Geometric Pooling: A User's Guide

Richard Pettigrew
University of Bristol

Jonathan Weisberg
University of Toronto

Abstract

Much of our information comes to us indirectly, in the form of conclusions others have drawn from evidence they gathered. When we hear these conclusions, how can we modify our own opinions so as to gain the benefit of their evidence? In this paper we study the method known as geometric pooling. We consider two arguments in its favour, raising several objections to one, and proposing an amendment to the other.

Some of your evidence about the world you gather directly for yourself, but much is gathered by others. Sometimes you obtain the second sort of evidence directly, when one of your fellows describes it explicitly. But often you learn only the effect it has had on their opinions. For instance, you might learn your doctor's view about what is causing your symptoms without learning all the background knowledge and detailed test data that underpins it. Or you might learn your fellow researcher's new probabilities for the hypotheses you're both investigating, rather than the data she's just collected, and on which she's just updated her probabilities.

When you encounter someone who has gathered their own evidence, and you learn not the evidence itself but only the opinions they now have, how should you update your own opinions? The Bayesian says you should treat such second-order evidence just like you treat any other evidence and update using Bayes' rule; and they have many arguments in favour of this prescription.[1] But sometimes you can't do that. After all, in this situation, Bayes' rule requires you to have prior probabilities in different hypotheses concerning the opinions of the other person, and likelihoods given those hypotheses; and you might simply not have set these. So we seek an alternative method.

In this paper, we consider a particular proposal: you should combine your probabilities with your fellow's using a method known as *geometric pooling*. We will consider two arguments in its favour. To the first, we will raise several objections; to the second, we will propose an amendment.

[1] See for example Brown (1976), Oddie (1997), Lewis (1999), Greaves and Wallace (2006), Briggs and Pettigrew (2018), and Gallow (2019).

1 The framework

Throughout, we'll assume that the opinions of you and your fellow concern different ways the world might be, and different hypotheses about the objective chance of its turning out each of these ways. This is an extremely common situation, both in the context of scientific research and in our everyday lives. For instance, the hypotheses might concern the value of a particular parameter in a scientific theory, such as the basic reproduction number (R value) for an infectious disease, while the states of the world might be distinguished by some observable feature, such as the pattern of illness in a given population. Each hypothesis about the R value tells you the chance of observing a given pattern of illness.

For ease of exposition, we will assume the hypotheses concern the chance that a particular coin will land heads, which we'll call its bias. And the states of the world are distinguished by different sequences of heads and tails that might result when the coin is tossed repeatedly. We'll assume the hypotheses form a partition, call it $\mathcal{H}$; and the states of the world form another, call it $\mathcal{S}$. You and your fellow each assign probabilities over $\mathcal{H}$ and over $\mathcal{S}$, and the Principal Principle (Lewis, 1980) says how they should relate: your probability for a sequence conditional on a hypothesis about the chances should be whatever probability the chance function that hypothesis posits assigns to that sequence.[2]

Now suppose you observe some tosses of the coin, your fellow observes some different tosses, and you each update your opinions in the light of this new evidence. You meet and your fellow tells you their new probabilities. How should you incorporate that information? It turns out it depends on which of their probabilities they share. Here are three sets of probabilities you might learn from your fellow: (1) their probability that the next toss will land heads; (2) their probabilities in each of the possible sequences of tosses; (3) their probabilities in each of the chance hypotheses. We'll take these in turn.

But first, let's survey two ways you might incorporate the information about your fellow's probabilities. The *linear* or *arithmetic pool* of probability functions P and Q splits the difference between the credences that each assigns. So the pooled probability in A is

$$L(P,Q)(A) = \frac{P(A) + Q(A)}{2}.$$

If P and Q are both probability functions, so is their arithmetic pool.

