{P}}(E) < \overline{\mathrm{P}}(E|H_i)$. By Proposition 4.3 we have $\underline{\mathrm{P}}(A) = \min\{p(A) : p \in \mathbb{P}_*\}$, $\underline{\mathrm{P}}(A|B) = \min\{p(A|B) : p \in \mathbb{P}_*\}$, $\overline{\mathrm{P}}(A) = \max\{p(A) : p \in \mathbb{P}_*\}$, and $\overline{\mathrm{P}}(A|B) = \max\{p(A|B) : p \in \mathbb{P}_*\}$ for every $A, B \in \mathscr{A}$ with $\underline{\mathrm{P}}(B) > 0$, so $\mathscr{B}$ dilates E with respect to $\mathbb{P}_*$. It follows from Proposition 5.1 that there are positive $(\underline{\varepsilon}_i, \overline{\varepsilon}_i)_{i\in I}$ in $\mathbb{R}$ such that $\underline{\mathbb{P}}_*(E|H_i,\underline{\varepsilon}_i) \subseteq \mathrm{S}^-_*(E,H_i)$ and $\overline{\mathbb{P}}_*(E|H_i,\overline{\varepsilon}_i) \subseteq \mathrm{S}^+_*(E,H_i)$ for every $i \in I$.

On the other hand, suppose that there are positive $(\underline{\varepsilon}_i, \overline{\varepsilon}_i)_{i\in I}$ in $\mathbb{R}$ such that for every $i \in I$, $\underline{\mathbb{P}}_*(E|H_i,\underline{\varepsilon}_i) \subseteq \mathrm{S}^-_*(E,F)$ and $\overline{\mathbb{P}}_*(E|H_i,\overline{\varepsilon}_i) \subseteq \mathrm{S}^+_*(E,H_i)$. Then by Proposition 5.1, $\mathscr{B}$ dilates E with respect to $\mathbb{P}_*$, so for every event $i \in I$, $\underline{\mathrm{P}}(E|H_i) < \underline{\mathrm{P}}(E) \leq \overline{\mathrm{P}}(E) < \overline{\mathrm{P}}(E|H_i)$, whence again by Proposition 4.3 it follows that $\mathscr{B}$ dilates E with respect to $\mathbb{P}$, as desired.

Clearly, that the radii $\underline{\varepsilon}_i$ and $\overline{\varepsilon}_i$ may be chosen in the way described follows from Proposition 5.1. The other implications are trivial consequences of what we have just shown. □

Proof of **Proposition** 5.3. On the one hand, if $\mathscr{B}$ strictly dilates E, then by Corollary 5.2 there are $(\underline{\varepsilon}_i)_{i\in I} \in \mathbb{R}^I_+$ and $(\overline{\varepsilon}_i)_{i\in I} \in \mathbb{R}^I_+$ such that for every $i \in I$, $\underline{\mathbb{P}}_*(E|H_i,\underline{\varepsilon}_i) \subseteq \mathrm{S}^-_*(E,H_i)$ and $\overline{\mathbb{P}}_*(E|H_i,\overline{\varepsilon}_i) \subseteq \mathrm{S}^+_*(E,H_i)$. Let

$$\varepsilon \quad =_{df} \quad \min\big\{\delta : \delta = \underline{\varepsilon}_i \text{ or } \delta = \overline{\varepsilon}_i \text{ for some } i \in I\big\}.$$

Then $\varepsilon > 0$, and clearly the inclusions still obtain. On the other hand, if (ii) obtains, then part (ii) of Corollary 5.2 obtains, so $\mathscr{B}$ strictly dilates E. $\square$

References

Couso, I., Moral, S., and Walley, P. (1999). Examples of independence for imprecise probabilities. In de Cooman, G., editor, *Proceedings of the First Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Ghent, Belgium.

Cozman, F. (2000). Credal networks. *Artificial Intelligence*, 120(2):199–233.

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

de Cooman, G. and Miranda, E. (2007). Symmetry of models versus models of symmetry. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, pages 67–149. King's College Publications, London.

de Cooman, G. and Miranda, E. (2009). Forward irrelevance. *Journal of Statistical Planning*, 139:256–276.

de Cooman, G., Miranda, E., and Zaffalon, M. (2011). Independent natural extension. *Artificial Intelligence*, 175:1911–1950.

de Finetti, B. (1974a). *Theory of Probability*, volume I. Wiley, 1990 edition.

de Finetti, B. (1974b). *Theory of Probability*, volume II. Wiley, 1990 edition.

Elga, A. (2010). Subjective probabilities should be sharp. *Philosophers Imprint*, 10(5).

Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics*, 75:643–69.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B*, 14(1):107–114.

Good, I. J. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science*, 17(4):319–321.

Good, I. J. (1974). A little learning can be dangerous. *The British Journal for the Philosophy of Science*, 25(4):340–342.

