Pooling, Products, and Priors

Richard Pettigrew
University of Bristol

Jonathan Weisberg
University of Toronto

Abstract

We often learn the opinions of others without hearing the evidence on which they're based. The orthodox Bayesian response is to treat the reported opinion as evidence itself and update on it by conditionalizing. But sometimes this isn't feasible. In these situations, a simpler way of combining one's existing opinion with opinions reported by others would be useful, especially if it yields the same results as conditionalization. We will show that one method—upco, also known as multiplicative pooling—is specially suited to this role when the opinions you wish to pool concern hypotheses about chances. The result has interesting consequences: it addresses the problem of disagreement between experts; and it sheds light on the social argument for the uniqueness thesis.

We often hear opinions without getting to hear the evidence behind them. Researchers report conclusions without sharing the underlying data; news stories omit testimony and statistics they relied on; and acquaintances share impressions, the basis for which they've long since forgotten. How should we modify our own opinions in these cases?

The orthodox Bayesian response is to treat the reported opinion as evidence itself, and update in the usual way: simply conditionalize on the fact that so-and-so holds such-and-such opinion. But sometimes this policy isn't feasible. We might not have the requisite priors, or we might have them but lack the cognitive wherewithal to calculate the corresponding posteriors. Or, we might be designing software that can't afford the time for the full computation.

In these situations, a simpler way of combining one's existing opinion with opinions reported by others would be useful. Especially if that method yields the same results conditionalization would.

We will show that one method—upco, also known as multiplicative pooling—is specially suited to this role when the opinions you wish to pool concern hypotheses about chances. Upco effectively aggregates the evidence behind opinions about chances, at least in typical cases. So using upco to fold someone else's opinion into your own is equivalent to conditionalizing on the evidence behind their opinion.
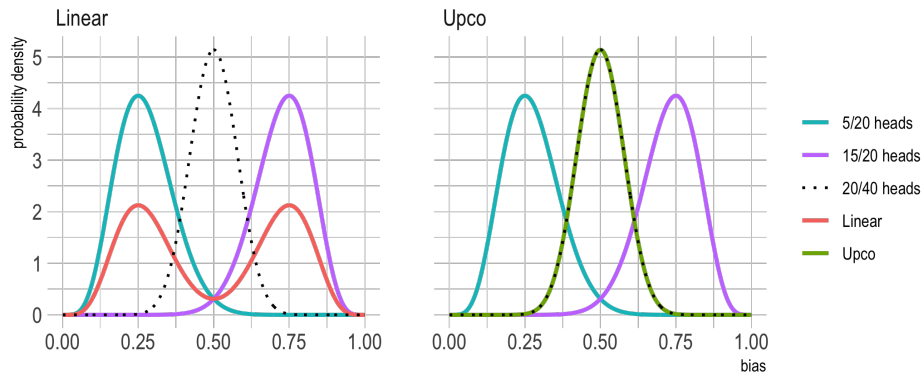
Figure 1: When pooling over hypotheses about the bias of a coin, linear pooling (red) has undesirable results, while upco (green) aggregates evidence.

This result has interesting consequences. It helps to address the problem of how to defer to experts who might disagree, as we show in Section 5. And it sheds light on the social argument for the uniqueness thesis, as we show in Section 6.

1 Background

If you assign to some proposition H the probability $P(H)$, and someone else reports a different probability $Q(H)$, a natural thought is to split the difference. That is, you might take the midpoint

$$\frac{P(H) + Q(H)}{2}$$

as your new probability for H. This is known as *linear pooling*. Linear pooling is intuitive and simple, but often gives undesirable results.

To illustrate, suppose you and a friend are interested in a coin of unknown bias. You both begin with a uniform prior over the $[0, 1]$ interval. Then, separately, you each perform 20 flips of the coin in private. Suppose you get 5 heads and they get 15. Then your posterior over the coin's bias will be the blue curve in the left panel of Figure 1, and theirs will be the purple curve. Combining these posteriors by linear pooling gives the camel shaped curve in red.

This is quite different from conditionalizing on the evidence behind your friend's posterior. That would yield the dotted curve in black instead. That's the distribution you'd get by conditionalizing your prior on the aggregate evidence, $5 + 15 = 20$ heads out of 40 flips total.[1]

---

[1]Winkler (1968) also makes this point: see his Figure 2.

2

How can we combine the blue and purple curves to get the desired, dotted curve? By multiplying instead of adding. Rather than add $Q(H)$ to your $P(H)$ and divide by 2 to renormalize, multiply $P(H)$ by $Q(H)$, then renormalize.

The renormalization step is a bit subtler now. It will depend on just which opinions Q shares with you. If you only learn their opinion about H and its negation $\overline{H}$, then the total amount of pre-normalization probability is $P(H)Q(H) + P(\overline{H})Q(\overline{H})$. So you must divide by this sum to renormalize. This makes your new opinion about H:

$$\frac{P(H)Q(H)}{P(H)Q(H) + P(\overline{H})Q(\overline{H})}.$$

We will use the notation $PQ(H)$ for this new opinion, as a mnemonic for its multiplicative origin.

In general, when Q shares their opinions over a countable[2] partition $\{H_i\}$, your new opinion about each H_i will be:

$$PQ(H_i) = \frac{P(H_i)Q(H_i)}{\sum_j P(H_j)Q(H_j)}.$$

This way of combining opinions is known as *multiplicative pooling*, or *upco*. We'll often write PQ for the distribution over $\{H_i\}$ that it generates.

Addition and multiplication are both simple, familiar functions that increase with both arguments. But linear pooling ends up being simpler than upco, because the denominator is always 2. Since the sum of probabilities over a partition is always 1, summing the terms $P(H_i) + Q(H_i)$ over any partition $\{H_i\}$ always yields the same value, 2. Whereas the sum of products $P(H_i)Q(H_i)$ varies depending on the partition.

This might seem like a strike against upco. But upco turns out to have many desirable properties, a number of which are laid out by Easwaran et al. (2016).[3] Our purpose in this section is to draw out another desirable feature, one that emerges when the H_i are chance hypotheses—about the bias of a coin, for example.

In the right-hand panel of Figure 1, upco combines the blue and purple curves to give the desired green curve. More generally, it effectively conditionalizes P's posterior on Q's data no matter how many heads and tails each has seen.[4] For example, in Figure 2, P's posterior is based on only 10 flips, while Q's is based on 20. The dashed curve is the posterior for their aggregate evidence, and the upco curve in green coincides perfectly.

