Demystifying Dilation

Gregory Wheeler
CENTRIA, Center for Artificial Intelligence
Universidade Nova de Lisboa

Abstract

Dilation occurs when upper and lower probability estimates of some event E are properly included in the upper and lower probability estimates of the probability of E conditional on another event F, resulting in a change from a more precise estimate of E to a less precise estimate of E upon learning F. Strict dilation occurs when E is dilated by every event in a partition, which means that sometimes E becomes less precise no matter how an experiment turns out. Many think that strict dilation is a pathological feature of imprecise probability models, while others have thought the problem is with Bayesian updating. However, a point often overlooked in critical discussions of dilation is that knowing that E is stochastically independent of F (for all F in a partition) is sufficient to avoid strict dilation. Since the most sensational alleged dilation examples are those which play up independence between *dilator* and *dilatee*, the sensationalism traces to mishandling imprecise probabilities rather than revealing a genuine puzzle about imprecise probabilities.

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 available evidence when estimating a probability.

Another Good idea, but not only 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 some may say, but instead only yield upper and lower constraints on probability

[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), and Henry Kyburg, Jr. (Kyburg 1961). 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).

estimates, or imprecise *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. Now, 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? If so, is it rational for you to pay someone to *not* tell you?

Critics find 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 imprecise probabilities. Conditioning on new information should reduce your imprecision, tradition tells us, unless the information is irrelevant, in which case we should expect there to be no change in your original position. However, strict dilation describes a case where the specific outcome of the experiment is irrelevant but imprecision is increased by conditioning, come what may.

The false allure of imprecise probabilities, conservatives lament, is a tale of trampled traditions and disregarded distinctions. Simply observing the difference between *objective* and *subjective* probabilities (White 2010, p. 163) or between *knowledge* and *belief* (Williamson 2007, pp. 176-7), and sticking to tried and true methods, like Laplace's *principle of indifference* (White 2010) or the *principle of maximum entropy* (Williamson 2010), these critics maintain, would avoid the whole hullabaloo. Even those who think that belief states should admit imprecision to "match the character" of the evidence despair of imprecise probability theory ever being of service to epistemology (Sturgeon 2008, 2010).

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 entirely by always selecting the most precise estimate available. There is no possibility for strict 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). However, the general idea of selective updating is not necessarily incompatable with orthodox Bayesian methods (Harper 1982). Indeed, one recent proposal to avoid dilation is to replace Good's first principle by a second-order principle to determine whether or not it pays you (in expectation) to update a particular hypothesis on a particular item of evidence (Grünwald and

[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) is the first published record. Seidenfeld and Wasserman's study is further developed in (Herron et al. 1994) and (Herron et al. 1997).

Halpern 2004).

One point often missed by dilation's detractors, conservatives and progressives alike, is that strict dilation requires that your evidence about the events in question leave open the possibility for an unknown interaction between the events (Seidenfeld and Wasserman 1993, Theorems 2.1 to 2.3). This is a key point, for the most sensational alleged cases of dilation—recent examples include (Sturgeon 2010, White 2010, Joyce 2011)—equivocate on whether the events in question are stochastically independent. If the events are completely stochastically independent of one another, then alleged cases where one event mysteriously dilates the estimate of the other are instances of mishandled imprecise probabilities rather than examples of dilation. Genuine cases of dilation are far less mysterious than critics contend since they arise when your probability model allows for the possibility for an unknown interaction between events. To conservatives, the message is *don't shoot the messenger*. Genuine dilation merely reports to you a warning about possible problematic interactions within your model. To progressives, the message is *there is more that goes into strategic updating than knowing your probabilities and your preferences*. Knowing what you do not know about the interaction between events is also an important consideration in deciding what your probability estimates should be in light of new evidence.

All of this is to say that dilation is information to reckon with, not defect of imprecise probabilities to be avoided at all costs.

Any discussion of dilation invariable includes a discussion of coin flips. The canonical example is due to Peter Walley (Walley 1991, pp. 298-9), which is reproduced as Example 1 in Section 3. A variation of Walley's example is presented in Example 2. As will be made clear, Example 2 is not a straightforward case of dilation, and the reason why boils down to a difference between precise and imprecise probability models over how probabilistic independence is handled. Within precise probability models independence, whether the concept is viewed as the factorization of some joint distribution by the product of its marginal distributions or is viewed as the outcome of one variable having no effect on the probability estimates of another, is a unitary notion. One can move freely between the two concepts, so long as some accommodation is made for conditioning on zero probability events. But within the imprecise probability setting there are several independence concepts. In this essay we will focus on three independence concepts. The upshot is that a sound inference from facts about independence within a precise probability model can turn out fallacious within an imprecise probability setting. Specifically, within an imprecise probability setting, observing that one event is irrelevant to another does not ensure that the two events are completely stochastically independent. What's more, irrelevance is not symmetric: observing that one event is irrelevant to the imprecise estimate of another does not ensure that the latter is irrelevant to the former. The relationship between these independence concepts is explored in Examples 2 and 3.

The aim of this essay is to defend the theory of imprecise probability by demystifying dilation. My contention is that distinguishing genuine dilation from counterfeit examples is crucial for recognizing the phenomenon, when it occurs, as a potential source of relevant information about your imprecise epistemic position. But for these points to fall into place, we will need to look at what a coin toss is and how genuine and bogus dilation examples alike manipulate those conditions.

