Updating on the evidence of others

Richard Pettigrew[1] · Jonathan Weisberg[2]

Abstract

One often learns the opinions of others without getting to hear the evidence behind them. How should you revise your own opinions in such cases? Dietrich (2010) shows that, for opinions about objective chance, the method known as upco effectively adds your interlocutor's evidence to your own. We provide a simple way of viewing upco that makes properties like Dietrich's easy to appreciate, and we do three things with it. First, we unify Dietrich's motivation for upco with another motivation due to Easwaran et al. (2016). Second, we show that laypeople can sometimes use upco to resolve expert disagreements. And third, we use it to cricitize the social argument for the uniqueness thesis.

1 Introduction

One often hears 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?

In this paper we study the method known as *upco*, or *multiplicative pooling*. As several authors have noted, when the opinions being combined concern objective chances, upco effectively aggregates the evidence behind those opinions (Dietrich, 2010; Morris, 1983; Winkler, 1968). In other words, using upco to fold someone else's opinions into your own is equivalent to conditionalizing on the evidence behind their opinion, in certain cases.

We provide a simple way of working with upco that makes its evidence-aggregating abilities especially easy to appreciate and work with. Then we apply

✉ Richard Pettigrew
  richard.pettigrew@bristol.ac.uk

  Jonathan Weisberg
  jonathan.weisberg@utoronto.ca

[1] University of Bristol, Bristol, England

[2] University of Toronto, Toronto, Canada

this perspective to three areas of philosophical interest. First, we unify upco's evidence-aggregating powers with another motivation for upco offered by Easwaran et al. (2016). Second, we identify cases where laypeople can use upco to resolve disagreements between experts. And third, we criticize an argument for the uniqueness thesis.

2 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 Fig. 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, namely $5 + 15 = 20$ heads out of 40 flips total.

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, instead multiply $P(H)$ by $Q(H)$, then renormalize (Winkler, 1968).
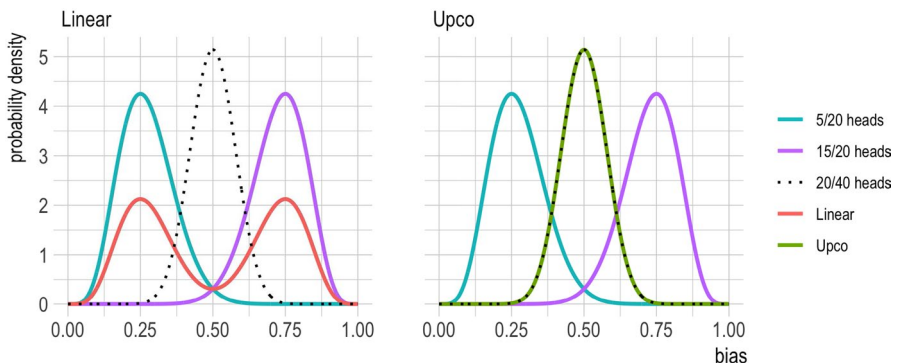

Fig. 1 When pooling over hypotheses about the bias of a coin, linear pooling (red) has undesirable results, while upco (green) aggregates evidence

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[1] 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* (Dietrich, 2010), or *upco* (Easwaran et al., 2016). We'll often write PQ for the distribution over $\{H_i\}$ that it generates.

Notice that, for this operation to be defined, the denominator cannot be zero. So there must be at least one H_i to which both parties assign positive probability. Otherwise, their opinions are too incompatible to be multiplicatively combined. Linear pooling does not have this limitation though, and it has at least one other advantage as well.

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, and on the ways P and Q are distributed over that partition.[2]

And yet, upco turns out to have many desirable properties, a number of which are laid out by Easwaran et al. (2016). Our purpose in this section is to illustrate another desirable feature due to Winkler (1968) and, in more general form, Dietrich (2010). This feature emerges when the H_i are chance hypotheses—about the bias of a coin, for example.

In the right-hand panel of Fig. 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.[3] For example, in Fig. 2, P's posterior is based on only 10 flips, while Q's is based on 20. The dashed

---

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

[2] This advantage is reversed, however, if we work on the odds scale, rather than the probability scale. Then upco amounts to simple multiplication, without any need for mucking around with normalizing constants: if we let $O_P(H) = P(H)/P(\overline{H})$, then $O_{PQ}(H) = O_P(H)\, O_Q(H)$.