---

[2]In the continuous case, the sum becomes an integral and probabilities become densities.

[3]But see Mulligan (2021) for criticism, and an alternate approach drawing on Genest and Schervish (1985).

[4]Winkler (1968, B64–5) and Morris (1983, Section 6) make similar observations; see also Babic et al. (manuscript).
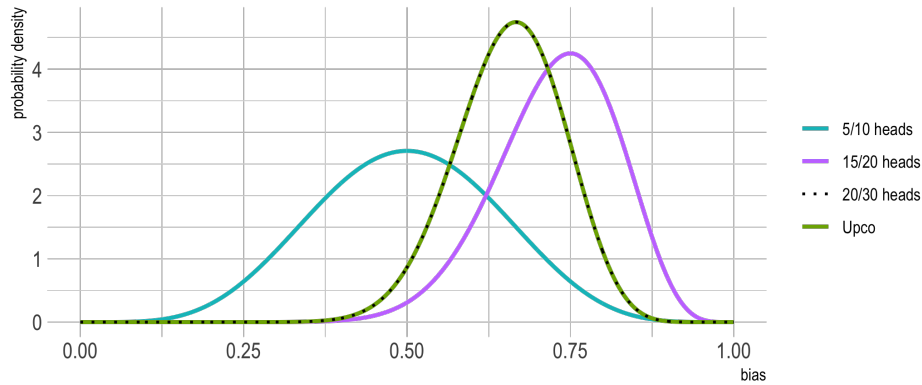
Figure 2: Upco works even when one agent has more evidence, e.g. 20 *observations vs.* 10.

How general is this feature of upco? When can it be used to effectively aggregate evidence? To a first approximation the answer is: when the H_i are chance hypotheses that render P's evidence independent of Q's. But this answer needs to be developed and refined. The next three sections undertake this development. Later sections then use the results to illuminate further questions.

2 A First Result

Two features of the coin tossing example contribute to upco's success. The first is that Q had a uniform prior over $\{H_i\}$. Though we'll see how to do without this assumption later. The second, more essential feature is that tosses are independent once we specify the coin's true bias.

In the general case, the evidence being aggregated can be anything. The important thing is that we can think of the H_i as chance hypotheses according to which P's evidence is independent of Q's. That is, each H_i posits a chance function C_i such that $C_i(EF) = C_i(E)C_i(F)$, where E and F are the bodies of evidence gathered by P and Q, respecively. Assuming P and Q defer to these chances per the Principal Principle (Lewis, 1980), the following two conditions hold:

$$P(EF \mid H_i) = P(E \mid H_i)P(F \mid H_i), \tag{1}$$
$$P(F \mid H_i) = Q(F \mid H_i). \tag{2}$$

When these conditions hold, and Q's prior is uniform, P can use upco to effectively conditionalize on Q's evidence.

We'll use the shorthand P_E for P's posterior. In other words, P_E is the probability function defined by $P_E(-) = P(- \mid E)$. Likewise Q_F is Q's posterior: $Q_F(-) = Q(- \mid F)$. In this notation, the upco of P's and Q's

4

posteriors is denoted $P_E Q_F$. The formal statement of our first result is then the following (see the Appendix for all proofs):

Proposition 1. *Let Q be uniform over a partition* $\{H_i\}$ *such that* (1) *and* (2) *hold for all* H_i. *Then for all* H_i, $P_E Q_F(H_i) = P(H_i \mid EF)$.

Informally speaking, using upco to combine P and Q's posteriors is equivalent to conditionalizing P's posterior on Q's evidence, assuming (i) a uniform prior for Q, and (ii) chance hypotheses that render P and Q's data independent.

If we think of E and F as the outcomes of separate experiments, then assumption (ii) is natural, and common in actual practice. Chance hypotheses typically posit independent and identically distributed data, as in the coin tossing example we began with. But whether data are discrete or continuous, i.i.d. outcomes are a standard modeling assumption.

Even when models don't posit identically distributed data, they often still posit independence. For example, a climate model might predict different average observed temperatures from year to year. Still, conditional on a full specification of the model's parameters, the data gathered by one team of researchers measuring temperatures in one year will be independent of the data gathered by another team measuring temperatures the following year.

What about assumption (i) though? What if Q's prior isn't uniform over $\{H_i\}$? We'll generalize Proposition 1 to handle this case below. But first we need to establish some useful properties of upco, which we'll use repeatedly in the rest of the paper. The next section lays out these properties, then the following section applies them to the case of a non-uniform prior for Q.

3   The Algebra of Upco

When we introduced upco, we chose the notation PQ to evoke multiplication. In this section we'll push the multiplication analogy further. We'll see that we really can think of upco as a product operation, multiplying one distribution P by another Q, to give a new distribution PQ. This product operation obeys the same algebraic laws as the familiar multiplication operation on numbers, *e.g.* it is commutative and associative. And, crucially, this same product operation also captures updating by conditionalization.

Looking at the definition of upco from earlier, it's fairly straightforward to verify that $PQ = QP$ for any P and Q. In other words, upco is a commutative operation. With a bit more work, we can further verify that upco is associative too. That is, whether we combine P with Q and then with R, or first combine Q and R and then with P, the result is the same: $P(QR) = (PQ)R$. (Again, see the Appendix for all proofs.)

When multiplying numbers, the value 1 has a special role: multiplying by 1 has no effect, $x \cdot 1 = x$. The[5] uniform distribution behaves similarly under upco: pooling an arbitrary P with the uniform distribution just returns P. That is, $PU = P$, where U is uniform over $\{H_i\}$.[6] In the terminology of algebra, U is the identity element for the upco operation.

Another key fact about multiplying numbers is that, as long as x is nonzero, it has an inverse. That is, there exists a number $x^{-1} = 1/x$ such that $x \cdot x^{-1} = 1$. Again, something similar is true for upco. As long as P is "regular," it has an inverse. That is, if P assigns no zeros over $\{H_i\}$, then there is another distribution P^{-1} such that $PP^{-1} = U$. In fact, this inverse is obtained by associating with each H_i the value $1/P(H_i)$, and then renormalizing.[7]

So upco induces a genuine algebra on probability distributions. Like multiplication for numbers, upco "multiplies" distributions in a way that is commutative, associative, possesses an identity element (the uniform distribution), and provides an inverse to every nonzero distribution.

This would all be just a neat bit of abstraction, but for one further fact. Crucially, conditionalization is the very same product operation as upco. Conditionalizing P on E is equivalent to taking the upco of P's prior distribution over $\{H_i\}$, and another distribution corresponding to P's likelihood function, $P(E \mid -)$.