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 $1/2$ of the outcome heads occurring for each toss.

In this section I will explain each piece of this mathematical model. In the next I will discuss two variations, one that leads to dilation, and another which does not.

Probability A probability function, P, is defined with respect to a *probability space* $(\Omega, \mathscr{F}, P)$ such that $\mathscr{F}$ is a σ-algebra over a set Ω and $P : \mathscr{F} \longrightarrow [0,1]$ is a non-negative measure satisfying three conditions:

(P1) $P(E) \geq 0$, for all $E \in \mathscr{F}$,

(P2) $P(\Omega) = 1$,

(P3) $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=i}^{\infty} P(E_i)$, when E_i are countable, pairwise disjoint elements of $\mathscr{F}$.

In plain terms, P is a single probability function which assigns to every event in the algebra $\mathscr{F}$ a numerically determinate real number. When Ω is finite and E and F are elements of $\mathscr{F}$, a corollary of P3 is

(P3′) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Even though P is a single, well-defined measure, by strategically withholding information about $(\Omega, \mathscr{F}, P)$ one may be placed in a position of only being able to derive an interval constraint for a probability assignment rather than a numerically precise assignment. For example, suppose that $P(E) = 1/2$ and $P(F) = 1/2$ but whether E and F are disjoint elements in $\mathscr{F}$ is unknown. If asked to solve for $P(E \cap F) = \beta$, β may take any value within $[0, 1/2]$.

We can generalize this calculation for binary meets and binary joins when only the marginal probabilities of a pair of events are known by the following pair of rules.

Proposition 2.1 *If $P(E)$ and $P(F)$ are defined in $(\Omega, \mathscr{F}, P)$, then:*

1. $P(E \cap F) = \beta$ where

$$\max[0, (P(E) + P(F)) - 1] \leq \beta \leq \min[P(E), P(F)], \quad \textit{and}$$

2. $P(E \cup F) = \beta$ *where*

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

So, a first remark about imprecise probability assignments is that they arise naturally within the context of a standard probability measure. There is nothing exotic or unorthodox about them.

Now, asserting that the range of solutions for an event $P(E)$ is $p \leq \beta \leq q$ is to say that there is a set of probability functions, $\mathbb{P}$, which assign a real number for β between p and q, inclusive. Here each P_i in $\mathbb{P}$ is defined with respect to the same outcome space Ω and algebra $\mathscr{F}$, where the lower probability of E in $\mathbb{P}$ is p, and the upper probability of E in $\mathbb{P}$ is q, written $\underline{P}(E) = p$ and $\overline{P}(E) = q$, respectively, and defined by

(P4) $\underline{P}(E) = p = \inf_{P \in \mathbb{P}} P(E)$, and

(P5) $\overline{P}(E) = q = \sup_{P \in \mathbb{P}} P(E)$.[3]

If $\underline{P}(F) > 0$, then define conditional lower and upper probabilities by $\underline{P}(E \mid F) = \inf_{P \in \mathbb{P}} P(E \mid F)$ and $\overline{P}(E \mid F) = \sup_{P \in \mathbb{P}} P(E \mid F)$, respectively.

The lower probability and upper probability of an event are identical just in case that value is the probability of that event, that is

(P6) $P(E) = \underline{P}(E) = \overline{P}(E)$.

In general, the relationship between lower and upper probabilities is defined by the following conjugacy relation,

(P7) $\overline{P}(E) = 1 - \underline{P}(E^c)$.[4]

Also, lower probability is superadditive,

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

which means that, by (P7), upper probability is subadditive. Finally, the set of distributions yielding a lower probability need not be unique. We write $\underline{\mathbb{P}}(E)$ to denote the set of lower probability distributions of E, $\{P \in \mathbb{P} : P(E) = \underline{P}(E)\}$. This convention extends naturally for the set of upper probability distributions, and the corresponding sets of upper and lower conditional probability distributions.

Conditions (P4-P8) provide a barebones structure for sets of probabilities. If we stick to a single probability space, then the postulates tell us that a gap between lower and upper probability opens only by closing off some part of the algebra from view. In this game of peekaboo, there is no reason to deviate from what the fully

[3] Alternative approaches which induce lower and upper probability are discussed in Haenni et al. (2010).

[4] E^c is the complement of E.

defined measure says about events unless some information about the probability space is hidden from view.

Although this basic model works by hiding parts of the algebra, do not assume that every lower probability model has precise probabilities kept out of sight. For example, suppose that a sample of eligible voters are 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 in the same, for some voters may be undecided, choosing neither commit to Smith nor 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 preferences of the voters but to the presence of 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, in this case the proportion of votes cast for Smith will be precisely the proportion of votes not cast for his opponent. But the poll is designed to estimate voter attitudes about Smith, not to predict the vote count for Smith. The precision of the vote count is irrelevant to resolving the imprecision in the poll of pre-election attitudes. Indeed, often the very point of a pre-election poll is to identify who the undecided voters are as part of an effort to influence how those voters will be cast their ballots on election day.

Independence, part 1 The textbook definition of probabilistic or stochastic independence is given in terms of a single measure. Two events E and F in $\mathscr{F}$ are *stochastically independent* if and only if

(ID) $P(E \cap F) = P(E)P(F)$.