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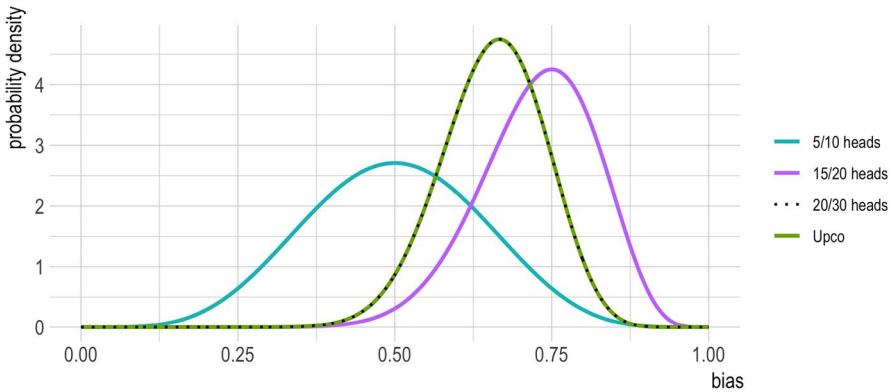
Fig. 2 Upco works even when one agent has more evidence, *e.g.* 20 observations vs. 10

curve is the posterior for their aggregate evidence, and the upco curve in green coincides perfectly.

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 (Dietrich, 2010). 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.

3 A special case

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, respectively. 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 posteriors is denoted $P_E Q_F$. The formal statement of our first result—which is a special case of Dietrich's (2010) Theorem 1—is then as follows (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's 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.

The restriction to chance hypotheses is significant, though. For example, suppose in the coin tossing case that Q were to report their opinion about heads on the next toss, rather than their opinions about the bias. Then Eq. (1) would fail, and upco would no longer serve to aggregate P's evidence with Q's. Instead, a slightly adjusted version of linear pooling would do the job.[4]

What about assumption (i)? What if Q's prior isn't uniform over $\{H_i\}$? We'll generalize Proposition 1 to address 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.

4 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

---

[4] Assuming P and Q have each observed n heads, the relevant function is

$$\frac{P(H) + Q(H)}{2}\left(\frac{n+2}{n+1}\right) - \frac{1}{2(n+1)}.$$

This function converges to linear pooling as $n \to \infty$. But unlike linear pooling, it is not convex. So it escapes the impossibility result due to Zhang (manuscript) discussed in Sect. 6.

associative. And, crucially, this same product operation also captures updating by conditionalization.

Looking at the definition of upco on page 899, 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: $P(QR) = (PQ)R$.[5]

When multiplying numbers, the value 1 has a special role: multiplying by 1 has no effect, $x \cdot 1 = x$. The[6] 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 $\{H_i\}$.[7] 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. 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.[8]

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.[9]

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 $\{H_i\}$:

---

[5] Notice that these properties are trivial if we work with odds rather than probabilities. Then upco just amounts to scalar multiplication (see footnote 2), which is of course commutative and associative. Similar remarks apply to the other properties of upco developed in the rest of this section: all are much more obvious on the odds scale. One way to think about the main result of this section, Proposition 2, is that it simply verifies these properties for the probability scale. This saves us the work of having to constantly translate between odds and probabilities in the rest of the paper, where we discuss ideas and results commonly presented in the probabilistic format.

[6] 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.

[7] 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.

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

[9] So the regular distributions form an abelian group, with upco as the product operation.

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

(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 following equation:[10]

$$P_E = PE_P.$$

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 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 Eq.* (3).

In the next section, we'll use these properties to address the epistemological problem that P faced at the end of Sect. 2.

5 When Q is not uniform

Recall where we left things at the end of Sect. 2. If Q's prior was uniform over $\{H_i\}$ and Eqs. (1) and (2) hold, then P can use upco on Q's posterior to effectively conditionalize on their evidence. 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.

---

[10] 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.

5.1 When *Q* is known

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}.[11] Then the algebraic properties developed in Sect. 3, together with assumptions (1) and (2), deliver:[12]

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

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:

Proposition 3 Let $\{H_i\}$ be a partition such that conditions (1) and (2) hold, and $Q(H_i) > 0$ for all H_i. Then for all H_i, $P_E Q_F Q^{-1}(H_i) = P(H_i \mid EF)$.

Notice that Proposition 1 is a special case of this result: when Q is uniform, so is Q^{-1}, so this term drops out. Indeed, Proposition 3 is in turn a special case of Dietrich's (2010) Theorem 1, alluded to earlier.

Nevertheless, Proposition 3 merits independent statement here, because it concerns an epistemologically distinct case of interest. It also illustrates the value of the algebraic perspective introduced by Proposition 2. Using that perspective, we can grasp epistemologically significant features of upco like Proposition 3 in just a single line of elementary algebra, as above.

One might question the epistemological interest of Proposition 3 on the grounds that P is unlikely to know what Q's prior was. And it is important to acknowledge that, quite often, we only get to hear what someone thinks now, and not what they thought in the past.

Still, cases where Q's prior is known may not be so uncommon. After all, Q might simply tell P what their prior was; it's not unusual to share one's perspective by saying something like, "I used to think X, but over time I've come to think Y instead." Often it's too hard to articulate all the evidence and experience that led to such a shift in opinion, so instead we describe the shift itself, and hope that this

---

[11] This solution requires some extra computation, but not as much as 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.