We will write E_P for the normalized likelihood function of E according to P. That is, E_P is the following probability distribution over $\{H_i\}$:

$$E_P(H_i) = \frac{P(E \mid H_i)}{\sum_j P(E \mid H_j)}. \tag{3}$$

Where P is the prior distribution over $\{H_i\}$, and P_E the posterior, the crucial equivalence between conditionalization and upco is captured by the equation:[8]

$$P_E = PE_P.$$

This tells us that P's posterior over $\{H_i\}$ can be factored into a prior distribution and a likelihood distribution. Which is important, because these factored terms can then be moved around thanks to commutativity and

[5]Many different probability functions can be uniform over $\{H_i\}$, but they all share the same distribution over $\{H_i\}$. So we can speak of "the" uniform distribution. The Appendix handles these matters more rigorously, but we allow ourselves some sloppiness in the main text for readability.

[6]Technically PU is only defined over $\{H_i\}$, while P may be defined over a larger algebra. Again, we handle this rigorously in the Appendix, but permit some slack here to ease the exposition.

[7]So $P^{-1}(H_i) = P(H_i)^{-1} / \sum_j P(H_j)^{-1}$.

[8]Strictly speaking, it's the restriction of P_E to $\{H_i\}$ that's equal to PE_P. But again, we permit ourselves some slack here, leaving a fully rigorous treatment for the Appendix.

associativity, and even canceled in some cases thanks to the existence of inverses.

But first, let's pause to summarize these properties of upco's algebra.

Proposition 2. *Fix a partition $\{H_i\}$ and write PQ for the upco of P and Q over $\{H_i\}$. Let U be uniform over $\{H_i\}$, and let P, Q, and R be arbitrary. Then*

(a) $PQ = QP$,

(b) $P(QR) = (PQ)R$,

(c) $PU = P$,

(d) $PP^{-1} = U$, *provided $P(H_i) > 0$ for all H_i so that P^{-1} is well-defined, and*

(e) $P_E = PE_P$, *where E_P is given by equation* (3).

In the next section, we'll use these properties to solve the epistemological problem that P faced at the end of Section 1.

4   When Q is Not Uniform

Recall where we left things at the end of Section 2. If Q's prior was uniform over $\{H_i\}$, then, when P pools their posterior with Q's posterior using upco, this is equivalent to conditionalizing on Q's evidence, by Proposition 1. The problem we left off with was, what if Q's prior wasn't uniform? Can P still use upco to acquire Q's evidence?

There are two cases to consider. If P knows what Q's prior was, then a simple adjustment to the upco calculation used in Proposition 1 solves the problem. But if P doesn't know Q's prior, things are trickier. P can still use upco to acquire Q's evidence, but only if they take Q's prior seriously, in a certain sense we'll explain below. But let's handle the easy case first.

Suppose that P does know what Q's prior was. Then all they have to do is include its inverse Q^{-1} in their upco calculation, to cancel out the offending prior Q. That is, in addition to "multiplying" their posterior P_E by Q's posterior Q_F, they must also multiply by Q^{-1}. Then the algebraic properties developed in Section 3, together with assumptions (1) and (2), deliver:[9]

$$P_E Q_F Q^{-1} = PE_P QF_Q Q^{-1} = PE_P F_Q = PE_P F_P = P_{EF}.$$

In other words, taking the upco of $P(- \mid E)$, $Q(- \mid F)$, and Q^{-1} is equivalent to conditionalizing P's prior on the aggregate evidence EF.

---

[9]The first equality uses property (e) from Proposition 2; the second uses (a), (c), and (d); the third uses assumption (2); and the last combines assumption (1) with property (e).

Theorem 3. *Let* $\{H_i\}$ *be a partition such that, for all* H_i, *conditions* (1) *and* (2) *hold and* $Q(H_i) > 0$. *Then for all* H_i,

$$P_E Q_F Q^{-1}(H_i) = P(H_i \mid EF).$$

Notice that this theorem has Proposition 1 as a special case. When Q is uniform, so is Q^{-1}, so the Q^{-1} term drops out.

This solution does require some extra computation, but much less than it first appears. To obtain $P(H_i \mid EF)$, it looks like P must first take the upco of their posterior with Q's posterior, then calculate the inverse of Q's prior, and then upco with that. But actually, the following much simpler calculation is equivalent:

$$\frac{P_E(H_i)Q_F(H_i)/Q(H_i)}{\sum_j P_E(H_j)Q_F(H_j)/Q(H_j)}.$$

So the only real cost is an extra division operation for each H_i. Otherwise, the computation is identical to the case where Q's prior is uniform.

Now let's turn to the more difficult case: suppose P does not know what Q's prior was. Then, P can only apply upco to the posteriors $P(- \mid E)$ and $Q(- \mid F)$. But if we re-run the same calculation we did in the case where Q was known, just without the Q^{-1} term included, we find that this yields

$$P_E Q_F = (PQ)(EF)_P.$$

This says that taking the upco of the posteriors is still equivalent to conditionalizing on the aggregate evidence EF. Except that the prior being conditionalized isn't P, but PQ—the upco of P's prior with Q's.[10]

Theorem 4. *Let* $\{H_i\}$ *be a partition such that* P *and* Q *satisfy conditions* (1) *and* (2). *Then for all* H_i,

$$P_E Q_F(H_i) = PQ(H_i \mid EF).$$

Like Theorem 3, this theorem also has Proposition 1 as a special case. Now the reason is that, if Q is uniform, then $PQ = P$.

Informally, this theorem says that upco and conditionalization commute, given assumptions (1) and (2). Taking the upco of P's and Q's posteriors is equivalent to first taking the upco of their priors, and then conditionalizing on the aggregate evidence.

So when P doesn't know Q's prior, they must compromise with Q to acquire their evidence via upco. Rather than conditionalizing their own

---

[10]Strictly speaking, PQ is only defined over the partition $\{H_i\}$: it's only a partial probability function, which can't be conditioned on EF. But it's straightforward to extend it using the Principal Principle. Each H_i specifies a chance $C_i(EF)$, which serves as the likelihood term $PQ(EF \mid H_i)$ in Bayes' theorem. So we will talk as if $PQ(H_i \mid EF)$ is defined.

prior on the aggregate evidence, upco will first combine their prior with Q's, and then conditionalize on EF.