Where the arithmetic pool of P and Q appeals to the arithmetic mean of the credences they assign, their geometric pool appeals to the geomet-

[2]If C_i is the chance function posited by H_i, then the Principal Principle requires of your prior P that $P(S \mid H_i) = C_i(S)$. And so, if you obtain evidence E, it requires of your posterior $P(- \mid E)$ that $P(S \mid H_iE) = C_i(S \mid E)$ (Hall, 2004, 101-2).

ric mean. The arithmetic mean of p and q is $(p+q)/2$; their geometric mean is $\sqrt{pq}$. However, we cannot take the pooled probability in A to be $\sqrt{P(A)Q(A)}$ because doing so does not always deliver a probability function. Instead, we must use the normalized geometric mean rather than the geometric mean itself. But that means we can define the geometric pool of P and Q only over a specified partition. If $\mathcal{A} = \{A_1, \ldots, A_n\}$ is a finite partition, then the pooled probability in A_i is

$$G_{\mathcal{A}}(P,Q)(A_i) = \frac{\sqrt{P(A_i)Q(A_i)}}{\sum_{j=1}^{n} \sqrt{P(A_j)Q(A_j)}}$$

Note that this is defined only if there is A_i in $\mathcal{A}$ such that $P(A_i), Q(A_i) > 0$. In this case, we say that P and Q have an overlapping support in $\mathcal{A}$. (We define geometric pooling over infinite partitions in the Appendix.)[3]

2 Pooling probabilities of the next toss

Suppose you meet your fellow and they tell you their probability in H, the prediction that the next toss will land heads. How should you update your probability in H?

Suppose P is your prior probability function and E is your evidence, while Q is your fellow's prior and F is their evidence. In many cases, the arithmetic pool of $P(H \mid E)$ and $Q(H \mid F)$ will approximate updating your prior on the aggregate evidence using Bayes' rule. After all, if E and F give the frequencies of heads in two large, disjoint samples of equal size, then $P(H \mid E)$ will closely match the frequency in your sample, and $Q(H \mid F)$ will closely match the frequency in their your fellow's. And in that case, the arithmetic mean $(P(H \mid E) + Q(H \mid F))/2$ will closely match the overall frequency in the aggregated sample you've amassed between you, and $P(H \mid EF)$ will closely match that too.

Why only "closely" match, why not exactly? Because we have to account for the influence of priors. Your opinion $P(H \mid E)$ doesn't just reflect the frequency of heads in your sample, it also reflects your beliefs about the coin from before you observed that sample, which are encoded in P. Likewise for your fellow's opinion, $Q(H \mid F)$. But if the sample is large, and your priors treated the flips as independent and identically distributed, then the observed frequency must be very close to the resulting opinion. In which case the arithmetic mean of opinions will closely match the opinion you would have if you were to conditionalize your prior on the full, aggregated sample.

[3]For some formal background on linear and geometric pooling, see Genest and Zidek (1986). For some recent philosophical discussion, see Dietrich (2010), Dietrich and List (2016), Russell, Hawthorne and Buchak (2015), and Baccelli and Stewart (2023).

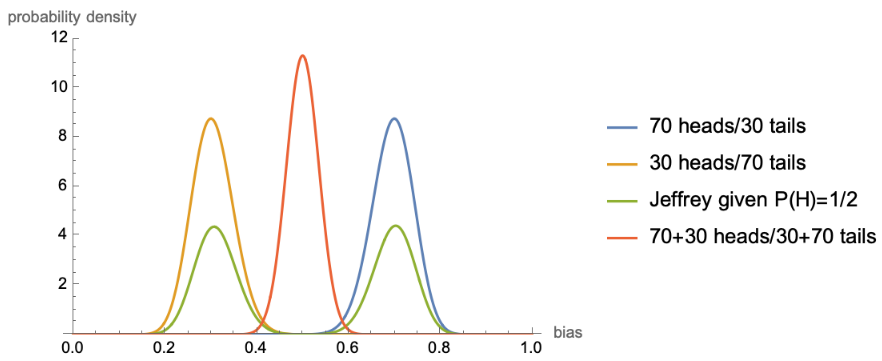

Figure 1: Suppose you and your fellow share the uniform prior over the possible biases. You witness 70 heads and 30 tails, while they witness 30 heads and 70 tails from a different set of tosses. Then the blue line gives your posterior distribution over the biases after updating on your evidence, and the yellow line gives your fellow's after updating on theirs. The red line gives the distribution obtained by conditionalizing your shared prior on the aggregate evidence, while the green line gives the distribution obtained by first taking the arithmetic pool of the posterior probabilities in H and then applying Jeffrey conditionalization to obtain the posteriors about the biases.