[12] 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).

conveys the kind of information that led to it. Proposition 3 can then guide P's interpretation of such a shift, when the topic concerns chance hypotheses.

There are other cases, too. For example, if there is a conventional prior commonly used in a certain domain or scientific field, then P might be able to count on Q having proceeded from that prior. Or, P might have good reason to think that Q's prior was the same as their own, if *e.g.* they have similar cultural backgrounds, epistemic tendencies, cognitive traits, etc.

Of course, P might only have an approximate idea of what Q's prior was in these kinds of cases. But—except in formal modeling contexts where probabilities are communicated precisely—the same will typically be true for Q's posterior. In most real-world cases, the opinions of others are never known exactly. So the entire project of using precise probabilistic rules to treat the problem of learning from the opinions of others is only an idealized model. Nevertheless, this idealized model can be useful in applications, where it can be used as an approximation. And it has theoretical interest, as we'll see in later sections.

5.2 When Q is not known

Now let's turn to the trickier case: suppose P does not know what Q's prior was. So P can only apply upco to the posteriors $P(- \mid E)$ and $Q(- \mid F)$. This yields

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

which 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.

Proposition 4 *Let $\{H_i\}$ be a partition such that P and Q satisfy conditions* (1) *and* (2). *Then for all H_i, $P_E Q_F(H_i) = PQ(H_i \mid EF)$.*[13]

Like Proposition 3, this result has Proposition 1 as a special case (this time the reason is that $PQ = P$ when Q is uniform). And also like Proposition 3, Proposition 4 is itself a special case of Dietrich's (2010) Theorem 1.

Informally, Proposition 4 says that, when P doesn't know Q's prior, they must compromise with Q to acquire their evidence via upco. Rather than conditionalizing P's 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

---

[13] 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.
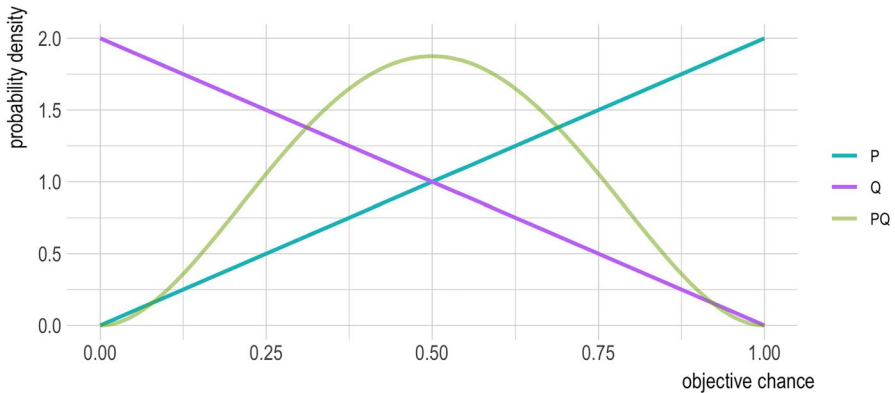
Fig. 3 A possible compromise between two priors

interpretation misses out. In Hall's (2004) terminology, P may partially defer to Q because they have some *analyst* expertise, not merely *database* expertise.

To illustrate, suppose that P and Q are contemplating the objective chance of some novel event. Neither of them has any relevant experience or data, so they must rely on purely *a priori* considerations. After some reflection, P favours a higher chance, Q a lower one. Specifically, their priors are the blue and purple lines in Fig. 3, respectively.[14]

Now suppose that P learns about Q's prior. Based on what they know of Q's epistemic prowess, they take Q's opinions here seriously. Not so seriously that they will simply adopt Q's prior in place of their own. But seriously enough to adjust their own prior in light of Q's. One possible adjustment is the compromise in green in Fig. 3, arrived at by upco. Then, upon learning that Q favours lower chances, P will dampen their expectations about the novel event in question.

But why might P adopt this particular compromise? Why not *e.g.* split the difference instead, making their revised prior the uniform one?

Well, P's prior was based on some *a priori* reason or argument for thinking that the chance of the event in question is high. So when they see Q's prior, they recognize it as the result of some similar *a priori* consideration, but favouring a low chance instead of a high one.

Now, in this particular example, the force of this *a priori* consideration of Q's is exactly equivalent to observing one prior chance event of the sort in question, and finding it negative. More precisely, Q's prior in purple is what you'd get as a posterior if you started with a uniform prior and observed one negative outcome.

So a natural way for P to respond to Q's prior is to combine Q's negative *a priori* insight with their own positive one, by updating as if they'd observed one prior instance of the sort of chance event in question and found it negative. And this is

---

[14] P's prior in blue is a Beta(2, 1), and Q's prior in purple is a Beta(1, 2). The upco compromise in green is thus a Beta(3, 3).

equivalent to using upco to combine their blue prior with Q's purple one, to arrive at the green compromise.

