

## LEARNING AND POOLING, POOLING AND LEARNING

RUSH T. STEWART AND IGNACIO OJEA QUINTANA

ABSTRACT. We explore which types of probabilistic updating commute with convex IP pooling (Stewart and Ojea Quintana, 2017). Positive results are stated for Bayesian conditionalization (and a mild generalization of it), imaging, and a certain parameterization of Jeffrey conditioning. This last observation is obtained with the help of a slight generalization of a characterization of (precise) externally Bayesian pooling operators due to Wagner (2009). These results strengthen the case that pooling should go by imprecise probabilities since no precise pooling method is as versatile.

### 1. INTRODUCTION

Bayesian conditionalization is the gold standard of probabilistic learning. Yet several authors advocate modifications of conditionalization for a number of reasons. For example, conditionalization entails assigning probability 1 to the evidence. Dissatisfied with such “dogmatic epistemology,” Richard Jeffrey proposed his *probability kinematics* as a way of updating on *uncertain* evidence (Jeffrey, 2004). To take another example, consider probabilistic *imaging*. One widely pursued goal in work on the logic of conditionals is to find a way of identifying the probability of a conditional with the corresponding conditional probability. Attempts to do so have been repeatedly frustrated by a series of triviality results. However, David Lewis introduces imaging and shows that a version of the identity holds when formulated in terms of imaging instead of in terms of conditionalization (Lewis, 1976). And there are other proposals, such as minimizing the Kullback-Leibler divergence (Kullback and Leibler, 1951). Here, the objective is to accommodate the evidence in such a way as to minimize a measure (the K-L divergence) of the difference between posterior and prior.

Probabilistic opinion pooling can be viewed as part of an important strand in Bayesian epistemology and statistics concerned with consensus. The received view is that personal probabilities are *subjective* (Ramsey, 1931; Savage, 1954; de Finetti, 1964), resulting in much fretting about the implications for scientific methodology. The worry is that the objectivity of scientific confirmation, explanation, inference, and the like is compromised to the extent that such probability plays a role. A prominent Bayesian response comes in the form of convergence and merging of opinions theorems, which show that, given agreement about probability 0 events and enough evidence, probabilities converge (almost surely) (Savage, 1954; Gaifman and Snir, 1982).<sup>1</sup> Conditionalization, that is, leads to consensus, “washing out” the problematically subjective priors leaving intersubjective agreement in their place.

---

*Date:* April 1, 2017.

<sup>1</sup>Not all merging of opinions results require probabilities to converge to certainty (Blackwell and Dubins, 1962). Under certain conditions, Bayesian conditionalizing can bring probabilities close even if they do not converge to 1 or 0.

Pooling offers a distinct way of reaching a consensus in probabilistic opinion, one available even when the opportunity to collect more evidence is not. Consensus is reached immediately *via* methods for *aggregating* probabilistic judgments instead of in the long run *via* conditionalization (Huttegger, 2015). After all, as Keynes astutely observes, we have reasons not to be particularly concerned with the long run. As with convergence arguments, intersubjective agreement stands in for objectivity in the pooling context. Still, in the literature on probabilistic opinion pooling, one of the constraints of central concern is *external Bayesianity*, which requires that pooling and Bayesian conditionalization on a common likelihood function commute (Madansky, 1964). That is, the result of pooling and then updating is the same as first updating and then pooling. The order of operations does not change the outcome. One natural question to ask in light of the aforementioned alternatives to conditionalization is, what about commutativity with alternative updating policies? This is the question that concerns us in the present essay.

Much of the focus in the pooling literature is on characterization and impossibility results. Such results are not the intended contribution of this paper (though Proposition 2 generalizes a characterization result due to Wagner to the imprecise probabilities setting). Instead, we continue an exploration of the potential of imprecise probabilities in the context of learning and pooling (Stewart and Ojea Quintana, 2017). There, we argue that collective opinion is more properly represented by imprecise probabilities (IP). We provide three arguments. First, if pooling is interpreted as reaching a *consensus* in probabilistic opinion, IP pooling is on firmer philosophical ground (Cf. Levi, 1985; Seidenfeld et al., 1989). The point, briefly, is that IP models allow for suspending judgment between some number of probability distributions by not ruling them out for use in deliberation and inquiry, and reflect the common ground among the group concerning which probability distributions *are* ruled out. Such a consensus constitutes a neutral initial position from which to launch further inquiry. Precise pooling functions, on the other hand, do not allow for an analogous suspense of judgment, and may yield collective probabilistic opinions endorsed by none of the group members. Second, there are IP pooling functions that jointly satisfy more of the standard pooling constraints than any precise pooling recipe can. Third, in the IP setting, the tension between a pooling method's being justified on epistemic or procedural grounds (Dietrich and List, 2014)—reflected in the tension between satisfying certain formal epistemic and procedural constraints—dissipates, an artifact of the assumption of precision.

The results that follow may be taken to contribute to that case for IP. Briefly put, we show that, while the form of updating for a given precise pooling method is quite restricted under the requirement of commutativity, relaxing the assumption that the collective opinion should take the form of a numerically determinate probability function enables us to lift many of those restrictions. Several revision methods are consistent with pooling understood the IP way. After introducing the mathematical pooling framework in the next section, we begin with the gold standard in Section 3. We remain neutral as to whether Jeffrey conditionalization and imaging ultimately admit of sufficient motivation, though we rehearse some of the standard motivations for and reservations about probability kinematics (Section 5) and imaging (Section 6), and state the commutativity results. Motivations for requiring commutativity of learning and pooling are discussed Section 4.

## 2. PRELIMINARIES

$\Omega$  denotes a sample space, a set of mutually exclusive and exhaustive possible states of the world.<sup>2</sup> In what follows, we assume that  $\Omega$  is countable. A function  $\mathbf{p} : \Omega \rightarrow [0, 1]$  is a *probability mass function* (pmf) iff  $\sum_{\omega \in \Omega} \mathbf{p}(\omega) = 1$ . An algebra  $\mathcal{A}$  of events over  $\Omega$  is a set of subsets of  $\Omega$  closed under complementation and finite unions (closure under countable unions yields a  $\sigma$ -algebra). We assume throughout the paper that  $\mathcal{A}$  is a  $\sigma$ -algebra. Given a pmf, we can define a *probability measure*, (abusing notation by using the same symbol as for pmfs)  $\mathbf{p}$ , on events in general:  $\mathbf{p}(E) = \sum_{\omega \in E} \mathbf{p}(\omega)$ .

Let  $\mathbb{P}$  be the set of all pmfs on  $\Omega$ . A *precise* pooling function is a function,  $F : \mathbb{P}^n \rightarrow \mathbb{P}$ , mapping a profile of  $n$  pmfs,  $(\mathbf{p}_1, \dots, \mathbf{p}_n)$ , to a single pmf,  $F(\mathbf{p}_1, \dots, \mathbf{p}_n)$ . Typically, the  $n$  pmfs are taken to represent the opinions of a set  $N$  of individuals, and  $F(\mathbf{p}_1, \dots, \mathbf{p}_n)$  is supposed to represent, in some sense, the aggregate or collective opinion. Various candidate interpretations of an opinion pool exist in the literature: a rational consensus (adopted genuinely by all members or adopted merely “for the sake of the argument”); a compromise adopted for the purpose of group decision making; the opinion a group member adopts after learning the opinion of her “epistemic peers”; the opinion an external agent adopts upon being informed of the  $n$  expert opinions, etc. (Genest and Zidek, 1986; Wagner, 2009). There are a number of concrete pooling functions discussed in the literature, but, by far, the two most prominent are linear and geometric pooling functions.

**Linear Opinion Pools.**  $F(\mathbf{p}_1, \dots, \mathbf{p}_n) = \sum_{i=1}^n \alpha_i \mathbf{p}_i$ , where  $\alpha_i \geq 0$  and  $\sum_{i=1}^n \alpha_i = 1$ .

A linear opinion pool is just a weighted arithmetic average of the  $n$  probability functions. A geometric pooling function takes the (weighted) *geometric* average of the  $n$  pmfs.

**Geometric Opinion Pools.**  $F(\mathbf{p}_1, \dots, \mathbf{p}_n) = c \prod_{i=1}^n \mathbf{p}_i^{\alpha_i}$ , where  $\alpha_i \geq 0$  and  $\sum_{i=1}^n \alpha_i = 1$ , and  $c = \frac{1}{\sum_{\omega' \in \Omega} [\mathbf{p}_1(\omega')]^{\alpha_1} \dots [\mathbf{p}_n(\omega')]^{\alpha_n}}$  is a normalization factor.<sup>3</sup>

The focus of this paper will be on a generalization of pooling functions to the IP setting:  $\mathcal{F} : \mathbb{P}^n \rightarrow \mathcal{P}(\mathbb{P})$ .<sup>4</sup> We use  $\mathcal{F}$ , opposed to  $F$ , to denote set-valued pooling operators. IP pooling functions are maps from profiles of probability measures to sets of probability measures. In particular, we consider pooling functions that map profiles of  $n$  pmfs to the convex hull of those functions:  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n) = \text{conv}\{\mathbf{p}_i : i = 1, \dots, n\}$ . The convex hull is the smallest convex set containing  $\mathbf{p}_i$  for  $i = 1, \dots, n$ . A set is convex if it satisfies the following property.