However, having updated your probabilities in H and $\overline{H}$ by linear pooling with your fellow, you must now find a way to update your probabilities in all the other propositions about which you have an opinion. Since H and $\overline{H}$ form a partition, a natural solution is to use Jeffrey conditionalization (Jeffrey, 1965). If you do, then your new probability in a chance hypothesis H_i from $\mathcal{H}$ is

$$L(P,Q)(H)P(H_i \mid HE) + L(P,Q)(\overline{H})P(H_i \mid \overline{H}E).$$

But, as we see in Figure 1, this does not come close to approximating the posterior obtained by conditionalizing on the aggregate evidence. So arithmetic pooling has some use in this situation, but its scope is rather limited.

3 Pooling probabilities of sequences

Now suppose you meet your fellow and they tell you not their probability that the next toss will land heads, but instead their probabilities for each of the possible sequences. How should you then update your probabilities in those sequences?

Baccelli and Stewart (2023) argue that geometric pooling is a good strategy in this case. Let's suppose you and your fellow shared the same prior probability function P before you started collecting evidence. The crucial point is that the evidence you have collected since then takes the form of a disjunction of states of the world from $\mathcal{S}$; and similarly for your fellow. If you witnessed heads, tails, then heads, then your evidence is the disjunction of every state of the world in which the sequence of tosses begins in that way.

We write E for your evidence and F for your fellow's. Then Baccelli and Stewart note the following striking fact:

Proposition 1. *If $P(EF) > 0$, then, for any state S in $\mathcal{S}$,*

$$G_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S) = P(S \mid EF).$$

That is, pooling your posterior probabilities with your fellow's gives the same probabilities over the sequences as updating your shared prior with the aggregate evidence.

What's more, if we use Jeffrey conditionalization to extend the geometric pool over $\mathcal{S}$ so that it assigns probabilities also to the chance hypotheses in $\mathcal{H}$, Proposition 1 extends as well. In order to apply Jeffrey conditionalization, we need likelihoods for each chance hypothesis H_i given each sequence S. But of course these will be different for P, $P(- \mid E)$ and $P(- \mid F)$. However, as we will see, it doesn't matter which of these we use. Let's see the calculation if we use P to give the likelihoods. If $G_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S)$ gives your new probability in S, then Jeffrey conditionalization says that your new probability in H_i is

$$\begin{aligned}
& \sum_{S \in \mathcal{S}} G_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S) P(H_i \mid S) \\
= & \sum_{S \in \mathcal{S}} P(S \mid EF) P(H_i \mid S) \ \text{(by Proposition 1)} \\
= & \sum_{S \in EF} \frac{P(S)}{P(EF)} \frac{P(H_i S)}{P(S)} \\
= & \sum_{S \in EF} \frac{P(H_i S)}{P(EF)} = P(H_i \mid EF)
\end{aligned}$$

It's straightforward to see that taking the likelihoods from $P(- \mid E)$ or $P(- \mid F)$ instead will give the same result.

Note that arithmetic pooling very much does not boast the feature described in Proposition 1. Suppose the coin will be tossed just two times, so that there are four possible sequences, HH, HT, TH, and HH. And suppose you and your fellow both assign prior probabilities as follows:

	HH	HT	TH	TT
$P(-)$	1/3	1/6	1/6	1/3

Suppose you observe that the first toss lands heads and your fellow observes that the second toss lands tails. Then your shared prior updated on the aggregate evidence places all probability on HT:

	HH	HT	TH	TT
$P(- \mid EF)$	0	1	0	0

But the arithmetic mean of your posteriors does not:

	HH	HT	TH	TT
$P(- \mid E)$	$2/3$	$1/3$	0	0
$P(- \mid F)$	0	$1/3$	0	$2/3$
$L(P(- \mid E), Q(- \mid F))$	$1/3$	$1/3$	0	$1/3$

However, while the feature described in Proposition 1 distinguishes geometric pooling from arithmetic pooling, the former is not unique in having it, and so it's not clear that the result provides quite the support for geometric pooling that Baccelli and Stewart suggest. Consider harmonic pooling, for instance: if $\mathcal{A} = \{A_1, \ldots, A_n\}$ is a partition, and P and Q have overlapping support among $\mathcal{A}$, the harmonic pool of P and Q is defined as

$$H_{\mathcal{A}}(P, Q)(A_i) = \frac{\frac{P(A_i)Q(A_i)}{P(A_i)+Q(A_i)}}{\sum_{j=1}^{n} \frac{P(A_j)Q(A_j)}{P(A_j)+Q(A_j)}}.$$

Then we have the following result, analogous to Proposition 1:

Proposition 2. *If $P(EF) > 0$, then, for any state S in $\mathcal{S}$,*

$$H_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S) = P(S \mid EF).$$

Now, as Baccelli and Stewart note, any pooling operator for which an analogue of Propositions 1 and 2 holds must agree with geometric pooling in the cases in question. That is, it must pool $P(- \mid E)$ and $P(- \mid F)$ in the same way geometric pooling does. But that's only because it must pool them to give $P(- \mid EF)$. Harmonic pooling does that; but it doesn't agree with geometric pooling in all cases—not by a long way.

Another reason that Proposition 1 gives only limited support to geometric pooling is that it doesn't do what you might want it to do in cases in which you and your fellow have different priors. Here is the generalization of Proposition 1 that covers such cases, where P is your prior and Q is your fellow's:

Proposition 3. *If $P(- \mid E)$ and $Q(- \mid F)$ have overlapping support among $\mathcal{S}$, then, for any state S in $\mathcal{S}$,*

$$G_{\mathcal{S}}(P(- \mid E), Q(- \mid F))(S) = G_{\mathcal{S}}(P, Q)(S \mid EF)$$

That is, if you take your prior updated on your evidence and your fellow's prior updated on theirs, and then combine them using geometric pooling, you get the same result as if you had combined your prior and your fellow's using geometric pooling and then updated on the aggregate evidence. But you might only be interested in your fellow's opinion to extract the evidence that informs it, preferring to retain your own prior rather than pool with theirs.

In fact, there are pooling operations that deliver this:

$$\Delta_{\mathcal{A}}(P,Q)(A_i) = \begin{cases} 0 & \text{if } Q(A_i) = 0, \\ \frac{P(A_i)}{\sum_{A_j : Q(A_j) > 0} P(A_j)} & \text{if } Q(A_i) > 0. \end{cases}$$

This rule ignores everything about your fellow's opinions, except whether they are zero or nonzero. Really, it just uses the zeros to identify their evidence, then conditionalizes your posterior on that evidence. Small wonder, then, that we have the following analogue of Proposition 1: if $P(EF) > 0$, then

$$\Delta_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S) = P(S \mid EF).$$

And indeed this rule will also usually favour your way of interpreting the evidence and ignore your fellow's, which geometric pooling does not. That is, if Q is regular over $\mathcal{S}$, so that $Q(S) > 0$ for all sequences S in $\mathcal{S}$, and if $P(EF) > 0$, then

$$\Delta_{\mathcal{S}}(P(- \mid E), Q(- \mid F))(S) = P(S \mid EF).$$

This observation points up a third problem with an argument for geometric pooling based on Proposition 1. In order to use geometric pooling this way, you have to know *all* your fellow's opinions about the sequences. They need to tell you $Q(S \mid F)$ for every sequence S in $\mathcal{S}$. But if they've communicated all that, and if their prior Q is regular, they've already told you what their evidence F is. So you might as well just conditionalize on F. It's strictly easier to ask what their evidence is and conditionalize on that than it is to get their full set of opinions over the possible sequences and take the geometric pool of those with yours.