In general, the idea here is to treat *a priori* considerations *as if* they were empirical data. When this analogy is apt, combining the *a priori* insights of others with your own is a matter of combining their "virtual data" with your own. And combining data is what upco does.

So the compromise that Proposition 4 requires between P's prior and Q's will be apt in such cases. In fact, we'll see in the next section that this compromise is what P would adopt upon learning Q's prior, when the proposed analogy holds. More precisely, whenever Q's prior matches the posterior you'd get by conditionalizing a uniform prior on data, then applying upco is equivalent to conditionalizing on the fact that Q holds that prior. Stated in the case of a two-cell partition $\{H, \overline{H}\}$ for simplicity,

$$P(H \mid Q(H) = q) = PQ(H),$$

when Q's prior $Q(H) = q$ can be viewed as if it derived from a uniform ur-prior by conditionalizing on data.

So even though Proposition 4 requires P to first adopt the compromise PQ before updating on the aggregate data EF, this might in fact be precisely what P wants. The compromise PQ might be just the prior they would have adopted if they had known what Q's prior was.

That said, there are certainly cases where this compromise is not one that P would endorse. After all, P might think Q's prior isn't worth taking seriously at all. Or they might take it so seriously that they would abandon their own prior entirely and adopt Q's instead if they knew what it was. And even in cases where they would instead compromise with Q, it needn't be the particular compromise that upco generates. The analogy between *a priori* considerations and empirical data needn't hold. We only claim that, when it does, Proposition 4 shows upco to have desirable results.

6 Updating on the credences of others

We've seen how upco can be used to conditionalize on the evidence *behind* someone's opinions. But Easwaran et al. (2016) use it to conditionalize on the opinions themselves. They show that, in certain cases, if P applies upco to Q's opinions, the result is the same as if P had conditionalized on the fact that Q holds those opinions.

In which cases though? The key is in the likelihoods P assigns to the opinions Q might hold. For simplicity, consider just a two-cell partition $\{H, \overline{H}\}$. Then what matters is $P(Q(H) = q \mid H)$, viewed as a function of q, and also on $P(Q(H) = q \mid \overline{H})$ viewed as a function of q. Specifically, these two functions must have the form

$$P(Q(H) = q \mid H) = c \cdot q \cdot f(q), \tag{4}$$

$$P(Q(H) = q \mid \overline{H}) = c \cdot (1 - q) \cdot f(q), \tag{5}$$

where c is a constant and f is a strictly positive function on [0, 1]. When P's likelihoods have this form, upco agrees with conditionalization on Q's opinion: $PQ(H) = P(H \mid Q(H) = q)$.

Conditions (4) and (5) are quite abstract, though. So it's natural to wonder, are there common or familiar cases where these conditions obtain, and thus where upco agrees with conditionalization?

Easwaran et al. offer examples where P's likelihoods are linear. In these cases, we can dispose of the function f by making it constant at 1, and we let c be the sum of the possible values of $Q(H)$. For example, if the possible values are $0, .1, .2, \ldots, .9, 1.0$, then $c = 1/55$.

This kind of example is simple mathematically. But epistemologically, it's somewhat mysterious. Under what circumstances would these be the possible values $Q(H)$ might take? And why would the probability of $Q(H)$ taking the value q be $q/55$ if H is true, and $(1 - q)/55$ if H is false? What sort of epistemic scenario might Q be facing such that these are the possible outcomes and likelihoods? Without an answer to this question, the utility of conditions (4) and (5) is in question.

As it turns out though, there are natural cases. Take the sorts of cases we've been considering, where P knows Q's prior over a partition of chance hypotheses, and then learns their posterior Q_F. Typically, P won't be able to infer what the proposition F is that Q conditionalized on. They can, however, infer its normalized likelihood distribution, by using the inverse of Q's prior: $Q^{-1}Q_F = Q^{-1}QF_Q = F_Q$. And by the Principal Principle, P's likelihood distributions are the same as Q's, so $F_Q = F_P$. Thus, learning Q's posterior is, for P, equivalent to learning that some proposition with likelihood distribution F_P is true.

Now, if the propositions with that likelihood distribution are mutually exclusive, then learning that one of them holds is equivalent to learning any one of them. So learning their disjunction is, for P, equivalent to learning F. Thus P's posterior is PF_P, which agrees with the upco calculation PQ_FQ^{-1} by Proposition 3. Conditionalization and upco thus agree: conditionalizing on Q's posterior yields the same distribution over $\{H_i\}$ as the upco calculation PQ_FQ^{-1}.