<sup>2</sup> $\Omega$  may be thought of as a partition of a space of agent-relative serious possibilities determined by consistency with a state of full belief. As is a state of full belief,  $\Omega$  is open to being revised, refined, etc., as judged appropriate (Levi, 1980).

<sup>3</sup>Notice that, due to the way geometric pooling is defined, there are profiles for which  $F(\mathbf{p}_1, \dots, \mathbf{p}_n)(\omega) = 0$  for all  $\omega \in \Omega$ —in violation of the probability axioms. Such a situation arises if for each  $\omega \in \Omega$  there is a  $\mathbf{p}_i \in (\mathbf{p}_1, \dots, \mathbf{p}_n)$  such that  $\mathbf{p}_i(\omega) = 0$ . Circumventing this problem, Wagner restricts the domain of pooling operators to the set of profiles for which this does not happen. That is, the domain of a pooling function is the set of profiles such that there is some  $\omega \in \Omega$  for which  $\mathbf{p}_i(\omega) > 0$  for all  $i = 1, \dots, n$ .

<sup>4</sup>See (Schervish and Seidenfeld, 1990; Herron et al., 1997) for studies of convergence relevant to IP.

**Convexity.** If  $\mathbf{p}_1, \mathbf{p}_2 \in \mathbf{P}$ , then  $\alpha\mathbf{p}_1 + (1 - \alpha)\mathbf{p}_2 \in \mathbf{P}$  for  $\alpha \in [0, 1]$ .<sup>5</sup>

Put another way,  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  is the set of all convex combinations of the individual probability functions. We let  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)(A)$  denote the set of probability values assigned to  $A$ :

$$\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)(A) = \{\mathbf{p}(A) : \mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)\}$$

There are alternative IP formats, including set-based formats that do not require convexity, interval-valued probability functions, or, more operationally, sets of desirable gambles, for instance. In our view, all are worthy of extensive study. Convex sets are quite commonly employed in the theory of IP, including in sophisticated decision theories. Since we do not intend to settle complex debates internal to IP theory here, and the results to come do not hinge on convexity in the sense that there are alternative IP formats for which they hold, we will simply restrict our attention to convex sets as an illustration of the potential of IP in theorizing about pooling.

### 3. EXTERNAL BAYESIANITY

Essential to Bayesian methodology is the assumption of a *prior* probability distribution on the algebra of events (or propositions) of concern. Learning proceeds by *conditionalizing* the prior on the evidence, yielding a *posterior* distribution. Conditionalization of a probability function,  $\mathbf{p}$ , on evidence,  $E$ , results from setting the posterior probability for any event  $A \in \mathcal{A}$  equal to the prior conditional probability  $\mathbf{p}(A|E)$ .

$$\mathbf{p}^E(A) = \mathbf{p}(A|E) = \frac{\mathbf{p}(A \cap E)}{\mathbf{p}(E)}, \text{ when } \mathbf{p}(E) > 0.$$

The posterior,  $\mathbf{p}^E$ , can be thought of as the result of learning  $E$ . In the context of sets of probability functions, conditionalization can be generalized by conditionalizing each member of the set:

$$\mathcal{F}^E(\mathbf{p}_1, \dots, \mathbf{p}_n) = \{\mathbf{p}^E : \mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n), \mathbf{p}(E) > 0, \text{ and } \mathbf{p}^E(\cdot) = \mathbf{p}(\cdot|E)\}$$

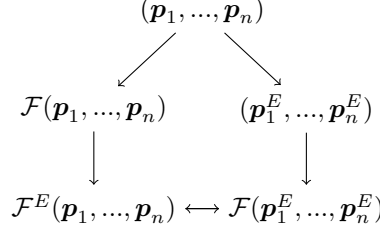
Call  $\mathcal{F}^E(\mathbf{p}_1, \dots, \mathbf{p}_n)$  the *prior-by-prior* conditionalization of  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  (when  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  is convex,  $\mathcal{F}^E(\mathbf{p}_1, \dots, \mathbf{p}_n)$  is called the *convex* conditionalization of  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ ).<sup>6</sup> We define prior-by-prior conditionalization generally, allowing  $\mathbf{p}(E) = 0$  for some  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  (Cf. Kyburg, 1987, p. 279). But when we first update the  $\mathbf{p}_i$ , we assume  $\mathbf{p}_i(E) > 0$  for  $i = 1, \dots, n$ ; otherwise,  $\mathcal{F}(\mathbf{p}_i^E, \dots, \mathbf{p}_n^E)$  is not defined. Though not stated in the language of probabilistic opinion pooling, proofs of the commutativity of convexifying a set of probability functions and conditionalization exist in the literature.

**Theorem 1.** (*Levi, 1978; Girón and Ríos, 1980*) *Convex IP pooling commutes with conditionalization.*

<sup>5</sup>Within the IP research community, convexity is a matter of some controversy. For attacks on the requirement, see (Seidenfeld et al., 1989; Kyburg and Pittarelli, 1992; Seidenfeld et al., 2010). For defenses, see (Levi, 1990, 2009).

<sup>6</sup>In the IP setting, conditionalization can actually lead to *greater* uncertainty in the short-run, a very interesting phenomenon known as *dilation* (Seidenfeld and Wasserman, 1993; Pedersen and Wheeler, 2014).

FIGURE 1. Commutativity of Pooling and Conditionalization



Importantly, linear opinion pooling does not commute with conditionalization, though geometric pooling does (Genest, 1984; Russell et al., 2015). As we will see, linear pooling does commute with imaging, though geometric pooling does not.

As standardly pointed out, *external Bayesianity* is a generalization of the requirement that pooling and standard conditionalization commute (Wagner, 2009; Dietrich and List, 2014; Russell et al., 2015), because conditionalization on a common likelihood function generalizes standard Bayesian conditionalization on an event. A *likelihood function*,  $\lambda : \Omega \rightarrow [0, \infty)$ , is intended to encode, given any  $\omega \in \Omega$ , how expected some evidence is with the number  $\lambda(\omega)$ . The conditionalization of a pmf,  $\mathbf{p}$ , on a likelihood function,  $\lambda$ , is given by the following formula.

$$\mathbf{p}^\lambda(\omega) = \frac{\mathbf{p}(\omega)\lambda(\omega)}{\sum_{\omega' \in \Omega} \mathbf{p}(\omega')\lambda(\omega')}, \text{ when } \sum_{\omega' \in \Omega} \mathbf{p}(\omega')\lambda(\omega') > 0$$

For the special case of Bayesian conditionalization on an event  $E$ , define  $\lambda$  as the indicator function for that event:

$$\lambda(\omega) = \begin{cases} 1, & \text{if } \omega \in E \\ 0, & \text{otherwise.} \end{cases}$$

(We verify the claim with routine substitutions in the footnote.<sup>7</sup>)

External Bayesianity requires that updating the individual probabilities on a common likelihood function and then pooling is the same as pooling and then updating the pool on that likelihood function.

**External Bayesianity.** For every profile  $(\mathbf{p}_1, \dots, \mathbf{p}_n)$  in the domain of  $F$  and every likelihood function  $\lambda$  such that  $(\mathbf{p}^\lambda, \dots, \mathbf{p}_n^\lambda)$  remains in the domain of  $F$ ,  $F(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda) = F^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n)$ .