4 Pooling probabilities of chance hypotheses

Finally, suppose you meet your fellow and they tell you—not their probability that the next toss will land heads, nor their probability for each possible sequence—but their probability for each chance hypothesis. How should you update your probabilities in those hypotheses then?

We will begin by describing how geometric pooling behaves in this case, extending some observations from the literature. Then, in Section 6, we will use what we've discovered to scrutinise an argument that appeals to geometric pooling's performance in computer simulations.

As before, suppose you and your fellow begin with a common prior. Then you each observe your own series of coin flips, and you each conditionalize on what you've observed. Then your fellow shares their posterior opinions about the coin's bias. In Section 3, we saw that if you geometrically pool posterior credences in the different possible sequences, you can

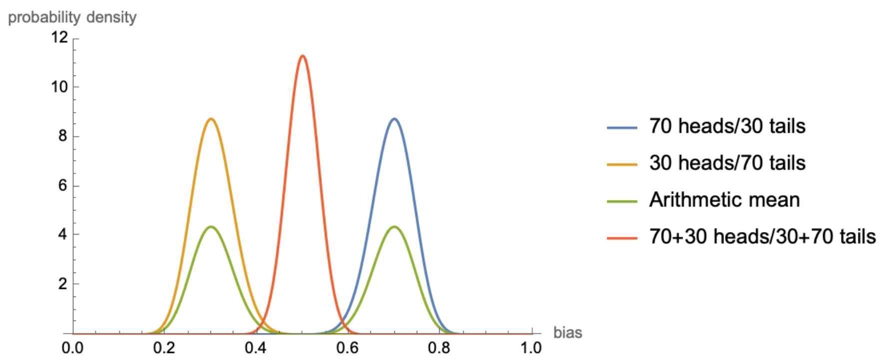

Figure 2: The red line gives the distribution obtained by conditionalizing your shared prior on the aggregate evidence, while the green line gives the arithmetic pool of your posterior and your fellow's.

incorporate your fellow's evidence perfectly; but we noted that it's in fact easier for them just to tell you their evidence than to tell you their credences in each of the sequences. But it's not so difficult for them to share their credences in the different chance hypotheses, particularly if these are given by a standard probability distribution determined by a few parameters, such as a Beta distribution.[4] So suppose they do share that. What happens when you pool your probabilities with theirs?

If you use arithmetic pooling in this case, it doesn't perform well.[5] To illustrate, suppose you and your fellow both start with a uniform prior over the possible biases. You then observe 70 heads out of 100 flips, while they observe 30 heads out of a different 100 flips. If you then take the arithmetic pool of your two distributions over the possible biases, the result will be the bimodal distribution in green in Figure 2. But the desired result is the unimodal distribution in red, since that is what conditionalizing your uniform prior on the aggregate data would give.

However, when you use geometric pooling in this case, it performs considerably better. It won't be exactly the same as conditionalizing your shared prior on the aggregate evidence, but it will be similar. The proportions will be right, but some information will be lost: it will be as if the sample sizes were cut in half. To illustrate, suppose again that you observe 70 heads out of 100 flips, while your fellow observes 30 heads out of a separate 100 flips. If you pool your posteriors over the biases, the result will be the same as if you had conditionalized your shared prior on 50 heads out of 100 flips.

Figure 3 illustrates this example in the case where the shared prior was

[4] Figure 4 illustrates some of the Beta distributions. We will meet them again shortly.

[5] This is noted by Winkler (1968, B68, Figure 2), who illustrates the point with a plot similar to Figure 2.