Notice that this argument requires assumptions with a more epistemic flavour than the ones we've relied on previously. For example, we assume here that P knows Q obeys the Principal Principle and will update by conditionalization. Whereas Proposition 3 only assumed that Q does these things—P needn't know that they do. The difference arises because P is now conditionalizing on Q's posteriors, which brings their expectations about how those posteriors are formed into play. Notice that Easwaran et al.'s (4) and (5) are, similarly, conditions on P's beliefs about Q.

The following, formal statement encodes these epistemic assumptions implicitly. For example, the random variable $\mathcal{Q}$ representing Q's posterior is arrived at by applying Bayes' theorem but with P's likelihoods. So P effectively takes for granted that Q obeys conditionalization and the Principal Principle.

Proposition 5 *Let $\{F_j\}$ be a countable partition, and let $\mathcal{Q}$ be a random variable whose value when $F \in \{F_j\}$ obtains is*

$$q_F = \frac{Q(H)P(F \mid H)}{Q(H)P(F \mid H) + Q(\overline{H})P(F \mid \overline{H})}. \tag{6}$$

If $Q_F(H) = q_F$, then $P(H \mid \mathcal{Q} = q_F) = PQ_FQ^{-1}(H)$.

This result generalizes to partitions $\{H_i\}$ with more than two cells, as we show in the Appendix. We state just the two-cell case here for simplicity, and for continuity with conditions (4) and (5). As we remark in the Appendix, (4) and (5) hold in the special case where Q's prior is uniform. In which case the Q^{-1} term drops out of the conclusion of Proposition 5.

Proposition 5 thus accomplishes two things. First, it bolsters Easwaran et al.'s motivation for upco, by providing a natural class of cases where their conditions hold and upco thus agrees with conditionalization. Second, it unifies two apparently distinct ways of motivating upco, namely Easwaran et al.'s and Dietrich's. Roughly speaking, upco conditionalizes on (the evidence behind) Q's opinion when Q's prior over a chance partition is known to be (or just is) obtained by conditionalizing on independent data.

That said, it should be acknowledged that conditions (4) and (5) are more general in a way. They aren't only satisfied in the special case of a uniform prior over a chance partition, and it would be interesting to identify other natural cases where they apply. But we leave that question for future work, and return instead to our main theme.

7 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 $\mathcal{Q}$ and $\mathcal{R}$ are random variables representing Q's and R's opinions about H:

$$P(H \mid \mathcal{Q} = q) = q, \tag{7}$$

$$P(H \mid \mathcal{R} = r) = r. \tag{8}$$

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

$$P(H \mid \mathcal{Q} = q, \mathcal{R} = r) = (q + r)/2. \tag{9}$$

But Dawid et al. ([1995](#)) show that these three conditions together imply $P(\mathcal{Q} = \mathcal{R}) = 1$.[15] In fact, they show that any weighted averaging rule with positive weights implies $P(\mathcal{Q} = \mathcal{R}) = 1$. 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 Fig. [1](#) always lies strictly in between the blue and purple curves, because linear pooling is strictly convex.

Formally, Zhang's generalization replaces ([9](#)) with the more general

$$P(H \mid \mathcal{Q} = q, \mathcal{R} = r) = f(q, r), \tag{10}$$

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 Eqs. ([7](#)), ([8](#)) and ([10](#)) again imply $P(\mathcal{Q} = \mathcal{R}) = 1$, 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.

This rules out several popular alternatives to linear pooling in the kind of case we have been discussing. When the experts opine only about the proposition H, the partition over which we are pooling is $\{H, \overline{H}\}$. And, in the special case of a two-cell partition, alternatives like geometric and harmonic pooling are both strictly convex. Importantly though, the same is not true for larger partitions. And indeed, geometric pooling escapes Zhang's impossibility result when the partition in question is the "ultimate" partition, *i.e.* the partition into singletons of worlds (Baccelli & Stewart, [2023](#)).

This leaves us with the question whether any simple pooling rule is capable of guiding a layperson faced with differing experts on coarser partitions, including even two-cell partitions. As we are about to show, the answer is yes: upco can.[16] In fact, it does so in a significant range of cases, which we can identify using Propositions [1](#) and [2](#).

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 Eq. ([7](#)) will hold. For P, learning Q's opinion is equivalent to learning how many heads and tails they observed. And

---

[15] See also Bradley ([2018](#)), and Gallow ([2018](#)) from whom we borrow this section's title.

[16] As Easwaran et al. ([2016](#)) note, upco is non-convex even on two-element partitions.

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, Eq. (8) 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

$$P(H \mid \mathcal{Q} = q, \mathcal{R} = r) = QR(H). \tag{11}$$

Now, crucially, it's entirely possible that Q and R will get different numbers of heads, and thus report different opinions. So $P(\mathcal{Q} = \mathcal{R}) \neq 1$ in this example. Thus upco is capable of serving two epistemic masters: Eqs. (7), (8) and (11) do not imply $P(\mathcal{Q} = \mathcal{R}) = 1$.