This compromise can be desirable, however. Often we aren't just interested in someone's opinion because they have some evidence that we don't. We may also think their interpretation of the evidence applies some insight, which our interpretation misses out. In Hall's (2004) terminology, P may partially defer to Q because they have some *analyst* expertise, not merely *database* expertise.

For us the interesting case is where this partial deference takes the following form: P's prior over $\{H_i\}$, conditional on Q's prior over $\{H_i\}$, is precisely PQ. In other words, if P had learned that Q initially assigned to $H_1, \ldots, H_n$ the values $q_1, \ldots, q_n$, then P would have assigned to each H_i the value dictated by upco:

$$P(H_i \mid Q(H_1) = q_1, \ldots, Q(H_n) = q_n) = PQ(H_i). \qquad (4)$$

If P's prior was deferential in this way, then adopting $PQ(H_i \mid EF)$ is equivalent to conditionalizing $P(- \mid E)$ on two further pieces of information: F, and the fact that Q's prior over $\{H_i\}$ was $\{q_i\}$. So upco again gives the desired result, when P partially defers to Q in this way.

But when would P partially defer to Q's judgment in this way? When is equation (4) plausible? When P views Q's analyst expertise as equivalent to possessing some (perhaps tacit) knowledge P lacks. Suppose that Q's prior can be represented as the conditionalization of a uniform prior Q_0 on some proposition F_0. Then, the upco of P and Q is the upco of $P(- \mid E_0)$ and $Q_0(- \mid F_0)$, where E_0 is a tautology. And Proposition 1 tells us this is equivalent to conditionalizing P on F_0. So if P thinks that Q's analyst expertise derive from knowledge of some such proposition F_0, they will partially defer to Q as in equation (4). Because doing so amounts to learning whatever information Q (tacitly) knows that gives them their expertise.

Notice that F_0 needn't be actual evidence acquired by Q, in the traditional sense. It may describe observations that Q has never actually made. It's enough that P thinks Q's analyst expertise derive from reasoning as if they had made those observations. For example, in our coin tossing example from earlier, Q may have no actual experience with this particular coin, but might have experience with other, similar coins, which leads them to begin with a certain, non-uniform prior. For example, they might begin with a fairly strong suspicion that this coin is fair. The resulting prior might be equivalent to a uniform prior conditionalized on, say, 10 observed flips evenly split between heads and tails.

<center>*   *   *</center>

Let's take stock. We've been considering the virtues of upco, as a way of responding to opinions about "models." That is, we've been thinking of the

<center>9</center>

H_i as chance hypotheses, according to which your evidence is independent of your interlocutor's.

We've seen that upco is equivalent to conditionalizing on your interlocutor's evidence, in two cases. First, when their prior over the partition was uniform, and second, when you know what their prior was and can thus "cancel" it using its inverse. But if their prior isn't uniform and you don't know what it was, you can't correct for it in this way.

However, if you partially defer to their judgment, upco can still deliver the desired result. Applying upco is equivalent to conditionalizing on both your interlocutor's prior and their evidence, provided your partial deference to their judgment has the right form. And, we've argued, this kind of partial deference is plausible in a range of cases.

In the remainder of the paper, we apply these results to other areas: expert disagreement in Section 5, and the uniqueness thesis in Section 6.

5   Serving Two Epistemic Masters

When experts differ, we laypeople face a conundrum. What opinion should we adopt as our own, given that there is no consensus opinion among the experts? It's tempting again to split the difference: to pool the experts' opinions linearly. Surprisingly, this turns out to be untenable.

Suppose you regard Q and R as experts about some proposition H. That is, if you learn Q's opinion, you will adopt it as your own, and likewise for R's opinion. The following two conditions hold then, where Q and R are now random variables representing these experts' opinions about H:

$$P(H \mid Q = q) = q, \tag{5}$$
$$P(H \mid R = r) = r. \tag{6}$$

If your policy is to split the difference should they differ, then we also have:

$$P(H \mid Q = q, R = r) = (q + r)/2. \tag{7}$$

But Dawid, DeGroot and Mortera (1995) show that these three conditions together imply $P(Q = R) = 1$.[11] Thus, to defer to Q and R individually, yet resolve any differences by linear pooling, you must be certain there won't be any differences to begin with.

In fact, Zhang (manuscript) shows that this result doesn't just hold for linear pooling, but for a large class of pooling rules. Assuming the domain of P is finite, it holds for any *strictly convex* pooling rule, *i.e.* any rule that always returns a value strictly between q and r (unless $q = r$). For example, the red curve in Figure 1 always lies strictly in between the blue and purple curves, because linear pooling is strictly convex.

---

[11]See also Bradley (2018) and Gallow (2018).

Formally, we are replacing (7) with the more general

$$P(H \mid Q = q, R = r) = f(q, r), \tag{8}$$

where f is any function that returns a number strictly between q and r when $q \neq r$, and returns q otherwise. Zhang shows that equations (5), (6) and (8) again imply $P(Q = R) = 1$, assuming P's domain is finite. So you can only plan to resolve any difference between Q and R by a strictly convex pooling rule if you are certain no such difference will arise.[12]

But upco is not strictly convex, as we can see from Figure 2: upco's green curve leaves the envelope enclosed by the blue and purple curves. And, indeed, upco can "serve two epistemic masters,"[13] reconciling the opinions of differing experts. In fact it does so in a significant range of cases, which we can identify using our earlier results.

Let's start with an example. Suppose a coin has two possible biases, described by the hypotheses H and $\overline{H}$. And suppose three agents all begin with the same prior P, which for now we'll assume is uniform over $\{H, \overline{H}\}$. One of these agents will flip the coin some number of times, and conditionalize on the result to arrive at a posterior we'll label Q. Another agent will perform a separate sequence of flips, arriving at R. The third agent, who so far still holds P, will then learn Q's and R's opinions about H.

If P knows these are the circumstances, then equation (5) will hold. For P, learning Q's opinion is equivalent to learning how many heads and tails they observed. And since P and Q share a common prior, P will draw the same conclusion from this information that Q did, *i.e.* adopt Q's opinion as their own. For exactly parallel reasons, equation (6) will hold too.

What about when P learns both experts' opinions? This is equivalent to learning how many heads and tails they observed between them. So P is effectively conditionalizing on the aggregate evidence. And we know from Proposition 1 that this is equivalent to taking the upco of Q's and R's posterior opinions. Thus

$$P(H \mid Q = q, R = r) = QR(H). \tag{9}$$

Now, crucially, it's entirely possible that Q and R will get different numbers of heads, and thus report different opinions. So $P(Q = R) \neq 1$ in this example. Thus upco is capable of serving two epistemic masters: equations (5), (6) and (9) do not imply $P(Q = R) = 1$.