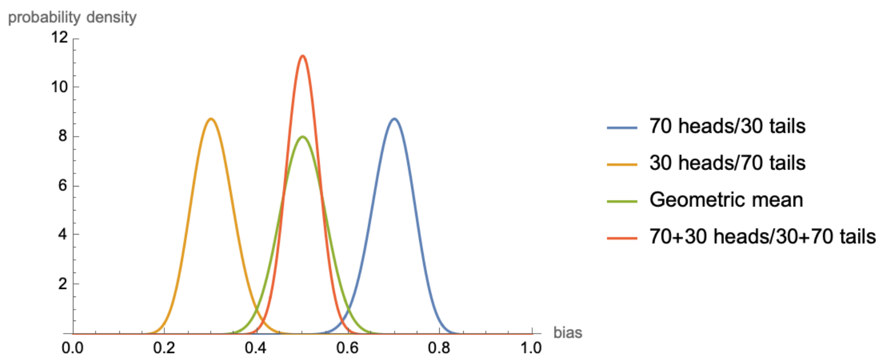

Figure 3: The red line gives the distribution obtained by conditionalizing your shared prior on the aggregate evidence, while the green line gives the geometric pool of your posterior and your fellow's.

uniform over the possible biases.[6] The geometric mean (red) only approximates conditionalizing on the aggregate evidence (green). But it does much better than arithmetic averaging did: compare Figure 2.

The general result being illustrated in Figure 3 is that pooling opinions about the coin's bias geometrically is equivalent to conditionalizing on half the aggregate data. If you observe k heads out of m flips, and your fellow observes l heads out of n flips, then the geometric pool of your posteriors is the same as conditionalizing on $(k+l)/2$ heads out of $(m+n)/2$ flips.[7]

Proposition 4. *Let X be the number of heads in the first m flips, Y the number of heads in the next n flips, and Z the number of heads in some sequence of $(m+n)/2$ flips. Where, recall, $\mathcal{H}$ is the partition of possible biases,*

$$G_{\mathcal{H}}(P(- \mid X = k), P(- \mid Y = l)) = P(- \mid Z = (k+l)/2).$$

One way to think about this result is that taking the geometric mean over the possible biases correctly gleans the direction the aggregate evidence points in, but with its force or magnitude understated.

What's more, Proposition 4 extends to propositions beyond the chance hypotheses. Once geometric pooling fixes your new probabilities in the possible biases of the coin, the Principal Principle steps in and does the rest, determining your new probabilities in each of the possible sequences in such a way that your new probability in a sequence is the probability you'd assign if you were to conditionalize your prior on the halved aggregate

[6]Again, this is noted by Winkler (1968, B68, Figure 2).

[7]See the Appendix for a more rigorous statement and proof: Proposition 4 is the special case of Theorem 6 where $\alpha = 1/2$ and $s = 2$. Theorem 6 generalizes results that are noted in passing by Winkler (1968, B64-5) and Morris (1983, Section 6), who consider the special case where the priors are Beta distributions.

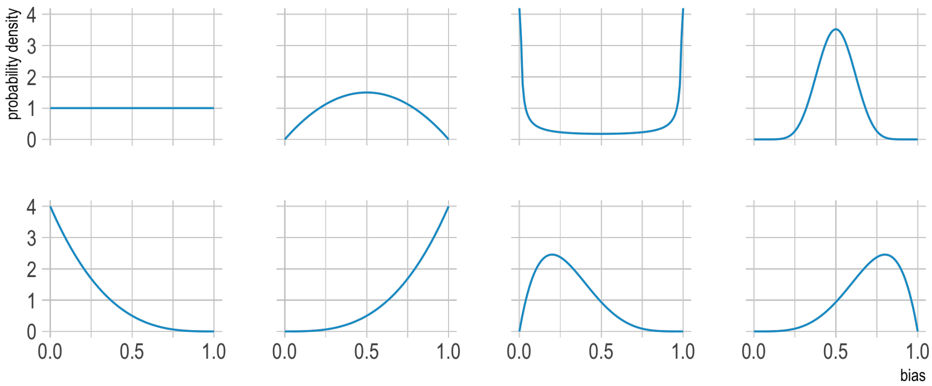

Figure 4: Some examples of Beta distributions

sample.[8]

Another way to think about Proposition 4 connects back to arithmetic pooling. Loosely speaking, taking the geometric pool over the biases is the same as taking the arithmetic mean of the underlying, aggregate data. Rather than combining your positive observations with your fellow's, you split the difference: $(k + l)/2$.