Grünwald, P. and Halpern, J. Y. (2004). When ignorance is bliss. In Halpern, J. Y., editor, *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI '04)*, pages 226–234, Arlington, Virginia. AUAI Press.

Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2011). *Probabilistic Logics and Probabilistic Networks*. Synthese Library. Springer, Dordrecht.

Halmos, P. R. (1950). *Measure Theory*. Van Nostrand Reinhold Company, New York.

Harper, W. L. (1982). Kyburg on direct inference. In Bogdan, R., editor, *Henry E. Kyburg and Isaac Levi*, pages 97–128. Kluwer Academic Publishers.

Herron, T., Seidenfeld, T., and Wasserman, L. (1994). The extent of dilation of sets of probabilities and the asymptotics of robust bayesian inference. In *PSA 1994 Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 1, pages 250–259.

Herron, T., Seidenfeld, T., and Wasserman, L. (1997). Divisive conditioning: further results on dilation. *Philosophy of Science*, 64:411–444.

Horn, A. and Tarski, A. (1948). Measures in boolean algebras. *Transactions of the AMS*, 64(1):467–497.

Joyce, J. (2011). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, 24(1):281–323.

Koopman, B. O. (1940). The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41(2):269–292.

Kyburg, Jr., H. E. (1961). *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, CT.

Kyburg, Jr., H. E. (1974). *The Logical Foundations of Statistical Inference*. D. Reidel, Dordrecht.

Kyburg, Jr., H. E. (2007). Bayesian inference with evidential probability. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, pages 281–296. King's College, London.

Kyburg, Jr., H. E. and Teng, C. M. (2001). *Uncertain Inference*. Cambridge University Press, Cambridge.

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

Levi, I. (1977). Direct inference. *Journal of Philosophy*, 74:5–29.

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

Rao, K. B. and Rao, M. B. (1983). *Theory of Charges: A Study of Finitely Additive Measures*. Academic Press.

Romeijn, J.-W. (2006). Analogical predictions for explicit similarity. *Erkenntnis*, 64:253–80.

Savage, L. J. (1972). *Foundations of Statistics*. Dover, New York.

Schlosshauer, M. and Wheeler, G. (2011). Focused correlation, confirmation, and the jigsaw puzzle of variable evidence. *Philosophy of Science*, 78(3):276–92.

Seidenfeld, T. (1994). When normal and extensive form decisions differ. In Prawitz, D., Skyrms, B., and Westerstahl, D., editors, *Logic, Methodology and Philosophy of Science*. Elsevier Science B. V.

Seidenfeld, T. (2007). Forbidden fruit: When Epistemic Probability may not take a bite of the Bayesian apple. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* King's College Publications, London.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2010). Coherent choice functions under uncertainty. *Synthese*, 172(1):157–176.

Seidenfeld, T. and Wasserman, L. (1993). Dilation for sets of probabilities. *The*

Annals of Statistics, 21:1139–154.

Shogenji, T. (1999). Is coherence truth conducive? *Analysis*, 59:338–45.

Smith, C. A. B. (1961). Consistency in statistical inference (with discussion). *Journal of the Royal Statistical Society*, 23:1–37.

Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, 42(1):139–65.

Sturgeon, S. (2010). Confidence and coarse-grain attitudes. In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, volume 3, pages 126–149. Oxford University Press.

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

Wayne, A. (1995). Bayesianism and diverse evidence. *Philosophy of Science*, 62(1):111–121.

Wheeler, G. (2006). Rational acceptance and conjunctive/disjunctive absorption. *Journal of Logic, Language and Information*, 15(1-2):49–63.

Wheeler, G. (2009a). Focused correlation and confirmation. *The British Journal for the Philosophy of Science*, 60(1):79–100.

Wheeler, G. (2009b). A good year for imprecise probability. In Hendricks, V. F., editor, *PHIBOOK*. VIP / Automatic Press.

Wheeler, G. (2012). Objective Bayesianism and the problem of non-convex evidence. *The British Journal for the Philosophy of Science*, 63(3):841–50.

Wheeler, G. (2013). Character matching and the envelope of belief. In Lihoreau, F. and Rebuschi, M., editors, *Epistemology, Context, and Formalism*, Synthese Library, pages 185–94. Springer. Presented at the 2010 APA Pacific Division Meeting.

Wheeler, G. and Scheines, R. (2013). Coherence, confirmation, and causation. *Mind*, 122(435):135–70.

White, R. (2010). Evidential symmetry and mushy credence. In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, volume 3, pages 161–186. Oxford University Press.

Williamson, J. (2007). Motivating objective Bayesianism: From empirical constraints to objective probabilities. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* College Publications, London.

Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford University Press.

Yule, U. (1911). *Introduction to the Theory of Statistics*. Griffin, London.