How general is this result? Quite general. The hypotheses and evidence can be anything really. P doesn't even need to be able to infer what Q's and

[12]This rules out several popular alternatives to linear pooling, notably geometric and harmonic pooling. While both of these alternatives are non-convex when pooling over a partition with three or more elements, they are strictly convex in the present case, where the partition $\{H, \overline{H}\}$ has only two elements.

[13]To borrow Gallow's (2018) phrase.

R's evidence was exactly, only that they acquired some evidence that warrants the reported opinions. The main thing for upco to be appropriate is the kind of conditional independence assumption we made in equation (1). The hypotheses *H* and $\overline{H}$ need to render *Q* and *R*'s evidence independent.

For instance, continue to assume our three agents begin with a common prior, *P*. One will learn the true element of some partition $\{E_i\}$, another the true element of a partition $\{F_j\}$. The third agent, who still holds *P*, knows all this, so they defer to *Q* and *R* as in (5) and (6). Now, for upco to be appropriate, we must assume that *Q*'s evidence is independent of *R*'s, conditional on each hypothesis. That is, for every E_i and F_j,

$$P(E_iF_j \mid H) = P(E_i \mid H)P(F_j \mid H),$$

and similarly given $\overline{H}$. Then, if *P* is uniform over $\{H, \overline{H}\}$, *P* will resolve any differences according to upco, *i.e.* (9) holds.

We can drop the uniform prior assumption much as we did in Section 4, by including its inverse. Somewhat ironically, this means that *P* must include the inverse of their own opinion, P^{-1}, in their upco calculation. This is because *P* is also the prior behind both *Q* and *R*'s opinions, and we don't want it to be "double counted." Combining *Q*'s and *R*'s posteriors in the present case amounts to combining PE_P with PF_P:

$$PE_PPF_P = P^2E_PF_P = P^2(EF)_P.$$

When *P* was uniform, we had $P^2 = P$ so there was no issue. But if *P* is not uniform, then $P^2 \neq P$ and we need to include a P^{-1} to cancel one of the *P*'s.

Bottom line: even in the case of a non-uniform prior, *P* can still resolve any difference between *Q*'s and *R*'s opinions by upco. They just have to include the inverse of the shared prior, P^{-1}. Our main result for this section is then formally stated as follows:

Theorem 5. *Let $\{E_i\}$ and $\{F_j\}$ be finite partitions. Let Q be a random variable that takes the value $P(H \mid E_i)$ in the event that E_i, and let $R = P(H \mid F_j)$ in the event F_j. Then (5) and (6) hold. If, furthermore, each pair E_i, F_j is conditionally indepenent given the elements of $\{H, \overline{H}\}$, then*

$$P(H \mid Q = q, R = r) = P^{-1}QR(H). \tag{10}$$

In the special case $P(H) = P(\overline{H})$, then (10) reduces to (9).

This result generalizes straightforwardly to partitions $\{H_i\}$ with more than two cells, as we show in the Appendix.

We could generalize further, dropping the assumption of a shared prior. But this would no longer be a case where *Q* and *R* must be experts *P* defers to in the sense of (5) and (6). While they would have strictly more evidence than *P*, their interpretation isn't necessarily one that *P* would endorse. *P*

might still defer to them partially, along the lines of equation (4). In which case upco would still apply, similar to Theorem 4. But we won't pursue the details here. Instead, we turn to another application of our earlier results.

6   The Social Argument for Uniqueness

According to the *uniqueness thesis*, there is always just one correct way to interpret a body of evidence (Feldman, 2006). No two agents with the same total evidence should ever disagree. The alternative to this view is *permissivism*, which holds that agents with the same evidence may have different views, at least in some cases.

In the Bayesian framework, uniqueness is equivalent to there being a privileged prior. Two conditionalizers who begin with the same prior and get the same evidence must end up with the same posteriors. Bayesians who believe in a privileged prior that all rational agents must begin with are called *objective* Bayesians.

We've seen that there is a special relationship between upco and a certain prior: the uniform prior over chance hypotheses. And, indeed, this prior has a long history in the objective Bayesian tradition. It was used by Laplace (1774 [1986]) to derive the Rule of Succession in his classic response to Hume's problem of induction. And it recurs in more general forms in other classic works by De Morgan, Johnson, and Carnap (see Zabell, 1989, for an overview of the history).

Our results above thus suggest a kind of social argument for the uniqueness thesis. A uniform prior over chance hypotheses is privileged, making it possible to aggregate evidence easily, without having to share that evidence explicitly. Instead, we can report the conclusions drawn from our evidence, and apply upco. So each of the members of a community gains an epistemic advantage if all members adopt that prior. They can glean one another's evidence from their conclusions, without having to know one another's evidence or priors.

In fact, an argument along the same basic lines is advanced by Dogramaci and Horowitz (2016), just without the Bayesian formal specifics. Dogramaci and Horowitz argue that uniqueness enables a division of epistemic labour. Assuming uniqueness, agents can gather and process evidence in parallel, then come together to share their conclusions. If everyone interprets evidence the same way, then they can just adopt one another's conclusions without having to hear the supporting evidence. But not if interpretations differ. Then they can't take one another's conclusions on board unquestioningly. Instead, they must either communicate all their evidence, or do their best to guess what evidence lies behind one another's opinions.

Two key questions are left open by Dogramaci and Horowitz's informal

presentation of their argument. We'll explain how these questions are answered by our results above. But we'll also see that these answers raise a third issue, which may weaken the social argument for uniqueness.

The first question is how to deal with defeating evidence. To illustrate, suppose you and a friend part ways to collect evidence. When you meet back up, your friend believes H. Assuming uniqueness, you would be obliged to believe H too, if you had their evidence. However, it might be that their evidence, when combined with yours, no longer supports H. So, even assuming uniqueness, you can't just take someone's testimony that H on board on the grounds that they are rational. That would amount to disregarding your own evidence in favour of theirs.

This problem is solved by upco in the Bayesian framework, assuming the conditions of Proposition 1 are satisfied. Then, two agents who begin with a uniform prior over $\{H, \overline{H}\}$ will effectively conditionalize on one another's evidence by applying upco. So even if your evidence favoured $\overline{H}$ while theirs favoured H, you can take their testimony on board without having to hear what their evidence was. Upco will ensure that your ultimate conclusion reflects both bodies of evidence correctly.