This perspective suggests that pooling the probabilities over the chance hypotheses geometrically will have the effect of pooling the probabilities over the predictions arithmetically, at least approximately. In fact, when the two agents' samples are of equal size ($m = n$), it has exactly this effect for a wide range of priors known as the Beta distributions (which include the uniform distribution).

Proposition 5. *Let X, Y, Z, and S be as in Proposition 4, and let H be the event of heads on an unobserved flip. If $m = n$ and P has a Beta distribution over the possible biases in the partition $\mathcal{H}$, then*

$$G_{\mathcal{H}}(P(- \mid X = k), P(- \mid Y = l))(H) = \frac{P(H \mid X = k) + P(H \mid Y = l)}{2}.$$

Insofar as pooling predictions arithmetically is attractive, this result helps to make pooling over chances geometrically attractive as well. The two methods go hand-in-hand in a range of interesting cases. To give a sense of the result's scope, Figure 4 illustrates the variety of Beta distributions by way of some examples.

[8]See Footnote 2 for the statement of the Principal Principle. Note that, if you conditionalize your prior on the halved aggregate sample, *i.e.*, $Z = (k + l)/2$, then the Principal Principle requires of your posterior that $P(S \mid H_i \,\&\, Z = (k + l)/2) = C_i(S \mid Z = (k + l)/2)$.

5 Three generalizations

Proposition 4 tells us that geometric pooling over the chance hypotheses effectively averages the data acquired by the two parties, provided they share a common prior. This result raises three natural questions, which we'll answer in this section by generalizing Proposition 4 in three ways.

First, what if there are more than two possible outcomes? So far we've only discussed coin flips, but the result of Proposition 4 extends to processes with more than two possible outcomes. For example, suppose you roll a six-sided die behind a screen 100 times and conditionalize on what you observe. Then your fellow does 100 rolls of their own and conditionalizes on what they observe. Let the data you each observe be as follows:

	⚀	⚁	⚂	⚃	⚄	⚅
Your Data	10	10	10	20	20	30
Their Data	30	20	20	10	10	10

If you then geometrically pool your posteriors over the possible biases, the result will be the same as if you had conditionalized your common prior on a sample of 100 rolls where the counts are:

	⚀	⚁	⚂	⚃	⚄	⚅
Average Data	20	15	15	15	15	20

We'll state a general theorem that captures this momentarily.

Second, what if you and your fellow begin with different priors? Once again, Proposition 4 generalizes in a natural way: the result of geometric pooling is still the same as conditionalizing on the average data, except that the prior being conditionalized is the geometric pool of your prior and your fellow's. Informally, conditionalizing and then pooling has the same effect as pooling and then conditionalizing on the averaged data.[9]

This generalization is formally elegant, but it won't always be philosophically satisfying. The issue is the same one we raised for Proposition 3: you might only be interested in your fellow's opinion for the data that informs it. In which case, you'll want to conditionalize your own prior, not the geometric pool of your respective priors.

Still, we are often interested in others' opinions for more than just the evidence that informs them. And if you think there's something to your fellow's way of *interpreting* data, then you might want to incorporate some of that into your own prior. Suppose specifically that, if you were to learn what their prior was, then you would adopt the geometric pool of your

[9]Note that this is different from the "external Bayesianity" property introduced by Madansky (1964). External Bayesianity concerns the case where both agents learn the same evidence, while here they learn different evidence.

respective priors. Then this generalization is just right. It amounts to conditionalizing on what your fellow's prior was, and then conditionalizing on the average data.