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 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 Eq. (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 (7) and (8). 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,

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

and similarly given $\overline{H}$. Then, if P is uniform over $\{H, \overline{H}\}$, P will resolve any differences according to upco, *i.e.* (11) holds.

We can drop the uniform prior assumption much as we did in Sect. 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:

$$PE_P PF_P = P^2 E_P F_P = P^2 (EF)_P.$$

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:

Proposition 6 *Let* $\{E_i\}$ *and* $\{F_j\}$ *be finite partitions. Let* $\mathcal{Q}$ *be a random variable that takes the value* $P(H \mid E_i)$ *in the event that* E_i*, and let* $\mathcal{R} = P(H \mid F_j)$ *in the event* F_j*. Then* (7) *and* (8) *hold. If, furthermore, each pair* E_i, F_j *is conditionally indepenent given the elements of* $\{H, \overline{H}\}$*, then*

$$P(H \mid \mathcal{Q} = q, \mathcal{R} = r) = P^{-1}QR(H). \tag{12}$$

In the special case $P(H) = P(\overline{H})$*, then* (12) *reduces to* (11)*.*

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

It might seem like a severe limitation of Proposition 6 that it only applies when the two experts begin with the same prior as P. But this isn't a limitation of upco; rather, it's what makes Q and R experts of the kind we're interested in here. If Q and R's priors were different from P's, they would not be experts in P's eyes: Eqs. (7) and (8) would no longer hold. Even though Q and R would still have strictly more evidence than P, making them database experts, P would not agree with their analyses of that evidence. So they would not be analyst experts for P, and P would not trust their judgment in the manner of Eqs. (7) and (8). Zhang's impossibility result would no longer apply.

What's more, P can still use upco to make use of Q and R's database expertise, even in the case where their priors are different from P's, provided P knows what those priors are. Instead of including the inverse P^{-1} of their own prior in Eq. (12), they can include the inverses of Q's and R's priors. In other words, they can make use of the idea behind Proposition 3. And even if they don't know Q's and R's priors, they might still be able make use of their database expertise in the manner of Proposition 4. Although, as we noted at the end of Sect. 4, this depends on P giving a specific sort of partial deference to Q's and R's priors.

8 The social argument for uniqueness

In this section we bring Proposition 2 to bear on an argument for the uniqueness thesis, the claim that there is only one correct way to interpret a body of evidence (Feldman, 2006). On this view, two agents with the same total evidence are never permitted to disagree. The alternative view, known as permissivism, holds that agents with the same evidence can reach different conclusions, at least in some cases.

Here we are concerned with a particular argument for the uniqueness thesis, due to Dogramaci and Horowitz (2016). The argument begins with the observation that we have a social practice of pressuring one another to be rational, a practice that presumably has some value. But why, they ask, is it valuable? What is the good in promoting rationality in others? The best explanation, they argue, is one that presupposes the uniqueness thesis.

In their view, promoting rationality is valuable because it aids in a division of epistemic labour. If there is a unique, correct way of interpreting evidence, and everyone follows it, then we can get the benefits of one another's evidence-gathering simply by hearing the conclusions drawn from that evidence. When someone tells you H is true,

you needn't worry about whether you would have drawn the same conclusion from whatever evidence led them to conclude H. You can just go ahead and believe H, since that's the right conclusion to draw from whatever their evidence was. So promoting rationality makes it possible to share the work of gathering and evaluating evidence.[17]

One problem with this story is that it neglects potential interactions between their evidence and yours. For example, suppose some recent polling has made you 70% confident that a majority of voters favour Party X in the upcoming election. Then you encounter someone you know to be rational who is 80% confident. Should you adopt their view as your own? Maybe, if you happen to know that their evidence includes your own. If their 80% is based on the same polling data you saw, plus some additional data, then you should join them at 80%.

But if their 80% is based on an entirely separate body of polling data, then you should become even *more* than 80% confident. Between the two of you, you have an even larger body of data supporting a Party X majority. So you shouldn't adopt your interlocutor's 80%, but rather something higher.

In general, you can't just adopt the views of other rational agents on the grounds that they're rational. It matters what their evidence for their view is, and how that evidence relates to your own.

Still, Dogramaci and Horowitz's story does seem to work in cases where you happen to know that your interlocutor's data doesn't overlap with your own. Suppose Q begins with a uniform prior over a partition $\{H_i\}$ of chance hypotheses, and goes off to gather data. They make some novel observation F, and then report back with their posterior, Q_F. Now, P won't typically be able to determine the proposition F from this posterior. But they know its likelihood distribution—it's the same as the reported posterior, since $Q_F = UF_Q = F_Q$. And since the H_i are chance hypotheses, P's likelihoods for F are the same: $F_Q = F_P$. So P can get the benefit of Q's evidence-gathering labour, by combining their prior P with the likelihood distribution F_Q gleaned from Q's report. The result $PF_Q = PF_P = P_F$ is the same as P's prior conditionalized on Q's evidence.