When pooling is presumed to produce a numerically determinate probability function for the group, generalized geometric pooling functions uniquely satisfy external

<sup>7</sup>For any  $A \in \mathcal{A}$ ,  $\mathbf{p}^E(A) = \frac{\mathbf{p}(A \cap E)}{\mathbf{p}(E)} = \frac{\sum_{\omega \in A \cap E} \mathbf{p}(\omega)}{\sum_{\omega \in E} \mathbf{p}(\omega)}$ . By the definition of a probability measure,  $\mathbf{p}(A) = \sum_{\omega \in A} \mathbf{p}(\omega)$ , so  $\sum_{\omega \in A} \mathbf{p}^\lambda(\omega) = \frac{\sum_{\omega \in A} \mathbf{p}(\omega)\lambda(\omega)}{\sum_{\omega' \in \Omega} \mathbf{p}(\omega')\lambda(\omega')}$  gives us  $\mathbf{p}^\lambda(A)$ . We show that these two fractions are equal by showing the equality of both the numerators and denominators. Since, for all  $\omega \in A$ ,  $\mathbf{p}(\omega)\lambda(\omega) = \mathbf{p}(\omega)$  if  $\omega \in E$  and 0 otherwise,  $\sum_{\omega \in A} \mathbf{p}(\omega)\lambda(\omega) = \sum_{\omega \in A \cap E} \mathbf{p}(\omega) = \mathbf{p}(A \cap E)$ . Hence, the numerators are equal. And since, for all  $\omega' \in \Omega$ ,  $\mathbf{p}(\omega')\lambda(\omega') = \mathbf{p}(\omega')$  if  $\omega' \in E$  and 0 otherwise, we have  $\sum_{\omega' \in \Omega} \mathbf{p}(\omega')\lambda(\omega') = \sum_{\omega' \in E} \mathbf{p}(\omega') = \mathbf{p}(E)$ . Hence, the denominators are equal, too. So,  $\mathbf{p}^E = \mathbf{p}^\lambda$ .

Bayesianity (Genest et al., 1986; Nau, 2002). Extended to the IP setting, the constraint requires  $\mathcal{F}(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda) = \mathcal{F}^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n)$ . The requirement is still that learning by updating on a common likelihood function and pooling commute, but the format of the pool is altered. What is not altered, we submit, are the compelling aspects of the constraint. The following observation can be shown.

**Proposition 1.** (*Stewart and Ojea Quintana, 2017, Proposition 2*) *Convex IP pooling functions are externally Bayesian.*

(The *propositions* in this paper are our results. When provided, proofs are relegated to the appendix.) And since Bayesian conditionalization is a special case of updating on a likelihood function, it follows that any IP pooling function (not necessarily convex) that is externally Bayesian also satisfies commutativity with conditionalization.

The fact that updating on a common likelihood generalizes updating on an event may serve to show that the assumption of a common likelihood function is not quite as strong as it may appear initially, since the conditionalization of the  $\mathbf{p}_i$  on some event drops out as a special case. That is, learning the same event is an instance of a shared likelihood function. It is also worth pausing to consider why Bayesians would deal in likelihood functions in the first place if updating with a likelihood function presents ways of learning not reducible to conditionalization.<sup>8</sup> One reason is that, under certain conditions, there is a way of regarding updating with a likelihood function as a case of Bayesian conditionalization by *refining* the algebra so that there is an event such that conditionalizing on it yields the same results as updating with the likelihood function on the coarser algebra. We return to this point—which is relevant to Jeffrey conditionalization and imaging as well—at the close of Section 5.

#### 4. COMMUTATIVITY

But why should it matter if a pooling method is externally Bayesian? More generally, why should we insist on the commutativity of learning and pooling? A few motivations, which we now briefly survey, are offered in the literature.

In introducing the external Bayesianity constraint, Madansky points out that the decisions of a group with common interests employing an externally Bayesian pooling operator will appear to outsiders as the decisions of a single Bayesian agent (1964). How? A Bayesian agent conditionalizes. So, given group opinion,  $F(\mathbf{p}_1, \dots, \mathbf{p}_n)$ , the updated group opinion should result from the group prior by conditionalization,  $F^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n)$ . If the group employs a pooling operator that is not externally Bayesian, and learning happens at the level of individuals, the posterior group opinion may not result from the prior group opinion by conditionalization:  $F(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda) \neq F^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n)$ . If the relevant learning happens at the level of group opinion, then the posterior group opinion may not be the result of applying the (non-externally Bayesian) pooling method that allegedly gives us the way of arriving at group opinion when applied to individual opinions: again,  $F(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda) \neq F^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n)$ .

Of course, there are a number of interpretations of the pool of individual opinions, including as “a rough summary” of the  $n$  pmfs (Wagner, 2009). The properties that are appropriate for a pooling function to exhibit depend on the interpretation

<sup>8</sup>Thanks to Paul Pedersen for emphasizing this point to us.

of the pool and the use to which it is put. When pooling is interpreted as a method of reaching either a compromise or a genuine consensus for use in group decision making, it may be important to ensure adherence of “group opinion” to norms of individual rationality. For views according to which groups can be agents subject to the same rationality conditions, for example, failing to satisfy external Bayesianity raises problems insofar as Bayesian conditionalization is rationally mandatory. More generally, those problems hold for failure of commutativity of pooling with any rule of learning held to be normatively compelling.

Furthermore, Russell et al. charge pooling methods that fail to commute with conditionalization with being subject to a diachronic Dutch book (2015). When the group posterior does not result from the group prior by conditionalizing on what the individuals learn, the conditions for a diachronic Dutch book are met. Echoing Raiffa (1968, pp. 221-226), Dietrich and List offer other strategic considerations in favor of external Bayesianity. If a pooling method is not externally Bayesian, collective opinion is open to manipulation. By disclosing relevant information at the appropriate time, someone could affect collective opinion by increasing the influence of certain opinions, for example (2014). There are, in short, possible manipulations besides those of a clever bookie.

Perhaps most uncontroversially, pooling operators for which learning and pooling commute save us the trouble of having to figure out whether updating should come before or after pooling, whether susceptibility to a diachronic book is damning for the pooling method, how and when to safeguard against manipulation, etc. In any event, the main position argued for in this paper can be understood as a conditional: *if* one finds commutativity of learning (of various types) and pooling compelling, *then* one has reasons to seriously consider IP pooling formats.

## 5. JEFFREY CONDITIONALIZATION

As indicated in the introduction, standard Bayesian conditionalization requires that the event conditionalized upon receives probability 1 in the posterior distribution. Jeffrey’s point is that not all learning experiences are like that. Sometimes observation leads to a revision in subjective probability even when there is no proposition (event)  $E$  that is learned “for certain.” Jeffrey’s famous candle light example serves to illustrate his point. Suppose you observe your friend’s coat, but only under candle light. The coat looks blue, but you are not quite sure. The impact of this observation is a shift in your subjective probabilities concerning the color of the coat, but none of the options goes to 1. This sort of scenario, some Jeffrey sympathizers claim, is “the normal case” (e.g., Spohn, 2012, p. 38). Improved lighting generally only shifts probabilities a bit more.

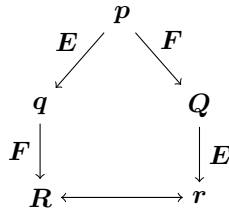
Let  $\mathbf{E} = \{E_i\}$  be a countable family of pairwise disjoint events partitioning  $\Omega$ . In the candle light example above, the partition of concern consists of the possible colors of the coat. A posterior,  $\mathbf{q}$ , comes from a prior,  $\mathbf{p}$ , by Jeffrey conditionalization by updating on the new probabilities for the cells of  $\mathbf{E}$ ,  $\mathbf{q}(E_i)$ , in the following way:

$$\mathbf{q}(A) = \sum_i \mathbf{q}(E_i) \mathbf{p}(A|E_i)$$

The  $\mathbf{q}(E_i)$  express the *direct* effect of an observation on subjective probabilities for the cells of the partition. When  $\mathbf{q}(E_i) = 1$  for some  $E_i$ , Jeffrey conditionalization reduces to standard Bayesian conditionalization.

A fact about Jeffrey conditionalization that many have found problematic is the failure of learning sequences to commute (Skyrms, 1986).

FIGURE 2. Commutativity of Updatings



Suppose that the learning experiences prompting the revision from  $p$  to  $q$  and the revision of  $Q$  to  $r$  are the same, as are those leading to the revision of  $q$  to  $R$  and  $p$  to  $Q$  (reflected in the updating partitions  $E$  and  $F$ , respectively). Jeffrey conditionalization can yield  $R \neq r$ . That is, switching the order in which a pair of observations is learned can yield different probabilities in the end. Van Fraassen complains:

Two persons, who have the same relevant experiences on the same day, but in a different order, will not agree in the evening even if they had exactly the same opinions in the morning. Does this not make nonsense of the idea of learning from experience? (1989, p. 338)

It is because of the issue of commutativity of learning experiences (as well as the nice off-the-shelf result of Wagner's (Theorem 3) we appeal to below) that we present here a particular parameterization of Jeffrey conditionalization intended to address the commutativity difficulty.

Hartry Field offers a fix, identifying conditions that are sufficient to ensure that, for finite partitions,  $R = r$  in Figure 2 above (Field, 1978). Wagner generalizes the result to countable partitions (Wagner, 2002). We introduce some useful notation. Where  $A$  and  $B$  are events and  $q$  is a revision of  $p$ , the *Bayes factor* is the ratio of new to old odds:

$$\mathcal{B}(q, p; A : B) = \frac{q(A)/q(B)}{p(A)/p(B)}$$

Instead of being reflected in identical *posteriors*, the proposal on the table is to understand identical learning as reflected in identical Bayes factors. Wagner points out that identifying identical learning with identical Bayes factors has a distinguished pedigree in Bayesian thinking (Good, 1983; Jeffrey, 2004).<sup>9</sup> What Field shows is that Jeffrey conditionalization is commutative when *identical learning* is interpreted as identical Bayes factors.