The second question left open by the informal argument is: what is the uniquely correct way of interpreting evidence? In Bayesian terms, what is the uniquely privileged prior? The informal argument suggests only that there is value in everyone using the same prior. But it leaves open the possibility that any choice of a common prior is as good as any other—that the choice of prior is a mere matter of convention, like choosing which side of the road everyone will drive on.

But if upco is used to reintegrate the epistemic labour divided among the members of the community, then one prior stands out as a particularly good choice: the uniform prior over chance hypotheses. After all, as we have seen, if the shared prior is the uniform one, we need only apply upco to our posterior and our testifier's; if the shared prior is something else, we need to apply the inverse of that shared prior as well to undo the double counting. So, in the Bayesian framework, Proposition 1 gives a definite answer to the second question.

So far we've seen how Proposition 1 strengthens the social argument for uniqueness. Now for the bad news: even with a uniform prior, upco has undesirable results when pooling repeatedly.

To see the problem, suppose two agents begin with a uniform prior over $\{H_i\}$ and part ways to gather evidence. One learns E, the other learns F, and they meet to upco their posteriors. Assuming the conditions of Proposition 1 hold, the result for both agents is

$$U E_P F_Q = E_P F_P = (EF)_P.$$

Now suppose our two agents part ways again, this time learning E' and F', respectively. Then they meet up once more. Before they meet, one agent

14

has $(EF)_P E'_P$ and the other has $(EF)_P F'_Q$. So when they upco, the result will be

$$(EF)_P E'_P (EF)_P F'_Q = (EF)_P^2 (E'F')_P.$$

Notice, the evidence EF from the first round of investigation has been double-counted! In terms of coin tossing, it's as if each agent had done twice as many flips in their first experiment as they actually did. The sample size from the first round has been inflated to twice its true size.

Now, we know how to correct for double-counting from Section 5. We can include $(EF)_P^{-1}$ in the second round of pooling, to cancel one of the extra $(EF)_P$ terms. But there's a significant cost to this solution. It doesn't just entail some extra computation, it also requires agents to keep track of what they've learned from one another. For example, P needs to remember that, the last time they pooled with Q, Q walked away "carrying a copy" of $(EF)_P$ with them. Only then will P know to include $(EF)_P^{-1}$ the next time, to avoid double-counting. So this solution makes significant storage demands.

Is it unduly demanding though? We aren't sure. Plausibly, we do keep track of the information we learn from one another, in some form or other. When someone expresses an opinion on one occasion, we don't treat it as fresh news when we hear them express the same opinion later. So it may be plausible that humans do something like what this account demands.

But even so, we suspect this weakens the social argument for uniqueness significantly. It suggests that the division of epistemic labour isn't as efficient as it first seems. If the uniform-over-chances prior enables a division of epistemic labour that is more efficient than what we would have otherwise, but not *that* much more efficient, its rational privilege is less stark.

7 Conclusion

Upco is well-suited to pooling opinions about chance hypotheses. When the chances render your evidence independent of your interlocutor's, you can use upco to effectively acquire their evidence. Since chance theories often render data independent, this result has broad application. Inquirers who gather data separately can effectively communicate their observations just by reporting their conclusions.

This result is also useful for reconciling disagreements beetween experts. As Zhang shows, neither linear pooling nor any other strictly convex rule can be used to combine the opinions of differing experts. But we've seen that upco can be used, when the experts' opinions concern chance hypotheses that, again, render their data independent.

Finally, the social argument for the uniqueness thesis is simultaneously bolstered and challenged by these results. Upco works especially well as a

way to aggregate evidence when the prior over chance hypotheses is uniform. This suggests that the classic, objective Bayesian approach initiated by Laplace serves an important social-epistemic function. At the same time though, upco is liable to double-count evidence when pooling repeatedly with the same person. This effect can be corrected for using inverses, but at the cost of a significant storage burden.

So upco has broad application for learning from the opinions of others, and it delivers unexpected insights in related areas. We conclude that upco is a valuable and illuminating pooling rule, meriting equal consideration alongside more standard options like linear pooling.

8   Appendix

Here we give formal statements and proofs of the results in the main text. For simplicity we state all results concerning only two agents, but all generalize to larger groups in the natural way.

Throughout, let $\{H_i\}$ be a finite partition of size n, and let P, Q, and R be probability functions. Associate with P the vector $\mathbf{p} = (p_1, \ldots, p_n)$ whose entries are $p_i = P(H_i)$. Likewise let $\mathbf{q}$ have entries $q_i = Q(H_i)$, and $\mathbf{r}$ the entries $r_i = R(H_i)$. Note that $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$ are *probability vectors*, *i.e.* their entries are nonnegative and sum to 1.

We'll write $\mathbf{p}_E$ for the probability vector with entries $P(H_i \mid E)$, $\mathbf{p}_{EF}$ for the vector with entries $P(H_i \mid EF)$, and so on. We'll also write $\mathbf{e}_P$ for the normalized likelihoods of E according to probability function P:

$$(\mathbf{e}_P)_i = \frac{P(E \mid H_i)}{\sum_j P(E \mid H_j)}.$$

Similarly, $\mathbf{f}_Q$ is the normalized likelihood distribution of F according to Q. Finally, we'll let G be a shorthand for EF, so that $\mathbf{g}_P$ denotes the normalized likelihood distribution of EF according to P:

$$(\mathbf{g}_P)_i = \frac{P(EF \mid H_i)}{\sum_j P(EF \mid H_j)}.$$

The upco of two probability functions can be viewed as a product operation on the associated vectors.

Definition 1 (Upco product)**.** *The upco product of $\mathbf{p}$ and $\mathbf{q}$ is defined*

$$\mathbf{pq} = (p_1 q_1, \ldots, p_n q_n)/\mathbf{p} \cdot \mathbf{q}.$$

This operation is defined as long as $\mathbf{p} \cdot \mathbf{q} > 0$, in which case it always returns another probability vector. If $\mathbf{p}$ and $\mathbf{q}$ are *regular*, meaning their entries are all positive, then $\mathbf{pq}$ is also regular.

We now give a formal statement and proof of Proposition 2.