In a standard probability space which defines a precise probability function P, E and F are stochastically independent just in case conditioning on F is irrelevant to estimating E, and vice versa. We say that F is *epistemically irrelevant* of E if and only if

(IR) $P(E \mid F) = P(E)$, when $P(F) > 0$,

and that

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

Although (ID), (IR) and (EI) are equivalent in precise probability models, it will turn out that these are distinct concepts within imprecise probability models. We will return to this point in the second section on independence below.

We may also characterize the degree to which two events diverge from stochastic independence, if they diverge at all, by appealing to an idea dating back to Yule:

(S) $S(E,F) =_{df} \frac{P(E \cap F)}{P(E)P(F)}$.

$S(E,F) = 1$ just in case E and F are stochastically independent, $S(E,F) > 1$ when E and F are positively correlated, and $S(E,F) < 1$ when E and F are negatively correlated.[5] We may extend S to sets of measures $\mathbb{P}$ as well, by defining

$$(\mathbb{S})\quad \mathbb{S}^+(E,F) = \{P \in \mathbb{P} : S(E,F) > 1\},$$
$$\mathbb{S}^-(E,F) = \{P \in \mathbb{P} : S(E,F) < 1\},$$
$$\mathbb{I}(E,F) = \{P \in \mathbb{P} : S(E,F) = 1\}.$$

$\mathbb{I}$, the set of measures in $\mathbb{P}$ whereby E and F are stochastically independent, is also called *the surface of independence* for E and F.

We will revisit independence again in part 2.

Bernoulli trials A Bernoulli trial is an experiment designed to yield one of two possible outcomes, success or failure, which naturally may be coded as 'heads' or 'tails'. A normal coin toss therefore takes as an outcome space $\Omega = \{heads, tails\}$, and a probability function P may be defined with respect to the algebra

$$\mathscr{F} = \{\emptyset, \{heads\}, \{tails\}, \{heads, tails\}\},$$

which contains the null event of neither side appearing face up after a toss, the event of heads appearing, the event of tails appearing, and the sure-thing event of either heads or tails appearing, respectively.

To discuss a series of coin tosses, define the random variable C_i as the outcome of the ith coin toss. For example, the probability that the outcome of the second toss is heads given that the outcome of the first is tails is expressed by

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

We may compress this notation by letting H_i refer to the outcome heads on toss i and T_i refer to the outcome tails on toss i. With this shorthand, Equation (1) is compressed to

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

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

$$\mathscr{B} = \{\{heads\}, \{tails\}\}$$

in the algebra $\mathscr{F}$ partitions the outcome space Ω if the outcome of a coin toss must be in $\mathscr{B}$. This follows trivially from modeling a coin toss as a Bernoulli trial. We say that $\mathscr{B}$ is a measurable partition just in case every event has non-zero probability, which is reasonable to maintain unless the coin is double-sided.

[5] 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). See Wheeler (2009) for discussion.

∴ A fair coin toss is a Bernoulli trial with probability that H is 1/2, and a series of fair coin tosses is a stochastically independent sequence of fair coin tosses, so-defined.

3 Dilation

Lower probability was introduced by having you imagine that part of the algebra of events was blocked from view. It is a short step to interpreting lower probability as a representation of your attitude about events. To make this shift clear in the examples 'You' will denote an arbitrary intentional system, and $\mathbb{P}_Y$ that system's set of credal probabilities. I invite you to play along.[6]

Dilation. An event F *dilates* the event E just in case

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

In words, outcome F dilates E just in case the range of unconditional probability assignments to E, $[\underline{P}(E), \overline{P}(E)]$, is a proper subset of the range of the probability assignments to E given F, $[\underline{P}(E \mid F), \overline{P}(E \mid F)]$.

Now suppose that $\mathscr{B}$ is a measurable partition of $(\Omega, \mathscr{F}, P)$. Then, this partition of possible outcomes $\mathscr{B}$ *strictly dilates* the event E just in case

$$\underline{P}(E \mid F) < \underline{P}(E) \leq \overline{P}(E) \leq \overline{P}(E \mid F), \quad \text{for all } F \in \mathscr{B}.$$

The remarkable thing about strict dilation is the specter of turning a more precise estimate into a less precise estimate, *no matter the outcome*.

Coin Example 1. 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 the outcome may depend on the outcome of the first. Nothing is known about the type or degree of the possible dependence. Let H_1, T_1, H_2, T_2 denote the possible outcomes of each coin toss.[7]

We know the coin is fair and that the first toss is a fair toss, so Your estimate for the first toss is precise. The interaction between the tosses is unknown, but in the extreme the first toss may determine the outcome of the second. This likewise puts a precise constraint on your estimate of the second toss prior to the experiment. Hence, by P6 Your upper and lower marginal probabilities are precisely 1/2:

[6] Alas, 'You', you will find, is also a Good idea. The convention has been followed by de Finetti and Walley, among others.