So the argument seems to work, in the case of chance partitions and non-overlapping data. What's more, it doesn't just deliver the conclusion that a unique, rational prior exists. It says what that prior is: the uniform-over-chances prior. And this prior just so happens to have a long history in the objective Bayesian tradition, the version of Bayesianism that embraces the uniqueness thesis.[18]

On closer inspection though the argument fails, because the uniform-over-chances prior isn't actually what's driving P's successful exploitation of Q here. Even if Q uses some other prior, P can still solve for F_Q, as long as they know what that prior was. When P learns the posterior $Q_F = QF_Q$, they can solve for F_Q using

---

Q's inverse: $Q^{-1}(QF_Q) = F_Q$. And from there they can get the benefit of Q's evidence, just as before.

The moral is that it's not important what prior Q uses. What matters is that P *knows* what prior they used. That's enough for them to "factor out" Q's prior from their posterior, and isolate the import of their evidence, whatever it was. The division of epistemic labour Dogramaci and Horowitz propose doesn't rely on there being a unique rational prior that everyone uses, but on everyone *knowing* what priors others are using.

Of course, one way to ensure everyone knows each other's priors is to have a social convention requiring everyone to use the same prior. But then this shared prior would be just that: a social convention. The choice of prior would be like deciding which side of the road to drive on; one choice is as good as another, so long as everyone chooses the same.[19]

9 Conclusion

We've been studying upco's ability to aggregate evidence in certain cases, especially cases of opinions about objective chance. We developed a simple, algebraic way of viewing upco that makes its evidence-aggregating powers especially easy to appreciate and work with. And, using that algebraic frame, we've seen that upco's ability to aggregate evidence is closely related to its ability to mimic conditionalization, noted by Easwaran et al. (2016). We've also used that frame to identify cases where laypeople can use upco to resolve disagreements between experts. And we've used it to criticize an argument for the uniqueness thesis.

We conclude that viewing upco as a way of aggregating evidence is a fruitful perspective to take. It both improves our understanding of upco itself, and exposes applications to areas of philosophical interest.

Appendix

Here we give formal statements and proofs of the results in the main text. 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.

---

[19] One might respond that the uniform-over-chances prior would make an especially good convention on objective grounds, since it saves P having to use Q^{-1} to cancel Q in solving for F_Q. This not only saves on computation, but also improves accuracy by eliminating opportunities for calculation errors. In reply, we would note how radically our sights are being lowered here. We already lowered them once, when we restricted the benefits of a privileged prior to cases of chance hypotheses and non-overlapping data. Now they are being lowered still further; the value being claimed for the uniquely rational prior is just that it saves on computation and reduces opportunities for calculation errors. If this is all the social argument for uniqueness comes to, we are content to leave it there.

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*

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

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. $\qquad\square$

Remark 1 As noted in footnotes 2 and 5, the above result becomes fairly trivial if we work with odds rather than probabilities. Corresponding to each regular probability vector $\mathbf{p}$ is an odds vector whose i^{th} entry is $p_i / \sum_{j \neq i} p_j$. It's straightforward to verify that the odds vector corresponding to $\mathbf{pq}$ is the elementwise product of those corresponding to $\mathbf{p}$ and to $\mathbf{q}$. The commutativity and associativity of upco is then even easier to verify, as are the other properties stated in the above proposition. Nevertheless, we choose to work with probabilities here, for the sake of continuity with existing literature.

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

Proposition 7 *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 $\mathbf{g}_P$. $\qquad\square$

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

Proposition 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

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

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. $\square$

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

Proposition 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

$$(\mathbf{pe}_P)(\mathbf{qf}_Q) = (\mathbf{pq})(\mathbf{e}_P \mathbf{f}_Q) = (\mathbf{pq})(\mathbf{e}_P \mathbf{f}_P) = (\mathbf{pq})(\mathbf{ef})_P.$$

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. $\square$

Next we prove the general form of Proposition 5.

Proposition 8 *Let $\{H_i\}$ and $\{F_j\}$ be countable partitions. Let $\mathbf{Q}$ be a random vector whose value when $F \in \{F_j\}$ obtains is the vector $\mathbf{q}^F$ defined by*

$$q_i^F = \frac{Q(H_i)P(F \mid H_i)}{\sum_k Q(H_k)P(F \mid H_k)}. \tag{13}$$

If $Q_F(H_i) = q_i^F$ for all i, then $P(H_i \mid \mathbf{Q} = \mathbf{q}^F) = PQ_F Q^{-1}(H_i)$ for all i.