Proposition 2 (formal). *Suppose* $\mathbf{p} \cdot \mathbf{q} \cdot \mathbf{r} > 0$, *so that* $\mathbf{pq}$, $(\mathbf{pq})\mathbf{r}$, *and* $\mathbf{p}(\mathbf{qr})$ *are defined. Let* $\mathbf{u} = (1/n, \ldots, 1/n)$, *and if* $\mathbf{p}$ *is regular let*

$$\mathbf{p}^{-1} = \frac{1}{\sum_i 1/p_i} \left(1/p_1, \ldots, 1/p_n\right).$$

Then

(a) $\mathbf{pq} = \mathbf{qp}$,

(b) $\mathbf{p}(\mathbf{qr}) = (\mathbf{pq})\mathbf{r}$,

(c) $\mathbf{pu} = \mathbf{p}$,

(d) $\mathbf{pp}^{-1} = \mathbf{u}$ *if* $\mathbf{p}$ *is regular, and*

(e) $\mathbf{p}_E = \mathbf{pe}_P$.

Proof. Part (a) follows immediately from the commutativity of scalar multiplication and of dot products.

For part (b), compare the ith entries:

$$(\mathbf{p}(\mathbf{qr}))_i = \frac{p_i \frac{q_i r_i}{\mathbf{q} \cdot \mathbf{r}}}{\mathbf{p} \cdot (\mathbf{qr})} = \frac{p_i q_i r_i}{(\mathbf{p} \cdot \mathbf{qr})(\mathbf{q} \cdot \mathbf{r})},$$

$$((\mathbf{pq})\mathbf{r})_i = \frac{\frac{p_i q_i}{\mathbf{p} \cdot \mathbf{q}} r_i}{(\mathbf{pq}) \cdot \mathbf{r}} = \frac{p_i q_i r_i}{(\mathbf{p} \cdot \mathbf{q})(\mathbf{pq} \cdot \mathbf{r})}.$$

In both cases the ith entry is proportional to $p_i q_i r_i$. Since probability distributions with identical proportions are identical, $\mathbf{p}(\mathbf{qr}) = (\mathbf{pq})\mathbf{r}$, as desired.

For (c), the ith entry of $\mathbf{pu}$ is:

$$(\mathbf{pu})_i = \frac{p_i(1/n)}{\sum_j p_j(1/n)} = \frac{p_i}{\sum_j p_j} = p_i.$$

For (d), first observe that $\mathbf{p}^{-1}$ is a probability vector because

$$\sum_i \frac{1}{p_i \sum_j 1/p_j} = \frac{1}{\sum_j 1/p_j} \sum_i \frac{1}{p_i} = 1.$$

Moreover, the ith entry of $\mathbf{pp}^{-1}$ is proportional to

$$p_i \frac{1}{p_i \sum_j 1/p_j} = \frac{1}{\sum_j 1/p_j}.$$

So the entries of $\mathbf{pp}^{-1}$ are constant, hence must be $1/n$.

Finally, for (e), by definition $p_i = P(H_i)$ and $(\mathbf{e}_P)_i \propto P(E \mid H_i)$. So $(\mathbf{pe}_P)_i \propto P(H_i)P(E \mid H_i)$. By Bayes' theorem, $q_i \propto P(H_i)P(E \mid H_i)$ as well. So $\mathbf{q}$ and $\mathbf{pe}_P$ have the same proportions, hence must be identical. ∎

We will have several occasions to use the fact that conditional independence implies $\mathbf{g}_P = \mathbf{e}_P \mathbf{f}_P$.

Proposition 6. *If condition* (1) *holds for all H_i, then $\mathbf{g}_P = \mathbf{e}_P \mathbf{f}_P$.*

Proof. The entries of $\mathbf{e}_P$ are proportional to the $P(E \mid H_i)$, and the entries of $\mathbf{f}_P$ are proportional to the $P(F \mid H_i)$. So the entries of $\mathbf{e}_P \mathbf{f}_P$ are proportional to $P(E \mid H_i)P(F \mid H_i) = P(EF \mid H_i)$, hence to $\mathbf{g}_P$. ∎

Next we prove Theorem 3, which we restate here for convenience.

Theorem 3 (restatement). *Suppose that for all H_i, conditions* (1) *and* (2) *hold and $Q(H_i) > 0$. Then for all H_i,*

$$P_E Q_F Q^{-1}(H_i) = P(H_i \mid EF).$$

Proof. By condition (1), $\mathbf{e}_P \mathbf{f}_P = \mathbf{g}_P$. And by (2), $\mathbf{f}_Q = \mathbf{f}_P$. So

$$(\mathbf{p}\mathbf{e}_P)(\mathbf{q}\mathbf{f}_Q)\mathbf{q}^{-1} = (\mathbf{p}\mathbf{q}\mathbf{q}^{-1})(\mathbf{e}_P \mathbf{f}_Q) = \mathbf{p}(\mathbf{e}_P \mathbf{f}_P) = \mathbf{p}\mathbf{g}_P = \mathbf{p}_{EF}.$$

Since the distribution on the left gives the $P_E Q_F Q^{-1}(H_i)$ values, and the entries of $\mathbf{p}_{EF}$ are the $P(H_i \mid EF)$ values, this completes the proof. ∎

Now we prove Theorem 4, which we also restate for convenience.

Theorem 4 (restatement). *Suppose that for all H_i, conditions* (1) *and* (2) *hold. Then for all H_i,*

$$P_E Q_F(H_i) = PQ(H_i \mid EF).$$

Proof. By condition (1), $\mathbf{e}_P \mathbf{f}_P = \mathbf{g}_P$. And by (2), $\mathbf{f}_Q = \mathbf{f}_P$. So

$$(\mathbf{p}\mathbf{e}_P)(\mathbf{q}\mathbf{f}_Q) = (\mathbf{p}\mathbf{q})(\mathbf{e}_P \mathbf{f}_Q) = (\mathbf{p}\mathbf{q})(\mathbf{e}_P \mathbf{f}_P) = (\mathbf{p}\mathbf{q})(\mathbf{e}\mathbf{f})_P.$$

The left hand side gives the values of $P_E Q_F(H_i)$, and the right gives the $PQ(H_i \mid EF)$ values. So this completes the proof. ∎

We now turn to Theorem 5, which we'll prove in the more general form of Theorem 8. The proof really has two separate pieces, one that depends on the specifics of upco and a second which has nothing to do with upco. The first is quick using upco's algebra.