[7] This is Walley's canonical dilation example (Walley 1991, pp. 298–9), except that here we are using lower probabilities instead of lower previsions. The basic difference is that (i) our measure is defined over events rather than gambles, and (ii) we do not assume that $\mathbb{P}$ is a closed convex set. The coin examples we will consider does not exercise either of these components of Walley's theory. Bernoulli variables are a gamble on a 0 or 1 outcome. So, since we are ignoring decision problems, the behavior of lower probability and lower previsions align. Likewise, convexity is principally needed for rational choice, which we are setting to one side in this essay.

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

However, aside from $P_Y(H_2 \mid H_1 + P_Y(H_2 \mid T_1) = 1 - P_2 \mid H_1) + P(T_2 \mid T_1)$, little is known about the direction or degree of dependence between the pair of tosses. Appealing to Proposition 2.1, model Your ignorance by

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

Suppose now that You learn that the outcome of the first toss is heads. The extremal points from b), $\underline{P}_Y(H_1 \cap H_2) = 0$ and $\overline{P}_Y(H_1 \cap H_2) = 1/2$, can be conditioned by Bayes' rule, yielding

(c) $\underline{P}_Y(H_2 \mid H_1) = \underline{P}_Y(H_1 \cap H_2)/\underline{P}_Y(H_1) = 0$, and
$\overline{P}_Y(H_2 \mid H_1) = \overline{P}_Y(H_1 \cap H_2)/\overline{P}_Y(H_1) = 1$.

So, although $P_Y(H_2) = 1/2$, $P_Y(H_2 \mid H_1)$ may take any value within $[0,1]$. An analogous argument holds if instead you learned that the outcome of the first toss is tails, i.e., $P_Y(H_2) = 1/2$ but $P_Y(H_2 \mid T_1) = [0,1]$. Since $\mathscr{B} = \{H_1, T_1\}$ partitions the outcome space, these two cases exhaust the possible observations. But in either case the probability that the second toss lands heads *dilates* from $1/2$ to the vacuous unit interval upon learning the outcome of the first toss. Your precise probability about the first toss become vacuous *no matter which way the second coin toss lands*. $\diamond$

One way to interpret the extremal points $\underline{P}_Y(H_2 \mid H_1) = 0$ and $\overline{P}_Y(H_2 \mid H_1) = 1$ is as two extreme competing hypotheses about how the second toss is conducted. One hypothesis asserts that the outcome of the second toss is certain to match the outcome of the first, hence $P_1(H_2 \mid H_1) = 1$, whereas the second hypothesis asserts that it is certain to land opposite the first, $P_2(H_2 \mid H_1) = 0$. Your probability estimate of the second toss is therefore precisely $1/2$ before observing the first toss, since the first toss is fair and determines the outcome of the second in one way or the other. However, even though you know the outcome of the first toss provides relevant information, observing the outcome of the first toss is insufficient to distinguish between the two competing hypotheses. So, the result of observing the outcome of the first toss is that Your estimate of the second is now maximally imprecise.

Independence, part 2 As made clear by (Seidenfeld and Wasserman 1993, Theorems 2.1 to 2.3), deviation from stochastic independence is an essential ingredient in dilation. Here are two of those observations.

Theorem 3.1 (Seidenfeld and Wasserman, 1993) *If $\mathscr{B}$ strictly dilates E, then for every $F \in \mathscr{B}$,*

$$\underline{\mathbb{P}}(E|F) \subseteq \mathbb{S}^-(E,F) \text{ and } \overline{\mathbb{P}}(E|F) \subseteq \mathbb{S}^+(E,F).$$

Proof. Pick a representative $P' \in \underline{\mathbb{P}}(E \mid F)$. Then $P'(E\cap F)/P'(F) = \underline{P}(E \mid F) < \underline{P}(E) \leq \overline{P}(E)$. Hence, $S'(E,F) < 1$ (defined on P'). So, $\underline{P}(E|F) \subseteq \mathbb{S}^-(E,F)$. An analogous argument for a representative $P'' \in \overline{\mathbb{P}}(E \mid F)$ yields $S''(E,F) > 1$, so $\overline{P}(E|F) \subseteq \mathbb{S}^+(E,F)$ as well. $\diamond$

Theorem 3.2 (Seidenfeld and Wasserman, 1993) *If for every $F \in \mathscr{B}$,*

$$\underline{\mathbb{P}}(E) \cap \mathbb{S}^-(E,F) \neq \emptyset \;\; and \;\; \overline{\mathbb{P}}(E) \cap \mathbb{S}^+(E,F) \neq \emptyset,$$

then $\mathscr{B}$ strictly dilates E.

Proof. Pick a $P' \in \underline{\mathbb{P}}(E) \cap \mathbb{S}^-(E,F)$. Then, $P'(E) = \underline{P}(E)$ and, since $S'(E,F) < 1$, $P'(E \cap F) < P'(E)P'(F)$. So, $\underline{P}(E) = P'(E) > P'(E \mid F) \geq \underline{P}(E \mid F)$. A similar argument holds for $P'' \in \overline{\mathbb{P}}(E) \cap \mathbb{S}^+(E,F)$ only if $P''(E) > P''(E|F) \geq \overline{P}(E \mid F)$. ⋄

In plain terms, Theorem 3.1 states that whenever the event E is strictly dilated by the partition of events $\{F, F^c\}$, then there exist $P \in \mathbb{P}$ on which the probability of E given F deviates from independence to some degree in one direction, and some $P \in \mathbb{P}$ on which the probability of E given F deviates from independence in the other direction. While Theorem 3.1 gives a necessary condition for strict dilation, Theorem 3.2 gives a sufficient condition. If lower probability of E is also such that E and F are negatively associated, and upper probability of E is also such that E and F are positively associated, then $\{F, F^c\}$ strictly dilates E. Seidenfeld and Wasserman's results are about dependence of particular values, not about dependence of variables. Independence of variables implies independence of all their respective values, but not conversely.