<sup>9</sup>Wagner contends that identical learning should be thought of as identical Bayes factors rather than identical posteriors. One alleged reason is that posteriors are tainted by the prior, whereas Bayes factors are an uncontaminated measure of the impact of the evidence. How do Bayes factors measure the impact of the evidence in isolation from the prior? Consider the case in which  $q$  comes from  $p$  by Bayesian conditionalization on  $E$ . Then,

$$q(A)/q(B) = \frac{p(A|E)}{p(B|E)}$$



**Theorem 2.** (Wagner, 2002, Theorem 3.1) Consider the revision scheme of Figure 2. If

$$\mathcal{B}(\mathbf{q}, \mathbf{p}; E_{i_1} : E_{i_2}) = \mathcal{B}(\mathbf{r}, \mathbf{Q}; E_{i_1} : E_{i_2}) \text{ for all } i_1, i_2$$

and

$$\mathcal{B}(\mathbf{R}, \mathbf{q}; F_{j_1} : F_{j_2}) = \mathcal{B}(\mathbf{Q}, \mathbf{p}; F_{j_1} : F_{j_2}) \text{ for all } j_1, j_2,$$

then  $\mathbf{R} = \mathbf{r}$ .

Wagner further shows that Jeffrey conditionalization has an equivalent parameterization in terms of Bayes factors (2009, Theorem 3.2).  $\mathbf{q}$  gives us posteriors for atomic events,  $b_k = \mathcal{B}(\mathbf{q}, \mathbf{p}; E_k : E_1)$ ,  $k = 1, 2, \dots$ , and  $[\omega \in E_k]$  is the characteristic function of the set  $E_k$ :

$$[\omega \in E_k] = \begin{cases} 1, & \text{if } \omega \in E_k \\ 0, & \text{otherwise.} \end{cases}$$

Wagner's parameterization, then, is the following.

$$\mathbf{q}(\omega) = \mathbf{p}_J^E(\omega) = \frac{\sum_k b_k \mathbf{p}(\omega) [\omega \in E_k]}{\sum_k b_k \mathbf{p}(E_k)}$$

There are two very nice features of this parameterization that are relevant. First, as we have been explaining, it responds to the complaints about commutativity because the result of a sequence of updates is invariant under permutations of that sequence when Jeffrey conditionalization is understood this way, with identical learning reflected in identical Bayes factors instead of identical posteriors.

The second nice feature, as Wagner shows, is that with his parameterization, we can articulate a version of commutativity with Jeffrey conditionalization that provides us with a characterization of externally Bayesian pooling operators in terms familiar to formal epistemologists and philosophers of science. We call Wagner's version of commutativity with Jeffrey  $CJC_W$ .

$CJC_W$ . For all partitions  $\mathbf{E} = \{E_k\}$  of  $\Omega$ , all profiles  $(\mathbf{p}_1, \dots, \mathbf{p}_n)$  in the domain of  $F$ , the Jeffrey update of the pool,  $F_J^E(\mathbf{p}_1, \dots, \mathbf{p}_n) = \frac{\sum_k b_k F(\mathbf{p}_1, \dots, \mathbf{p}_n) [\omega \in E_k]}{\sum_k b_k F(\mathbf{p}_1, \dots, \mathbf{p}_n)(E_k)}$ , is identical to  $F(\frac{\sum_k b_k \mathbf{p}_1 [\omega \in E_k]}{\sum_k b_k \mathbf{p}_1(E_k)}, \dots, \frac{\sum_k b_k \mathbf{p}_n [\omega \in E_k]}{\sum_k b_k \mathbf{p}_n(E_k)}) = F(\mathbf{p}_{1J}^E, \dots, \mathbf{p}_{nJ}^E)$ , the pool of the (Jeffrey updated) posteriors.<sup>10</sup>

and

$$\mathcal{B}(\mathbf{q}, \mathbf{p}; A : B) = \frac{\mathbf{p}(A|E)/\mathbf{p}(B|E)}{\mathbf{p}(A)/\mathbf{p}(B)}.$$

So,  $\mathcal{B}(\mathbf{q}, \mathbf{p}; A : B)$  is a measure of the change the evidence,  $E$ , induces in favor of  $A$  over  $B$ .  $\mathcal{B}(\mathbf{q}, \mathbf{p}; A : B)$  can also be rearranged using Bayes' theorem.

$$\frac{\mathbf{q}(A)}{\mathbf{q}(B)} = \frac{\mathbf{p}(A|E)}{\mathbf{p}(B|E)} = \frac{\frac{\mathbf{p}(A)\mathbf{p}(E|A)}{\mathbf{p}(E)}}{\frac{\mathbf{p}(B)\mathbf{p}(E|B)}{\mathbf{p}(E)}} = \frac{\mathbf{p}(A)\mathbf{p}(E|A)}{\mathbf{p}(B)\mathbf{p}(E|B)} = \frac{\mathbf{p}(A)}{\mathbf{p}(B)} \times \frac{\mathbf{p}(E|A)}{\mathbf{p}(E|B)}$$

Dividing now by  $\frac{\mathbf{p}(A)}{\mathbf{p}(B)}$ , the denominator of  $\mathcal{B}(\mathbf{q}, \mathbf{p}; A : B)$ , gives us

$$\mathcal{B}(\mathbf{q}, \mathbf{p}; A : B) = \frac{\mathbf{p}(E|A)}{\mathbf{p}(E|B)}$$

The quantity  $\mathbf{p}(E|A)/\mathbf{p}(E|B)$  is sometimes referred to as the *likelihood ratio*. So, the Bayes factor is a ratio of the non-prior quantities involved in Bayes' theorem, the quantities that revise the prior.

<sup>10</sup>Wagner's version of commutativity with Jeffrey conditionalization involves some additional technical assumptions. First, that  $\mathbf{p}_i(E_k) > 0$  for all  $i$  and all  $k$ . Second, that  $b_1 = 1$  and

Crucially, the Bayes factors,  $b_k$  for  $k = 1, 2, \dots$ , are held fixed across the  $\mathbf{p}_i$  (and also used in updating  $F(\mathbf{p}_1, \dots, \mathbf{p}_n)$ ). This is quite different from holding fixed a common posterior distribution,  $\mathbf{q}$ , in Jeffrey conditionalizing the  $\mathbf{p}_i$ . Put another way, he shows that External Bayesianity is equivalent to  $CJC_W$ .

**Theorem 3.** (*Wagner, 2009, Theorem 3.3*) *A (precise) pooling operator is externally Bayesian iff it satisfies  $CJC_W$ .*

We take  $\mathcal{F}_J^E(\mathbf{p}_1, \dots, \mathbf{p}_n)$  to be given by Jeffrey conditionalization of each element of  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  on the partition  $\mathbf{E}$ , holding fixed the Bayes factors  $b_k$  for  $k = 1, 2, \dots$  for  $\mathbf{p}_i$  for  $i = 1, \dots, n$ . A very slight mathematical generalization allows us to extend Wagner’s result to IP pooling functions in general.

**Proposition 2.** *Let  $\mathcal{F} : \mathbb{P}^n \rightarrow \mathcal{P}(\mathbb{P})$  be an IP pooling function (not necessarily convex).  $\mathcal{F}$  is externally Bayesian iff  $\mathcal{F}$  satisfies  $CJC_W$ .*

In particular, moving to the convex IP setting does not break the equivalence between external Bayesianity and commutativity with Wagner’s parameterization of Jeffrey conditionalization. Putting Propositions 1 and 2 together, we obtain the following.

**Proposition 3.** *Convex IP pooling satisfies  $CJC_W$ .*

We have the following immediate corollary (the proof is trivial given Proposition 3, and is omitted).

**Proposition 4.** *Convexity is preserved under Jeffrey conditionalization with common Bayes factors.*

To say that convexity is preserved means that, if we start with a convex set, we do not lose convexity in moving to the set of updated probability functions.

However, when  $\mathbf{q}_i$  comes from  $\mathbf{p}_i$  by standard Jeffrey conditionalization on some shared posterior distribution,  $\mathbf{q}$ , over a partition,  $\mathbf{E}$ , and the pool is updated likewise by updating each element of  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  on that same posterior distribution over  $\mathbf{E}$ , Jeffrey conditionalization and convex IP pooling do *not* commute.