Lemma 7. *Let $Q(-) = P(- \mid E)$ and $R(-) = P(- \mid F)$. If* (1) *holds and $P(H_i) > 0$ for all H_i, then*

$$P(H_i \mid EF) = P^{-1}QR(H_i).$$

Proof. By hypothesis, $\mathbf{q} = \mathbf{p}\mathbf{e}_P$, $\mathbf{r} = \mathbf{p}\mathbf{f}_P$, $\mathbf{g}_P = \mathbf{e}_P \mathbf{f}_P$, and $\mathbf{p}^{-1}$ exists. So,

$$\mathbf{p}\mathbf{g}_P = \mathbf{p}\mathbf{e}_P \mathbf{f}_P = (\mathbf{p}^{-1}\mathbf{p})\mathbf{p}\mathbf{e}_P \mathbf{f}_P = \mathbf{p}^{-1}(\mathbf{p}\mathbf{e}_P)(\mathbf{p}\mathbf{f}_P) = \mathbf{p}^{-1}\mathbf{q}\mathbf{r}.$$

Since the left hand side gives the $P(H_i \mid EF)$ values, and the right gives the $P^{-1}QR(H_i)$ values, the proof is complete. ∎

18

In the main text we stated Theorem 5 in terms of a two-cell partition $\{H, \overline{H}\}$. Theorem 8 addresses the more general case, where $\{H_i\}$ may have more than two elements.

Theorem 8. *Let $\{H_i\}$, $\{E_j\}$, and $\{F_k\}$ be finite partitions. Let $\mathbf{Q}$ be a random vector, whose i^{th} value when E_j obtains is $P(H_i \mid E_j)$, and let $\mathbf{R}$ be a random vector whose i^{th} value when F_k obtains is $P(H_i \mid F_k)$. Then*

$$P(H_i \mid \mathbf{Q} = \mathbf{q}) = q_i, \tag{11}$$
$$P(H_i \mid \mathbf{R} = \mathbf{r}) = r_i. \tag{12}$$

If, furthermore, $P(E_j F_k \mid H_i) = P(E_j \mid H_i) P(F_k \mid H_i)$ for all i, j, k, then

$$P(H_i \mid \mathbf{Q} = \mathbf{q}, \mathbf{R} = \mathbf{r}) = P^{-1} Q R(H_i). \tag{13}$$

Proof. Note that, with Lemma 7 proved, the remaining work has nothing to do with upco. The operative idea is just that P can infer Q and R's evidence from their opinions, or near enough.

Let $E_{\mathbf{q}}$ be the union of all E_j's such that $P(H_i \mid E_j) = q_i$ for all i. And let $F_{\mathbf{r}}$ be the union of all F_k's such that $P(H_i \mid F_k) = r_i$:

$$E_{\mathbf{q}} = E_{\mathbf{q}_1} \cup \ldots \cup E_{\mathbf{q}_m},$$
$$F_{\mathbf{r}} = E_{\mathbf{r}_1} \cup \ldots \cup E_{\mathbf{r}_n}.$$

Since $\mathbf{Q} = \mathbf{q}$ is equivalent to $E_{\mathbf{q}}$, and $\mathbf{R} = \mathbf{r}$ to $F_{\mathbf{r}}$, we have for all i:

$$P(H_i \mid \mathbf{Q} = \mathbf{q}) = P(H_i \mid E_{\mathbf{q}}) = P(H_i \mid E_{\mathbf{q}_1}) = q_i,$$
$$P(H_i \mid \mathbf{R} = \mathbf{r}) = P(H_i \mid F_{\mathbf{r}}) = P(H_i \mid F_{\mathbf{r}_1}) = r_i,$$

establishing (11) and (12).

Now observe that $\mathbf{Q} = \mathbf{q}, \mathbf{R} = \mathbf{r}$ is equivalent to $\bigcup_{x,y} (E_{\mathbf{q}_x} \cap F_{\mathbf{r}_y})$. So

$$P(H_i \mid \mathbf{Q} = \mathbf{q}, \mathbf{R} = \mathbf{r}) = P(H_i \mid \bigcup_{x,y} (E_{\mathbf{q}_x} \cap F_{\mathbf{r}_y})) = P(H_i \mid E_{\mathbf{q}_1} \cap F_{\mathbf{r}_1}).$$

By conditional independence then, Lemma 7 implies that for all H_i,

$$P(H_i \mid \mathbf{Q} = \mathbf{q}, \mathbf{R} = \mathbf{r}) = P^{-1} Q R(H_i),$$

as desired. ■

References

Babic, Boris, Anil Gaba, Ilia Tsetlin and Robert L. Winkler. manuscript. "Resolute and Correlated Bayesians.".

Bradley, Richard. 2018. "Learning From Others: Conditioning Versus Averaging." *Theory and Decision* 85(1):5–20.

Dawid, Alexander P., Morris H. DeGroot and Julia Mortera. 1995. "Coherent Combination of Experts' Opinions." *Test* 4(2):263–313.

Dogramaci, Sinan and Sophie Horowitz. 2016. "An Argument for Uniqueness About Evidential Support." *Philosophical Issues* 26(1):130–47.

Easwaran, Kenny, Luke Fenton-Glynn, Christopher Hitchcock and Joel D. Velasco. 2016. "Updating on the Credences of Others: Disagreement, Agreement, and Synergy." *Philosophers' Imprint* 6(11):1–39.

Feldman, Richard. 2006. Reasonable Religious Disagreements. In *Philosophers Without Gods: Meditations on Atheism and the Secular Life*. Oxford University Press pp. 194–214.

Gallow, J. Dmitri. 2018. "No One Can Serve Two Epistemic Masters." *Philosophical Studies* 175(10):2389–2398.

Genest, Christian and Mark J. Schervish. 1985. "Modeling Expert Judgments for Bayesian Updating." *Annals of Statistics* 13(3):1198–1212.

Hall, Ned. 2004. "Two Mistakes About Credence and Chance." *Australasian Journal of Philosophy* 82(1):93–111.

Laplace, Pierre Simon. 1774 [1986]. "Memoir on the Probability of the Causes of Events." *Statistical Science* 1(3):359–363.

Lewis, David. 1980. A Subjectivist's Guide to Objective Chance. In *Studies in Inductive Logic and Probability*, ed. Richard C. Jeffrey. Vol. II University of California Press.

Morris, Peter A. 1983. "An Axiomatic Approach to Expert Resolution." *Management Science* 29(1):24–32.

Mulligan, Thomas. 2021. "The Epistemology of Disagreement: Why Not Bayesianism?" *Episteme* 18(4):587–602.

Winkler, Robert L. 1968. "The Consensus of Subjective Probability Distributions." *Management Science* 15(2):B61–75.

Zabell, Sandy L. 1989. "The Rule of Succession." *Erkenntnis* 31(2/3):283–321.

Zhang, Snow. manuscript. "Coherent Combinations of Multiple Experts' Opinions: Another Impossibility Result.".