Another important point, which will be made clear in the next two examples, is that the familiar equivalence between independence of a joint distribution as the product of its marginal distributions (ID) and independence as irrelevant information (EI) does not hold in an imprecise model. This 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 another. However, this step, from observed irrelevance of one event to the probability estimate of another to concluding that the one event is stochastic independent of the other, is fallacious within imprecise probability models. What this means is that the familiar route to constructing a joint distribution, such as the joint distribution of two coin tosses landing heads, from estimates of the marginal distributions for each toss plus the knowledge that the two tosses are "independent" is problematic when at least one of the coin tosses has an imprecise estimate as to whether heads will occur. The problem within an imprecise probability setting stems from determining which notion of independence is operating in the model.

Suppose that $X \in \{0,1\}$ and $Y \in \{0,1\}$ are binary random variables. Define $\mathbb{P}_1$ as the set of marginal distribution of X, $\mathbb{P}_2$ as the set of marginal distributions of Y, and $\mathbb{P}$ as the set of joint distributions for $X \cap Y$. There are several ways to construct $\mathbb{P}$ from estimates of X defined in terms of the upper and lower probabilities of X in $\mathbb{P}_1$, estimates of Y defined in terms of upper and lower probabilities of Y in $\mathbb{P}_2$, and judgements that "X is independent of Y"—differences owing to distinct independence concepts this judgment can invoke. Here are three important concepts.

(𝕀ℝ) Learning whether Y (i.e., whether $Y = 0$ or $Y = 1$) is *epistemically irrelevant* to X when $\mathbb{P}(X \mid Y \in \{0,1\}) = \mathbb{P}_1(X)$, for $\overline{\mathbb{P}_2}(Y \in \{0,1\}) > 0$, where

$$\mathbb{P}(X \mid Y \in \{0,1\}) = \{P(\cdot \mid Y \in \{0,1\}) : P \in \mathbb{P}(X \cap Y) \text{ and } P(Y \in \{0,1\}) > 0\}.$$

(𝔼𝕀) X is *epistemically independent* of Y just when Y is epistemically irrelevant to X, and X is epistemically irrelevant to Y.

(𝕀𝔻) X and Y are *completely stochastically independent* if, for all $P \in \mathbb{P}$,

$$P(X \in \{0,1\} \cap Y \in \{0,1\}) = P(X \in \{0,1\})P(Y \in \{0,1\}),$$

which also guarantees that

$$P(X \in \{0,1\} \mid Y \in \{0,1\}) = P(X \in \{0,1\}), \text{ when } P(Y \in \{0,1\}) > 0.$$

So, within an imprecise probability setting we see that (𝕀𝔻) ⇒ (𝔼𝕀) ⇒ (𝕀ℝ). However, (𝕀ℝ) $\not\Rightarrow$ (𝔼𝕀): epistemic irrelevance is asymmetric, whereas epistemic independence is symmetric. Also (𝔼𝕀) $\not\Rightarrow$ (𝕀𝔻): judging that two experiments are epistemically independent does not ensure that their underlying uncertainty mechanisms are stochastically independent. A joint set of distributions may satisfy epistemic independence without being factorizable. The "fat coins" example (below) demonstrates a case where (𝔼𝕀) is satisfied but (𝕀𝔻) is not.

Sorting out the riddle of dilation hinges on making explicit what independence concepts are invoked in a problem. With this observation in mind, compare Example 1, which explicitly states that the interaction between the tosses is unknown, to the next example, Example 2, which states that the tosses are independent but without specifying explicitly *which* notion of independence is operative.

Coin Example 2. Suppose that there are two coins rather than one.[8] Both are tossed normally, but only the first is a fair coin toss. The second coin is of unknown bias.

(d) $P_Y(H_1) = 1/2$,

(e) $0 < \underline{P}_Y(H_2) \leq \overline{P}_Y(H_2) < 1$, written $P_Y(H_2) \in (0,1)$.[9]

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

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

[9] 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.

(f) $\underline{P}_Y(H_1 \cap H_2) = \underline{P}_Y(H_1)\underline{P}_Y(H_2) \approx 0$, and

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

Now suppose both coins are tossed and the outcomes are known to me but not to You. I 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) I announce that "H_1 iff H_2" (M) or I announce that "H_1 iff T_2" ($\neg M$).

Since You are told that the first and the second tosses are independent, and initially Your estimate that the outcomes are matched or split is $1/2$, then

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

and

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

Also, it follows that

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

We also observe that

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

Equation 5 says that my announcing whether the two tosses are matched or split is epistemically irrelevant to Your estimate of the second toss landing heads. Analogously, one might also think that announcing whether the two tosses are matched or split is epistemically irrelevant to estimating the first toss. Surely it seems strange that my announcement is irrelevant to changing Your view of a coin You know nothing about but nevertheless relevant to changing Your view about a fairly tossed coin, in which case we might think that

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

should hold as well.