**Proposition 5.** *Convex IP pooling does **not** commute with Jeffrey conditionalization on a common posterior.*

Instead of holding fixed common Bayes factors, here we assume a fixed posterior distribution on  $\mathbf{E}$ . Neither is commutativity of Jeffrey conditionalization so formulated and pooling guaranteed in the precise setting (Wagner, 2009, pp. 340-341). In particular, linear and geometric pooling fail to commute with Jeffrey in general. Furthermore, certain “objective” Bayesian updating methods, like minimizing the Kullback-Leibler divergence between posterior and prior, generalize Jeffrey conditionalization (Diaconis and Zabell, 1982). A corollary of Proposition 5, then, is that minimizing the Kullback-Leibler divergence does not commute with convex IP pooling. Minimizing the Kullback-Leibler divergence is also a generalization of Jaynes’ Maximum Entropy formalism (e.g., Williams, 1980). While there are many advocates of the Kullback-Leibler approach (e.g., Hartmann, 2014), even in

---

$\sum_k b_k \mathbf{p}_i(E_k) < \infty$  for  $i = 1, \dots, n$ . Third, where  $\mathbf{q}_i(\omega) = \frac{\sum_k b_k \mathbf{p}_i(\omega)[\omega \in E_k]}{\sum_k b_k \mathbf{p}_i(E_k)}$ , it is the case that  $0 < \sum_k b_k F(\mathbf{p}_1, \dots, \mathbf{p}_n)(E_k) < \infty$ . In the IP setting, this last assumption may be adjusted to be a requirement for each  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ .

the precise setting, a number of Bayesian-style objections to MaxEnt methods have been voiced in the literature (see, e.g., Seidenfeld, 1986; Gaifman and Vasudevan, 2012).

How much ground does Jeffrey conditionalization ultimately gain over standard Bayesian conditionalization? In certain cases—characterized by the so-called *superconditioning criterion*—Jeffrey conditionalization can be represented as Bayesian updating on *certain* evidence in a richer probability space (Diaconis and Zabell, 1982, Theorem 2.1). In fact, Diaconis and Zabell’s point is general, applying to any revision method, including imaging, so long as the superconditioning criterion holds. For finite algebras, the effect of Jeffrey conditionalization can always be so represented by Bayesian conditionalization (fn. 1, Wagner, 2002, p. 268).<sup>11</sup> Kyburg attributes an early, informal version of this result to Levi (Levi, 1967; Kyburg, 1987, Lemma A.5). The point, as Kyburg puts it, is not that we make effable the ineffable observational input motivating Jeffrey conditionalization: “This is not to say that we need to *specify* that evidence; it is that there is an algorithm by means of which the impact of the uncertain evidence can be represented as the impact of other ‘certain’ evidence” (Kyburg, 1987, p. 280). The obvious question is whether or when the superconditioning move is a philosophically legitimate one.

## 6. IMAGING

An hypothesis occupying the attention of many scholars working on the logic of conditionals asserts that the probability of a conditional is identical to the relevant conditional probability:  $\mathbf{p}(A \rightarrow B) = \mathbf{p}(B|A)$ . There are, of course, a number of ways to interpret the components of the claim. There are, for instance, various ways to define a conditional probability,  $\mathbf{p}(B|A)$ , just as there are various ways to specify an interpretation of a conditional connective,  $\rightarrow$ . Suppose we take conditional probability to be defined standardly as in Section 3. Is there an interpretation of  $\rightarrow$  such that the desired identity holds for all  $\mathbf{p}$ ? No, on pain of triviality, as Lewis proved (1976). That is, a conditional satisfying the identity exists only for trivial probability models. Similar triviality results hold for alternative readings of the identity. For example, “for any  $\mathbf{p}$  there exists some  $\rightarrow$  such that  $\mathbf{p}(A \rightarrow B) = \mathbf{p}(B|A)$ ” runs into similar problems. An impressive battery of such triviality results has been obtained for different ways of reading the identity. A helpful overview of much of the relevant literature can be found in (Hájek and Hall, 1994). Though it fails in general when formulated in terms of conditionalization, Lewis shows that a version of the identity holds if formulated in terms of *imaging* instead. We turn now to a brief presentation of imaging and Lewis’ possibility result.

Robert Stalnaker specified the semantics of the so-called *Stalnaker conditional*,  $>$ , in terms of possible worlds.  $\Omega$  is interpreted as a set of possible worlds.<sup>12</sup> Propositions are subsets of  $\Omega$ . We assume that  $\mathcal{A}$  is a  $\sigma$ -algebra of subsets of  $\Omega$ , the set of relevant propositions. For any  $\omega \in \Omega$ , let  $\omega_A$  be the “most similar” possible world at which  $A$  holds, the “closest”  $A$ -world to  $\omega$ . Say that  $A > B$  is

<sup>11</sup>In finite spaces, any revision method can be represented as conditionalization in a richer space *via* superconditioning provided the posterior probability is *absolutely continuous* with respect to the prior.

<sup>12</sup>A metaphysically deflationary conception of possible worlds has it that a possible world is just a maximally complete set of sentences in some propositional language, instead of a “possible totality of facts.”

true at  $\omega$  iff  $B$  is true at  $\omega_A$  (when the antecedent is impossible,  $A > B$  is taken to be vacuously true at all worlds). Lewis tailors a probabilistic revision scheme to the Stalnaker semantics.

For any non-empty  $E \in \mathcal{A}$ , imaging shifts the probability from each  $\omega' \in \Omega$  to its “image” atom,  $\omega \in E$ . If  $\omega' \in E$ , then  $\omega'$  is its own image atom. Lewis offers an interpretation in terms of possible worlds. On the assumption that for each world there is a unique “most similar” world in  $E$ , imaging can be thought of as the process of revising probabilities by shifting the total probability of each world to its most similar world in  $E$ . Relaxing the uniqueness assumption results in what is known as *general imaging*. General imaging allows the probability of each  $\omega' \in \Omega$  to be shifted to an image set, each element of which receives some fraction of the total probability of  $\omega'$ . Clearly, general imaging reduces to imaging when the image set is a singleton and the total probability of each  $\omega' \in \Omega$  is transferred to its image set.

Formally, we represent the relevant transfer of probability with a transfer function,  $T : \mathcal{A} \times \Omega \times \Omega \rightarrow [0, 1]$ , such that  $\sum_{\omega \in \Omega} T_E(\omega', \omega) = 1$  for all  $\omega' \in \Omega$ . For any  $E$  and all  $\omega, \omega' \in \Omega$ ,  $T_E(\omega', \omega)$  (times 100 percent) specifies the percentage of the total probability mass that is transferred from  $\omega'$  to  $\omega$ . It is sometimes assumed—e.g., by Lewis but not by Leitgeb—that  $\sum_{\omega \in E} T_E(\omega', \omega) = 1$ , so that  $E$  bears probability 1 after imaging on it. With  $T$  in place, we can formulate the recipe for general imaging. Say that  $\mathbf{q}$  comes from  $\mathbf{p}$  by *general imaging* if

$$\mathbf{q}(\omega) = \mathbf{p}(\omega||E) = \sum_{\omega' \in \Omega} \mathbf{p}(\omega') T_E(\omega', \omega)$$

The constraint on  $T$  of summing to 1 for each  $\omega'$  ensures that all probability mass is transferred, so no probability mass is created or destroyed, and the result of imaging is again a pmf. As before, the probability of an event  $A \in \mathcal{A}$  can be obtained by summing across  $\omega \in A$ . Lewis claims that conditionalization and imaging are both minimal revisions, but in different senses. While conditionalization “does not distort the profile of probability ratios, equalities, and inequalities among sentences that imply  $A$ ,” imaging “involves no gratuitous movement of probability from worlds to dissimilar worlds” (1976, p. 142). Lewis proves the following possibility result for (sharp) imaging and the probability of conditionals.

**Theorem 4.** (*Lewis, 1976, p. 142*) *The probability of a Stalnaker conditional with a possible antecedent is the probability of the consequent after imaging on the antecedent:  $\mathbf{p}(A > B) = \mathbf{p}(B||A) = \mathbf{q}(B)$ .*

More important for the purposes of the present paper is that, as Hannes Leitgeb observes, a result about general imaging due to Peter Gärdenfors can be restated in the language of pooling operators (2016). By *update mechanism*, Leitgeb means a function  $U : \mathbb{P} \times \mathcal{A} \rightarrow \mathbb{P}$  that maps a probability function and a (non-empty) proposition to a probability function. Gärdenfors shows that general imaging is the unique probabilistic revision method that preserves convex combinations of probability measures. Leitgeb repurposes this result, obtaining the following insight about pooling.

**Theorem 5.** (*Cf. Gärdenfors, 1982, Theorem 1*) *Update by general imaging (with respect to a fixed transfer function  $T$ ) is the unique update mechanism that commutes with linear pooling with respect to arbitrary coefficients.*

(Here, the transfer function,  $T$ , is invariant across priors.) Together with Leitgeb’s insight, Gärdenfors’ theorem makes showing the next result very easy.