Proof Fix F, and let J_F index the elements of $\{F_j\}$ that satisfy (13). So $\mathbf{Q} = \mathbf{q}^F$ just in case $\bigcup_{j \in J_F} F_j$, and thus

$$P(\mathbf{Q} = \mathbf{q}^F \mid H_i) = P(\bigcup_{j \in J_F} F_j \mid H_i) = \sum_{j \in J_F} P(F_j \mid H_i).$$

By Bayes' theorem then:

$$P(H_i \mid \mathbf{Q} = \mathbf{q}^F) \propto P(H_i)P(\mathbf{Q} = \mathbf{q}^F \mid H_i)$$

$$= P(H_i) \sum_{j \in J_F} P(F_j \mid H_i)$$

$$= P(H_i) \sum_{j \in J_F} \frac{q_i^F}{Q(H_i)} \sum_k Q(H_k)P(F_j \mid H_k)$$

$$\propto P(H_i)Q_F(H_i)/Q(H_i).$$

Similarly, $PQ_FQ^{-1}(H_i) \propto P(H_i)Q_F(H_i)/Q(H_i)$. So the two distributions have the same proportions, hence must be identical. $\qquad\square$

Remark 2 To recover Easwaran et al.'s conditions (4) and (5) in the above proof, consider the special case where $\{H_i\} = \{H, \overline{H}\}$ and $Q(H) = Q(\overline{H}) = 1/2$. There we set $c = 1$, and define

$$f(q_1^F) = 2 \sum_{j \in J_F} \sum_k Q(H_k)P(F_j \mid H_k).$$

We now turn to Proposition 6, which we'll prove in the more general form of Proposition 9. 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 1 *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, $\mathbf{q} = \mathbf{pe}_P$, $\mathbf{r} = \mathbf{pf}_P$, $\mathbf{g}_P = \mathbf{e}_P\mathbf{f}_P$, and $\mathbf{p}^{-1}$ exists. So,

$$\mathbf{pg}_P = \mathbf{pe}_P\mathbf{f}_P = (\mathbf{p}^{-1}\mathbf{p})\mathbf{pe}_P\mathbf{f}_P = \mathbf{p}^{-1}(\mathbf{pe}_P)(\mathbf{pf}_P) = \mathbf{p}^{-1}\mathbf{qr}.$$

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. $\qquad\square$

In the main text we stated Proposition 6 in terms of a two-cell partition $\{H, \overline{H}\}$. Proposition 9 addresses the more general case, where $\{H_i\}$ may have more than two elements.

Proposition 9 *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*

$$P(H_i \mid \mathbf{Q} = \mathbf{q}) = q_i, \tag{14}$$

$$P(H_i \mid \mathbf{R} = \mathbf{r}) = r_i. \tag{15}$$

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

$$P(H_i \mid \mathbf{Q} = \mathbf{q}, \mathbf{R} = \mathbf{r}) = P^{-1}QR(H_i). \tag{16}$$

Proof Note that, with Lemma 1 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 (14) and (15).

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 1 implies that for all H_i,

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

as desired. □

References

Babic, B., Gaba, A., Tsetlin, I., & Winkler, R. L. Resolute and Correlated Bayesians. INSEAD Working Paper No. 2022/20/DSC. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4077726#.

Baccelli, Jean, & Stewart, Rush T. (2023). Support for geometric pooling. *The Review of Symbolic Logic, 16*(1), 298–337.

Bradley, Richard. (2018). Learning from others: Conditioning versus averaging. *Theory and Decision, 85*(1), 5–20.

Dawid, Alexander P., DeGroot, Morris H., & Mortera, Julia. (1995). Coherent combination of experts opinions. *Test, 4*(2), 263–313.

Dietrich, Franz. (2010). Bayesian group belief. *Social Choice and Welfare, 35*(4), 595–626.

Dogramaci, Sinan, & Horowitz, Sophie. (2016). An argument for uniqueness about evidential support. *Philosophical Issues, 26*(1), 130–47.

Easwaran, Kenny, Fenton-Glynn, Luke, Hitchcock, Christopher, & Velasco, Joel D. (2016). Updating on the credences of others: Disagreement, agreement, and synergy philosophers. *Imprint, 6*(11), 1–39.

Feldman, Richard. (2006). *Reasonable religious disagreementsmeditations on atheism and the secular life* (pp. 194–214). In Philosophers Without Gods: Oxford University Press.

Gallow, J. Dmitri. (2018). No one can serve two epistemic masters. *Philosophical Studies, 175*(10), 2389–2398.

Hall, Ned. (2004). Two mistakes about credence and chance. *Australasian Journal of Philosophy, 82*(1), 93–111.

Laplace, P. S. (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.

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, S. Coherent Combinations of Multiple Experts' Opinions: Another Impossibility Result. Draft manuscript / personal communication.