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

$$P_Y(M) \neq P_Y(M \mid H_1) \in (0,1), \tag{7}$$

and a symmetric argument holds if instead the first coin lands tails.

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

The short answer is that the apparent contradiction between Equation 6 and Equation 7 stems from mixing together different concepts of independence. Equation 6 expresses that my announcement of whether 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. But, judging that one event is epistemically irrelevant to another does not ensure that the two events are epistemically independent when one of the events has an imprecise estimate.

In any case, it should be clear that (f) does not entail that my announcement is completely stochastically independent of the first toss. For suppose that β is the unknown bias of the second toss landing heads. Condition (f) says that,

$$\begin{aligned} P(H_1 \cap H_2) &= 1/2\beta \\ P(H_1 \cap T_2) &= 1/2(1-\beta) \\ P(T_1 \cap H_2) &= 1/2\beta \\ P(T_1 \cap T_2) &= 1/2(1-\beta), \end{aligned}$$

which entails $P_Y(H_1 \mid M) = \beta$. Therefore, if $0 \leq \beta \leq 1$, my 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 (Joyce 2011) response to Example 2, which is to reject $P_Y(H_1 \mid M) = P_Y(H_1)$ 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 $P_Y(H_1 \mid M) = P_Y(H_1) = P_Y(M) = 1/2$ holds everywhere. While proponents of [The principle of Indifference] will find this congenial, friends of imprecise probabilities will rightly protest that there is no justification for treating the events as independent. (Joyce 2011, my notation)

According to Joyce, 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" (2011).

On the other hand, 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 (White 2010) to view the conflict in Example 2 to be a counterexample to imprecise credal probability.

I will argue that there is a consistent imprecise probability model for Example 2 and that Joyce and White each have it half right. My claims are these:

- *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.

First, let us explore visually how to consistently represent these two claims together with the conditions spelled out 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

$$P(H_1 \cap H_2) = \alpha_1 \quad P(T_1 \cap T_2) = \alpha_4$$
$$P(H_1 \cap T_2) = \alpha_2 \quad P(T_1 \cap H_2) = \alpha_3.$$

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, with maximal probabilities for each of the four possible outcomes marking the four extremities,

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

In a fully specified precise probability model for independent tosses, the α_i's are identical and their value corresponds to a single point within the tetrahedron. (If the flips are both fair coin tosses, that value would be $1/4$.) At the other extreme, in a completely unconstrained imprecise probability model for the set $\mathbb{P}$ of measures, the entire tetrahedron may represent admissible values.[10] So, what You know initially about the coin tosses will translate to conditions that constrain the space of admissible probabilities within this polytope, and what You learn by updating will translate to some other region within the tetrahedron. Different independence concepts translate to different ways of rendering one event irrelevant to another, and 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 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 $P_Y(H_1) = 1/2$.

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

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

(f) *i*. The upper and lower probability of both tosses landing heads is the shaded pentahedron, where the base of the polytope defined by the coordinates α_2, α_3, α_4 represents the lower probability $\underline{P}_Y(H_1 \cap H_2) = 0$, 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 $\overline{P}_Y(H_1 \cap H_2) = 1/2$.

ii. Toss C_1 is independent of C_2,[11] so $\mathbb{I}(C_1, C_2) \neq \emptyset$ but $\mathbb{S}^+(C_1, C_2) = \mathbb{S}^-(C_1, C_2) = \emptyset$. 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

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

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

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

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.

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

On this parameterization, the pair of coin tosses are epistemic independent and this relationship is represented by the saddle-shaped surface of independence satisfying Equations 8 and 9. Furthermore, Your initial estimate that I 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.

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

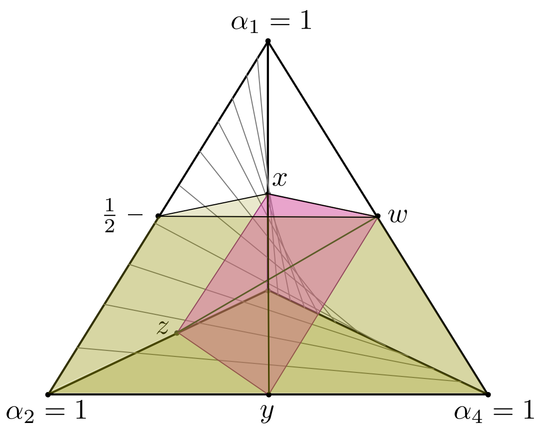

Figure 1: The constraints on Coin Example 2.

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 I announce "split", which is the edge $\alpha_2\alpha_3$. So, announcing whether the two tosses match or whether 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 an intuitive sense in which learning whether the outcomes match or split is epistemically irrelevant to Your estimate of heads. This raises two questions. First, whether there room within an imprecise probability model to accommodate this view; second; if so, whether it is 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 distribution 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 my 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 I 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, $\mathbb{P}_Y(M \cap H_1)$, cannot be the same as above, since $P_Y(M \mid H_1) \in (0,1) \neq P_Y(M)$.