**Proposition 6.** *Convex IP pooling commutes with general imaging.*

(Commutativity with imaging (*CI*) could be stated as an axiom for pooling operators.) An analogue of Proposition 4 follows immediately from Proposition 6: convexity is also preserved under general imaging.

Interest in imaging extends beyond natural language semantics and the philosophy of language. Part of the concern with identifying the probability of a conditional with the relevant conditional probability, after all, comes from attempts to give *acceptability* conditions for conditionals. Furthermore, some have argued that, while conditional probability represents *matter-of-fact* supposition in the context of probability, imaging represents *counterfactual* supposition.<sup>13</sup> Lewis himself thought that an interpretation like Stalnaker’s is right for *subjunctive* conditionals or counterfactuals, but not for indicative conditionals. Imaging finds crucial employment in James Joyce’s account of causal decision theory. Causal relationships are thought to be expressed by subjunctive conditionals, so the probability of such conditionals is of central concern on that view (1999). Baratgin and Politzer contend that empirical evidence indicates that general imaging has some claim as a description of actual revision of probabilistic judgment in dynamic environments (2010). Many authors, however, complain that a philosophically defensible interpretation of the requisite similarity relation among possible worlds has yet to be provided. Some see both the Stalnaker conditional and imaging as questionable shifts from epistemology to metaphysics (Arló-Costa, 2007). Moreover, beginning at least with Ramsey, an alternative line of research attempts to provide acceptability conditions for counterfactuals in terms of belief revision theory, eschewing construals of counterfactuals as bearing truth values (Levi, 1996). Nevertheless, imaging seems to have captured the fancy of many philosophers and others working on conditionals, counterfactual reasoning, and decision theory.

## 7. DISCUSSION

Propositions 3 and 4 have important implications outside of the context of opinion pooling. While IP models have been widely and convincingly advanced as superior to precise Bayesian representations of uncertainty, standard conditionalization *via* certain learning has been by and large retained as the relevant updating rule (Levi, 1978; Giron and Ríos, 1980). Proposition 4 shows that there is no *mathematical* necessity in that retention for convex Bayesians. For those compelled by Jeffrey’s vision of learning, they can have their convex sets of probabilities and their probability kinematics, too.<sup>14</sup> A similar point holds for imaging. Since convexity is preserved under imaging, imaging constitutes a possible “dynamics” for convex Bayesians.<sup>15</sup>

<sup>13</sup>Others, however, have offered more uniform accounts of supposition (e.g., Levi, 1996).

<sup>14</sup>Though, as Diaconis and Zabell’s aforementioned result shows us, in a range of cases there is no mathematical necessity in adopting Jeffrey conditionalization in order to obtain the results of Jeffrey conditionalization.

<sup>15</sup>Though it is not uncontroversial that conditionalization or some other type of updating of represents *learning*. Isaac Levi, for instance, writes, “All conditions of rationality are equilibrium conditions. In a sense they are synchronic conditions [...] Furthermore, in stating conditions of rational equilibrium, no prescription is made regarding the psychological path to be taken in

Furthermore, in the precise setting, only linear opinion pooling commutes with imaging. But linear opinion pooling does not commute with Bayesian conditionalization. It follows that no precise pooling method commutes with both imaging and Bayesian conditionalization. In this way, one’s hand is forced on the question of updating methods by commitments to pooling methods, and *vice versa*. Not so in the imprecise setting.

TABLE 1. Summary of Pooling and Updating Commutativity

	Linear	Geometric	Convex IP
External Bayesianity	X	✓	✓
CJC <sub>W</sub>	X	✓	✓
CI	✓	X	✓

Another way to put our point is that, if commutativity of learning and pooling is endorsed *and* more than one updating method is found acceptable (depending on context, say), then there may exist no accommodating precise pooling method.

Further limitations issue from results obtained in the literature. For example, suppose commutativity of pooling with both standard Bayesian conditionalization (or Jeffrey conditionalization) and marginalization is endorsed. In the precise setting, we are out of luck. Again, not so in the IP setting, as our results in conjunction with those of (Stewart and Ojea Quintana, 2017) attest. Philosophical positions that argue for or assume that pooling should be of a particular format are answerable for the limitations of those methods. For instance, in the epistemological debate about peer disagreement, a prominent position encourages peers to “split the difference” between their probabilistic opinions (Elga, 2007). So-called *conciliatory* views on disagreement generally counsel revising opinions in the direction of the dissenting opinions. The revision goes by equal-weight or near equal-weight linear pooling (Christensen, 2009). Some consequences of the failure of commutativity with conditionalization are highlighted in (Russell et al., 2015). As indicated above, Russell, et al. allege that a variant of a diachronic Dutch book can be made against parties following such a policy of disagreement resolution. Similar points can be made regarding other properties of particular pooling methods. For example, neither linear nor geometric pooling preserves probabilistic independence in general (Genest and Wagner, 1987), though convex IP pooling preserves Levi’s *confirmational irrelevance*, a generalization of probabilistic independence (Stewart and Ojea Quintana, 2017, Proposition 4). Seidenfeld, Schervish, and Kadane offer a decision-theoretic counterexample to the reasonableness of linear pooling on the basis of its failure to preserve independence (2010). Arguing similarly, Elkin and Wheeler present a variant of a Dutch book argument against resolving disagreements according to the equal weight view (2016). We submit that, not only are IP pooling functions more flexible formal tools, but they admit of stronger normative motivations when various prominent pooling criteria (including the commutativity criteria above) are taken as normative yardsticks.

---

moving from disequilibrium or from one equilibrium position to another. In other words, there are no norms prescribing rational learning processes” (Levi, 1970).

## APPENDIX: PROOFS

## PROOF OF PROPOSITION 2

*Proof.* We follow through Wagner's proof for the precise case (2009, Theorem 3.3), adapting it for IP where necessary.

( $\Rightarrow$ ) Assume that  $\mathcal{F}$  is externally Bayesian, i.e., for all profiles and any likelihood function,  $\mathcal{F}^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n) = \mathcal{F}(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda)$ . We want to show that, for all partitions  $\mathbf{E} = \{E_k\}$  of  $\Omega$  and all profiles in  $\mathbb{P}^n$ ,

$$\begin{aligned} \mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \dots, \mathbf{p}_n) &= \left\{ \frac{\sum_k b_k \mathbf{p}[\cdot \in E_k]}{\sum_k b_k \mathbf{p}(E_k)} : \mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n) \right\} \\ &= \mathcal{F} \left( \frac{\sum_k b_k \mathbf{p}_1[\cdot \in E_k]}{\sum_k b_k \mathbf{p}_1(E_k)}, \dots, \frac{\sum_k b_k \mathbf{p}_n[\cdot \in E_k]}{\sum_k b_k \mathbf{p}_n(E_k)} \right) \\ &= \mathcal{F}(\mathbf{p}_{1J}^{\mathbf{E}}, \dots, \mathbf{p}_{nJ}^{\mathbf{E}}) \end{aligned}$$

where the first and last equalities are definitional. Recall the definition of  $b_k$ :  $b_k = \mathcal{B}(\mathbf{q}, \mathbf{p}; E_k : E_1) = \frac{\mathbf{q}(E_k)/\mathbf{q}(E_1)}{\mathbf{p}(E_k)/\mathbf{p}(E_1)}$ ,  $k = 1, 2, \dots$ . Set  $\lambda(\omega) = \sum_k b_k [\omega \in E_k]$ . Wagner observes the following chain of equalities then obtains for  $\mathbf{p}_i, i = 1, \dots, n$  (2009, (3.10), p. 342):

$$(\star) \sum_{\omega \in \Omega} \lambda(\omega) \mathbf{p}_i(\omega) = \sum_{\omega \in \Omega} \mathbf{p}_i(\omega) \sum_k b_k [\omega \in E_k] = \sum_k b_k \sum_{\omega \in \Omega} \mathbf{p}_i(\omega) [\omega \in E_k] = \sum_k b_k \mathbf{p}_i(E_k)$$