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

Accommodating the conflicting intuitions about Example 2 involved noticing the distinction between epistemic independence, which is symmetric, and epistemic irrelevance, which is asymmetric. Even so, we also mentioned that there is 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 this difference, let’s first take a detour through another example which illustrates a set of distributions which satisfies epistemic independence but fails to satisfy complete stochastic independence.

Fat Coins Example. Imagine that there are two three-sided coins with the heads side painted black and the tails side painted white. The color of the remaining side is unknown to You: it may be painted black, painted white, or not painted at all. The bias of the coins are such that the probability of the first coin landing heads is 0.5, tails 0.2, and side 0.3. The probability of the second coin landing heads is 0.2, tails 0.4, and side 0.4. However the coins are tossed, suppose that all You know is both that the conditional probability of the second coin landing black is between 0.2 and 0.6, no matter the color of the first toss, and that the conditional probability of the first coin landing black is between 0.5 and 0.8, no matter the color of the second toss. In other words, all that You know is that the tosses are epistemically independent,

$$\begin{aligned} 0.5 \leq P(B_1) \leq 0.8 \qquad & 0.2 \leq P(B_2) \leq 0.6, \\ 0.5 \leq P(B_1 \mid B_2) \leq 0.8 \qquad & 0.2 \leq P(B_2 \mid B_1) \leq 0.6, \\ 0.5 \leq P(B_1 \mid W_2) \leq 0.8 \qquad & 0.2 \leq P(B_2 \mid W_1) \leq 0.6. \end{aligned}$$

Consider the largest set of joint probability distributions $P \in \mathbb{P}$ satisfying the extremal points $\overline{P}(B_1) = 0.8$, $\overline{P}(B_2 \mid B_1) = 0.6$, and $\underline{P}(B_2 \mid W_1) = 0.2$ from above:

$$\begin{aligned}
P(B_1 \cap B_2) &= \overline{P}(B_1)\overline{P}(B_2 \mid B_1) = 0.8 \cdot 0.6 = 0.48, \\
&= \overline{P}(B_2)\overline{P}(B_1 \mid B_2) = 0.6 \cdot 0.8 = 0.48. \\
P(B_1 \cap W_2) &= \overline{P}(B_1)\underline{P}(W_2 \mid B_1) = 0.8 \cdot 0.4 = 0.32, \\
&= \underline{P}(W_2)\overline{P}(B_1 \mid W_2) = 0.4 \cdot 0.8 = 0.32. \\
P(W_1 \cap B_2) &= \underline{P}(W_1)\overline{P}(B_2 \mid W_1) = 0.2 \cdot 0.6 = 0.12, \\
&= \overline{P}(B_2)\underline{P}(W_1 \mid B_2) = 0.6 \cdot 0.2 = 0.12. \\
P(W_1 \cap W_2) &= \underline{P}(W_1)\underline{P}(W_2 \mid W_1) = 0.2 \cdot 0.4 = 0.08, \\
&= \underline{P}(W_2)\underline{P}(W_1 \mid W_2) = 0.4 \cdot 0.2 = 0.08.
\end{aligned}$$

Even though the tosses are epistemically independent, they are not stochastically independent. For example, $\overline{P}(H_2 \mid H_1) \neq \underline{P}(H_2 \mid T_1)$. ◇

With this observation one might worry that we have merely traded a small problem for a larger one. The fat coin example shows that we cannot ensure that two processes are stochastically independent simply from knowing that two experiments are epistemically independent. But sometimes the reason for treating outcomes as epistemically independent is because we know they result from stochastically independent mechanisms.

To put this worry to rest, consider again Example 2 but with a twist. Suppose now that my announcement of either M or $\neg M$ is stochastically independent of the first toss and stochastically independent of the second toss as well. Notice on this telling my announcement is no longer a reliable report about the outcome of the pair of tosses. Like before, whether I announce "match" or announce "split" is irrelevant to Your estimate that the first coin toss lands heads. But now it also is true that learning how the first toss lands is irrelevant to your estimate of whether I will announce "match" or "split". Surely, You should be able to handle my irrelevant announcements *in either direction* without injecting phony precision into Your model to maintain stochastic independence.

Here the epistemic independence between the tosses and between tosses and my announcements is because we understand the underlying uncertainty mechanism to be stochastically independent from each other. To represent complete stochastic independence we consider another parameterization of the second coin example in Figure 2 that allows us to explicitly say that the pair of coin tosses described in the 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. 2010, §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 cube.

$$P(H_2 \mid H_1) = \{a, d\} \quad P(T_2 \mid T_1) = \{r, s\}$$
$$P(H_2 \mid T_1) = \{b, c\} \quad P(T_2 \mid H_1) = \{q, t\}.$$

One surface 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.

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*.

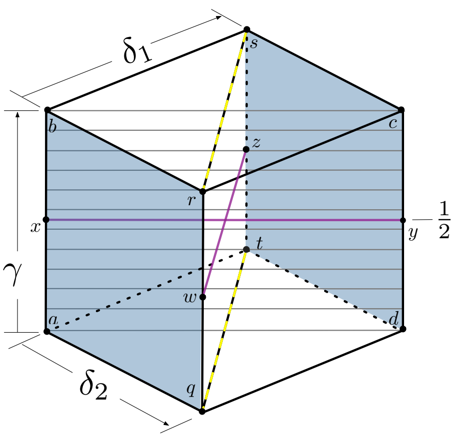

Figure 2: Independence constraints in Coin example 2.