Since each of the terms  $b_k \mathbf{p}_i(E_k)$  is positive and  $\sum_k b_k \mathbf{p}_i(E_k) < \infty$ ,  $\lambda$  is a likelihood function for  $\mathbf{p}_i, i = 1, \dots, n$ . Using  $(\star)$ , we can obtain

$$\mathcal{F}(\mathbf{p}_{1J}^{\mathbf{E}}, \dots, \mathbf{p}_{nJ}^{\mathbf{E}}) = \mathcal{F} \left( \frac{\mathbf{p}_1 \lambda(\cdot)}{\sum_{\omega' \in \Omega} \mathbf{p}_1(\omega') \lambda(\omega')}, \dots, \frac{\mathbf{p}_n \lambda(\cdot)}{\sum_{\omega' \in \Omega} \mathbf{p}_n(\omega') \lambda(\omega')} \right)$$

by substituting, for each  $i = 1, \dots, n$ ,  $\lambda(\cdot)$  for  $\sum_k b_k [\omega \in E_k]$  in the numerator and  $\sum_{\omega' \in \Omega} \mathbf{p}_i(\omega') \lambda(\omega')$  for  $\sum_k b_k \mathbf{p}_i(E_k)$  in the denominator. But by definition,

$$\mathcal{F} \left( \frac{\mathbf{p}_1 \lambda(\cdot)}{\sum_{\omega' \in \Omega} \mathbf{p}_1(\omega') \lambda(\omega')}, \dots, \frac{\mathbf{p}_n \lambda(\cdot)}{\sum_{\omega' \in \Omega} \mathbf{p}_n(\omega') \lambda(\omega')} \right) = \mathcal{F}(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda)$$

and by assumption  $\mathcal{F}(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda) = \mathcal{F}^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n)$ . By definition,  $\mathcal{F}^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n) = \{\mathbf{p}^\lambda : \mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)\}$ . But, for all  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ ,  $\mathbf{p}^\lambda = \frac{\sum_k b_k \mathbf{p}[\cdot \in E_k]}{\sum_k b_k \mathbf{p}(E_k)}$ . Hence,  $\mathcal{F}^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n) = \mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ . So,  $\mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \dots, \mathbf{p}_n) = \mathcal{F}(\mathbf{p}_{1J}^{\mathbf{E}}, \dots, \mathbf{p}_{nJ}^{\mathbf{E}})$  follows from the assumption.

( $\Leftarrow$ ) Suppose that  $\mathcal{F}$  satisfies  $CJC_W$  and that  $\lambda$  is a likelihood function for  $\mathbf{p}_i, i = 1, \dots, n$ . Let  $(\omega_1, \omega_2, \dots)$  be a list of all of those  $\omega \in \Omega$  such that  $\lambda(\omega) > 0$ , and let  $\mathbf{E} = \{E_1, E_2, \dots\}$ , where  $E_i := \{\omega_i\}$ . Setting  $b_k = \frac{\lambda(\omega_k)}{\lambda(\omega_1)}$  for  $k = 1, 2, \dots$ , it follows that  $b_k > 0$  and that  $b_1 = 1$ . Since  $\lambda$  is a likelihood for  $\mathbf{p}_i, i = 1, \dots, n$ , we have  $\sum_k b_k \mathbf{p}_i(E_k) < \infty, i = 1, \dots, n$ , and that  $(\mathbf{q}_1, \dots, \mathbf{q}_n) \in \mathbb{P}^n$ , where  $\mathbf{q}_i(\omega) := \frac{\sum_k b_k \mathbf{p}_i(\omega) [\omega \in E_k]}{\sum_k b_k \mathbf{p}_i(E_k)}$ . From  $CJC_W$ , it follows that 1)  $0 < \sum_k b_k \mathbf{p}(E_k) < \infty$  for all  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ , and that 2)  $\mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \dots, \mathbf{p}_n) = \mathcal{F}(\mathbf{p}_{1J}^{\mathbf{E}}, \dots, \mathbf{p}_{nJ}^{\mathbf{E}})$ . 1) implies that  $0 < \sum_{\omega \in \Omega} \lambda(\omega) \mathbf{p}(\omega) < \infty$  for all  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ , and 2) implies that  $\mathcal{F}^\lambda(\mathbf{p}_1, \dots, \mathbf{p}_n) = \mathcal{F}(\mathbf{p}_1^\lambda, \dots, \mathbf{p}_n^\lambda)$  (since substituting the definition of  $b_k$  in terms of  $\lambda$  in  $\frac{\sum_k b_k \mathbf{p}_i(\omega) [\omega \in E_k]}{\sum_k b_k \mathbf{p}_i(E_k)}$ ,

the formula for obtaining the  $\mathbf{q}_i$ , reduces that formula to the formula for updating on that  $\lambda$ ). □

### PROOF OF PROPOSITION 5

*Proof.* We provide a case in which convex IP pooling and Jeffrey conditionalization *as standardly construed* do not commute. Let  $\mathbf{q}_i$  come from  $\mathbf{p}_i$  by Jeffrey conditionalization, and let  $\mathbf{q}$  be a common posterior distribution over partition  $\mathbf{E}$  for  $\mathbf{p}_i$ ,  $i = 1, \dots, n$ . Let  $\mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  come from  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  by Jeffrey conditionalizing each  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  using  $\mathbf{q}$ , the common posterior distribution over  $\mathbf{E}$ . We offer a counterexample to commutativity in which  $\mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \dots, \mathbf{p}_n) \neq \mathcal{F}(\mathbf{q}_1, \dots, \mathbf{q}_n)$ .

Let  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ , and consider the following two pmfs.

TABLE 2. Priors

	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$
$\mathbf{p}_1$	1/4	1/4	1/4	1/4
$\mathbf{p}_2$	1/8	1/2	1/4	1/8

Let  $\mathbf{E} = \{E_1, E_2\}$  with  $E_1 = \{\omega_1, \omega_2\}$  and  $E_2 = \{\omega_3, \omega_4\}$  be a partition of  $\Omega$ . Jeffrey updating both pmfs using  $\mathbf{q}$ , where  $\mathbf{q}(E_1) = 2/3$  and  $\mathbf{q}(E_2) = 1/3$ , we obtain the following posteriors.

TABLE 3. Posteriors

	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$
$\mathbf{q}_1$	1/3	1/3	1/6	1/6
$\mathbf{q}_2$	2/15	8/15	2/9	1/9

Consider the .50 – .50 mixture of  $\mathbf{p}_1$  and  $\mathbf{p}_2$ ,  $\mathbf{p}^* = 0.5\mathbf{p}_1 + 0.5\mathbf{p}_2$ . It is clear that  $\mathbf{p}^* \in \mathcal{F}(\mathbf{p}_1, \mathbf{p}_2)$ . Jeffrey conditionalizing  $\mathbf{p}^*$  with  $\mathbf{q}$  gives us  $\mathbf{q}^*$ . In particular,  $\mathbf{q}^*(\omega_1) = 2/9$  and  $\mathbf{q}^*(\omega_3) = 4/21$ . It is clear that  $\mathbf{q}^* \in \mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \mathbf{p}_2)$ . Any  $\mathbf{q}_* \in \mathcal{F}(\mathbf{q}_1, \mathbf{q}_2)$  is of the form  $\mathbf{q}_* = \alpha\mathbf{q}_1 + (1 - \alpha)\mathbf{q}_2$  for  $\alpha \in [0, 1]$ .

Suppose that  $\mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \mathbf{p}_2) = \mathcal{F}(\mathbf{q}_1, \mathbf{q}_2)$ . Then, there is a  $\mathbf{q}_* \in \mathcal{F}(\mathbf{q}_1, \mathbf{q}_2)$  such that  $\mathbf{q}^* = \mathbf{q}_*$ . In particular,  $\mathbf{q}_*(\omega_1) = 2/9$  and  $\mathbf{q}_*(\omega_3) = 4/21$ . Letting  $\mathbf{q}_*(\omega_1) = 2/9$ , we can compute  $\alpha$ .

$$2/9 = \mathbf{q}_*(\omega_1) = \alpha\mathbf{q}_1(\omega_1) + (1 - \alpha)\mathbf{q}_2(\omega_1) = \alpha 1/3 + (1 - \alpha)2/15$$

Solving, we get  $\alpha = 4/9$ . However, we are supposed to have  $\mathbf{q}_*(\omega_3) = 4/21$ . For  $\alpha = 4/9$ , that is not the case.

$$\mathbf{q}_*(\omega_3) = \alpha\mathbf{q}_1(\omega_3) + (1 - \alpha)\mathbf{q}_2(\omega_3) = 4/9(1/6) + 5/9(2/9) = 16/81 > 4/21 = \mathbf{q}^*(\omega_3)$$

It follows that  $\mathcal{F}_J^{\mathbf{E}}(\mathbf{p}_1, \mathbf{p}_2) \neq \mathcal{F}(\mathbf{q}_1, \mathbf{q}_2)$ . □



## PROOF OF PROPOSITION 6

*Proof.* We want to show that  $\mathcal{F}(\mathbf{q}_1, \dots, \mathbf{q}_n) = \mathcal{F}_I^E(\mathbf{p}_1, \dots, \mathbf{p}_n)$ , where  $\mathbf{q}_i$  comes from  $\mathbf{p}_i$  by general imaging on  $E$ , and  $\mathcal{F}_I^E(\mathbf{p}_1, \dots, \mathbf{p}_n)$  comes from  $\mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  by general imaging each  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  on  $E$ . Again, we show both inclusions. In the proofs, we appeal to the fact any element of a convex set is some convex combination of the generating, extreme points: For any  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ ,  $\mathbf{p} = \sum_{i=1}^n \alpha_i \mathbf{p}_i$ , where  $\alpha_i \geq 0$  for  $i = 1, \dots, n$ , and  $\sum_{i=1}^n \alpha_i = 1$  (see, e.g., Stewart and Ojea Quintana, 2017, Lemma 1).