Now suppose I 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 my making an announcement, the specific issue is whether my announcement has the effect of dilating Your estimate of H_1.

Because since we have explicitly ensured stochastic independence between the tosses and my random announcements, we now see that conditioning only yields probabilities within the set $\mathbb{P}$ which are on the surface of independence. As a consequence, 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 me either announcing “match” or announcing“split,” just as before, but in addition You also know that Your estimate of M is not affected by knowing that the first toss is heads.

Is dilation reasonable?

The motor driving Examples 1 and 2 is imprecision in estimating the probability of the joint outcome of the two tosses, but the examples differ over the source of that imprecision. In the first example, the origin is the unknown interaction between the pair of tosses. We know the coin is fair, but we do not know how the second toss is performed. It is reasonable for you to dilate in this case since You may view the extremal points as two hypotheses about how the second coin is tossed, each asserting a causal relationship between the first toss and the second but to opposite effect.

The origin of imprecision in the second example is the unknown bias of the second coin. In this case we know the tosses are performed independently, but we considered a scenario in which announcing 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 added condition appears to make the tosses 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 I 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 asymmetric, in general, and in this case the announcement is epistemically irrelevant to estimating the first toss but the outcome of the first toss is epistemically relevant to estimating whether I will announce “match.” The error is compounded further by mistakenly assuming that epistemic independence is sufficient for stochastic independence, since the two tosses are epistemically irrelevant but are not stochastically independent.

Putting this together, we see that the first example is a genuine case of dilation, but the second example is mixed. In Example 2 the announcement of how the first toss turned out is relevant information to estimating whether both outcomes match or differ, and the outcome of the first toss would dilate Your estimate of whether the outcomes of both tosses match. But, the announcement of whether the tosses match *should not* dilate Your original estimate *even though my announcement and the first toss are associated*. In this direction, my announcement about how both coins landed is epistemically irrelevant to Your estimate of heads on the first toss.

In the last section we provided an explanation for how to give a coherent imprecise probability model of Example 2 interpreted in this manner. In this section we turn to consider an argument for why this model is rational. I will focus my remarks on the most controversial assumption, from an imprecise probability point of view, which is our endorsement of Equation 6.

Imagine that You maintain that $P_Y(H_1 \mid M) = P_Y(H_1) = P(M) = 1/2$ and are willing to maintain that the probability of the first toss landing heads is $1/2$ even after I 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.[12] You have a fair coin, I have a coin of unknown bias, and You have some strategy for guessing heads or tails for Your coin on the *n*th flip of an unbounded sequence of trials. I pay You €1.00 if Your guess is right, You pay me €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. I flip my coin, which has a fixed bias β for all trials. I observe the outcome of Your toss and my toss, then I announce to You whether they match. If I 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 my announcement that the coins match. Without Your knowledge, suppose I pick $\beta < 1/2$. Then my expected profit on each trial is zero: $1/2(1/2 - \beta) + 1/2(\beta - 1/2) = 0$.

If, before I pick the bias β of the coin, I know Your betting strategy, and it is other than flipping Your secondary coin, then I can bankrupt You by strategically choosing β. If you know β, then You can alter your betting strategy to bankrupt me. 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 $P_Y(H_2 \mid H_1) \neq P_Y(H_2)$ and are unwilling to maintain that the probability of the second toss landing heads is $1/2$ after I faithfully report to you that the first toss is heads.

In this case, You have a fair coin and I have a fair coin, but only Yours is tossed fairly. Just as before, I pay You €1.00 if Your guess is right, You pay me €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, I either arrange my coin to match the outcome of Your toss, or I arrange my 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 the same, $1/2$, as Your degree of belief that the outcome of my "toss" is heads. However, after I announce that Your coin

[12] Thanks to Clark Glymour for putting this argument to me.

landed heads, Your probability that my 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 unwilling 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.

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 discredit imprecise probability models by showing that events which have nothing to do with one another can nevertheless have mysterious effects on one or the other's probability estimates. But the scandal invariably rests on the assumption that *dilator* and *dilatee* are completely stochastically independent. In so far as this is true, the examples are not dilation examples. And when there is genuine dilation, then there is a possible interaction in the probability model that may or may not be relevant to Your probability estimates. Furthermore, 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 this is the best evidence available for a necessary action. Having this choice is useful, which counts against a blanket policy that either expunges or embraces dilation wholesale. Dilation can yield relevant information about Your probability model, but this isn't always true. In any event, reckoning with dilation is a good idea, even if not a Good idea.

Acknowledgements Thanks to Marco Alberti, Horacio Arló-Costa, Seamus Bradley, Stephan Hartmann, Jim Joyce, and Jon Williamson for their comments on earlier drafts, and thanks especially to Clark Glymour and Choh Man Teng for a long discussion on dilation which began one afternoon in a Lisbon café. This research was supported in part by award (LogICCC/0001/2007) from the European Science Foundation.

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.

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.* King's College Publications, Lon.

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

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(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., Romeyn, J.-W., Wheeler, G., and Williamson, J. (2010). *Probabilistic Logic 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. In *Philosophical Perspectives*. Blackwell.

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. 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.

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

White, R. (2010). Evidential symmetry and mush 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.