Let  $\mathbf{q} \in \mathcal{F}(\mathbf{q}_1, \dots, \mathbf{q}_n)$ . So,  $\mathbf{q} = \sum_{i=1}^n \alpha_i \mathbf{q}_i$ . Since  $\mathbf{q}$  is a linear pool of  $\mathbf{q}_i$  for  $i = 1, \dots, n$ , by Gärdenfors' result, Theorem 5,  $\mathbf{q}$  is also the result of imaging  $\mathbf{p} = \sum_{i=1}^n \alpha_i \mathbf{p}_i$  on  $E$ , because linear pooling and general imaging commute. Since  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ , it follows that  $\mathbf{q} \in \mathcal{F}_I^E(\mathbf{p}_1, \dots, \mathbf{p}_n)$ .

For the other direction, assume that  $\mathbf{q} \in \mathcal{F}_I^E(\mathbf{p}_1, \dots, \mathbf{p}_n)$ . So,  $\mathbf{q}$  is the result of general imaging some  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$  on  $E$ . For any  $\mathbf{p} \in \mathcal{F}(\mathbf{p}_1, \dots, \mathbf{p}_n)$ ,  $\mathbf{p} = \sum_{i=1}^n \alpha_i \mathbf{p}_i$ . By Gärdenfors' result,  $\mathbf{q} = \sum_{i=1}^n \alpha_i \mathbf{q}_i$ , where the  $\mathbf{q}_i$  come from the  $\mathbf{p}_i$  by general imaging on  $E$ , because general imaging and linear pooling commute. But then it follows that  $\mathbf{q} \in \mathcal{F}(\mathbf{q}_1, \dots, \mathbf{q}_n)$ . □

## REFERENCES

- Arló-Costa, H. (2007). The logic of conditionals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2014 ed.).
- Baratgin, J. and G. Politzer (2010). Updating: A psychologically basic situation of probability revision. *Thinking & Reasoning* 16(4), 253–287.
- Blackwell, D. and L. Dubins (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 882–886.
- Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass* 4(5), 756–767.
- de Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg and H. E. Smoklery (Eds.), *Studies in Subjective Probability*. Wiley.
- Diaconis, P. and S. L. Zabell (1982). Updating subjective probability. *Journal of the American Statistical Association* 77(380), 822–830.
- Dietrich, F. and C. List (2014). Probabilistic opinion pooling. In A. Hájek and C. Hitchcock (Eds.), *Oxford Handbook of Probability and Philosophy*. Oxford University Press.
- Elga, A. (2007). Reflection and disagreement. *Noûs* 41(3), 478–502.
- Elkin, L. and G. Wheeler (2016). Resolving peer disagreements through imprecise probabilities. *Noûs*.
- Field, H. (1978). A note on jeffrey conditionalization. *Philosophy of Science*, 361–367.
- Gaifman, H. and M. Snir (1982). Probabilities over rich languages, testing and randomness. *The Journal of Symbolic Logic* 47(03), 495–548.
- Gaifman, H. and A. Vasudevan (2012). Deceptive updating and minimal information methods. *Synthese* 187(1), 147–178.
- Gärdenfors, P. (1982). Imaging and conditionalization. *The Journal of Philosophy*, 747–760.
- Genest, C. (1984). A characterization theorem for externally bayesian groups. *The Annals of Statistics*, 1100–1105.

- Genest, C., K. J. McConway, and M. J. Schervish (1986). Characterization of externally bayesian pooling operators. *The Annals of Statistics*, 487–501.
- Genest, C. and C. G. Wagner (1987). Further evidence against independence preservation in expert judgement synthesis. *Aequationes Mathematicae* 32(1), 74–86.
- Genest, C. and J. V. Zidek (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 114–135.
- Girón, F. J. and S. Ríos (1980). Quasi-bayesian behaviour: A more realistic approach to decision making? *Trabajos de Estadística y de Investigación Operativa* 31(1), 17–38.
- Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. U of Minnesota Press.
- Hájek, A. and N. Hall (1994). The hypothesis of the conditional construal of conditional probability. *Probability and Conditionals: Belief Revision and Rational Decision*, 75.
- Hartmann, S. (2014). A new solution to the problem of old evidence. In *Philosophy of Science Association 24th Biennial Meeting*, Chicago, IL.
- Herron, T., T. Seidenfeld, and L. Wasserman (1997). Divisive conditioning: Further results on dilation. *Philosophy of Science*, 411–444.
- Huttegger, S. M. (2015). Merging of opinions and probability kinematics. *The Review of Symbolic Logic* 8(04), 611–648.
- Jeffrey, R. (2004). *Subjective Probability: The Real Thing*. Cambridge University Press.
- Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.
- Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 79–86.
- Kyburg, H. E. (1987). Bayesian and non-bayesian evidential updating. *Artificial Intelligence* 31(3), 271–293.
- Kyburg, H. E. and M. Pittarelli (1992). Some problems for convex bayesians. In *Proceedings of the Eighth international conference on Uncertainty in artificial intelligence*, pp. 149–154. Morgan Kaufmann Publishers Inc.
- Leitgeb, H. (2016). Imaging all the people. *Episteme* (DOI: 10.1017/epi.2016.14).
- Levi, I. (1967). Probability kinematics. *British Journal for the Philosophy of Science*, 197–209.
- Levi, I. (1970). Probability and evidence. In M. Swain (Ed.), *Induction, Acceptance, and Rational Belief*, pp. 134–156. New York: Humanities Press.
- Levi, I. (1978). Irrelevance. In C. Hooker, J. Leach, and E. McClellan (Eds.), *Foundations and Applications of Decision Theory*, Volume 1, pp. 263–273. Boston: Springer.
- Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, MA.
- Levi, I. (1985). Consensus as shared agreement and outcome of inquiry. *Synthese* 62(1), pp. 3–11.
- Levi, I. (1990). Pareto unanimity and consensus. *The Journal of Philosophy* 87(9), 481–492.
- Levi, I. (1996). *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference and Nonmonotonic Reasoning*. Cambridge University Press.
- Levi, I. (2009). Why indeterminate probability is rational. *Journal of Applied Logic* 7(4), 364–376.

- Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review* 85, 297–315.
- Madansky, A. (1964). Externally bayesian groups. Santa Monica, CA: RAND Corporation.
- Nau, R. F. (2002). The aggregation of imprecise probabilities. *Journal of Statistical Planning and Inference* 105(1), 265–282.
- Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis* 79(6), 1305–1342.
- Raiffa, H. (1968). Decision analysis: Introductory lectures on choices under uncertainty.
- Ramsey, F. P. (1931). Truth and probability. *The Foundations of Mathematics and Other Logical Essays*, 156–198.
- Russell, J. S., J. Hawthorne, and L. Buchak (2015). Groupthink. *Philosophical Studies* 172(5), 1287–1309.
- Savage, L. (1972, originally published in 1954). *The Foundations of Statistics*. New York: John Wiley and Sons.
- Schervish, M. and T. Seidenfeld (1990). An approach to consensus and certainty with increasing evidence. *Journal of Statistical Planning and Inference* 25(3), 401–414.
- Seidenfeld, T. (1986). Entropy and uncertainty. *Philosophy of Science*, 467–491.
- Seidenfeld, T., J. B. Kadane, and M. J. Schervish (1989). On the shared preferences of two bayesian decision makers. *The Journal of Philosophy* 86(5), 225–244.
- Seidenfeld, T., M. J. Schervish, and J. B. Kadane (2010). Coherent choice functions under uncertainty. *Synthese* 172(1), 157–176.
- Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *The Annals of Statistics* 21(3), 1139–1154.
- Skyrms, B. (1986). *Choice and Chance: An Introduction to Inductive Logic* (3rd Edition ed.). Belmont: Wadsworth Publishing Company.
- Spohn, W. (2012). *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford University Press, USA.
- Stewart, R. T. and I. Ojea Quintana (2017). Probabilistic opinion pooling with imprecise probabilities. *Journal of Philosophical Logic*, doi:10.1007/s10992-016-9415-9.
- van Fraassen, B. C. (1989). *Laws and Symmetry*. Clarendon Press Oxford.
- Wagner, C. (2002). Probability kinematics and commutativity. *Philosophy of Science* 69(2), 266–278.
- Wagner, C. (2009). Jeffrey conditioning and external bayesianity. *Logic Journal of IGPL* 18(2), 336–345.
- Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, 131–144.