Third and finally, what would it take to get the full body of aggregate data, rather than the average data where the sample size is cut in half? It helps here to revisit the definition of geometric pooling, with the radical sign rewritten as an exponent. That is, we write $\sqrt{x}$ as $x^{1/2}$:

$$G_{\mathcal{A}}(P,Q)(A_i) = \frac{(P(A_i)Q(A_i))^{1/2}}{\sum_{j=1}^{n} \left(P(A_j)Q(A_j)\right)^{1/2}}.$$

A natural thought is that the sample size gets cut in half because of the $1/2$ exponent. And this thought suggests a more general definition, where we let the exponent be any proportion $0 \leq \alpha \leq 1$:

$$M^{\alpha}_{\mathcal{A}}(P,Q)(A_i) = \frac{(P(A_i)Q(A_i))^{\alpha}}{\sum_{j=1}^{n} \left(P(A_j)Q(A_j)\right)^{\alpha}}.$$

For a given choice of α, call the resulting pooling rule *multiplicative$^{\alpha}$ pooling*.

Our third generalization says that α is the proportion of the aggregate data that gets conditionalized on. In terms of coin flips, if you observe k heads out of m tosses and your fellow observes l heads out of n tosses, then multiplicative$^{\alpha}$ pooling over the biases is the same as conditionalizing on $\alpha(k+l)$ heads out of $\alpha(m+n)$ tosses. That is,

$$M^{\alpha}_{\mathcal{H}}(P(- \mid X = k), P(- \mid Y = l)) = P(- \mid Z = \alpha(k+l)),$$

where Z is the number of heads in some sequence of $\alpha(m+n)$ tosses.

The answer to our third question then is: set $\alpha = 1$, rather than $1/2$. Then, pooling with your fellow will give the same result as conditionalizing on the full body of aggregated data, theirs and yours together. In Figure 3 for example, the yellow and blue curves will then combine to give the desired red curve, rather than the green approximation to it.[10]

Why bother with geometric pooling then? Why would we ever want to set $\alpha = 1/2$ instead of 1? Because we often want to pool with the same person on more than one occasion, and in that case choosing $\alpha = 1$ can be disastrous. Every time you pool with your fellow using $\alpha = 1$, all their data gets counted anew. And this means their data gets double counted the second time you pool with them, triple counted the third time, and so on.

[10]Setting $\alpha = 1$ gives the pooling rule that Easwaran et al. (2016) call *upco*, and Dietrich (2010) and Dietrich and List (2016) call *multiplicative pooling*. They show that it has several desirable properties, and in (Authors, manuscript) we derive several others relevant to evidence aggregation. But, since geometric pooling is our focus here, we only touch on multiplicative pooling briefly in this section.

For example, suppose you and your fellow each do one flip, pool, then flip again and pool again. And let's imagine the sequence you observe is HT, while they observe TT. If $\alpha = 1$, then the second round of pooling will effectively double count the flips from the first round. The final result after two rounds of pooling will be as if you had observed the sequence $HTHTTT$. So the sample size will actually be inflated, with two non-existent flips; and the frequency will be off too, with $1/3$ heads instead of the true $1/4$. Further iterations will compound this effect, yielding even more distorted results.

Choosing $\alpha = 1/2$, however, avoids this problem entirely. The frequency of heads and tails in the total, aggregate data is then always accurately reflected in the results of pooling. After repeated iterations of observing-and-pooling, if $1/3$ of the observed flips have been heads, then the agents' posteriors will be as if they had conditionalized on observing a sequence that is $1/3$ heads.

So there is a tradeoff here. When pooling just once, $\alpha = 1$ avoids leaving evidence on the table. But for repeated pooling, only $\alpha = 1/2$ avoids double-counting. The catch is that the sample size is cut in half. In the next section we'll look at an application where this tradeoff plays a crucial role.

Let's bring together the three points of this section. First, the result of Proposition 4, which we initially presented in terms of coin tosses, generalizes to die rolls and other events with more than two possible outcomes. Second, it also generalizes elegantly to the case where you and your fellow have different priors—although the philosophical significance of this generelaization depends on your appraisal of your fellow's way of interpreting data. Third and finally, the result also generalizes to multiplicative$^\alpha$ pooling: pooling over the chance hypotheses amounts to conditionalizing on a similar sample whose size is α of the true sample's size.

The following result incorporates all three generalizations, phrased in terms of rolls of an s-sided die.

Theorem 6. *Let $\mathbf{X} = (X_1, \ldots, X_s)$ be the vector of counts from the first m rolls, $\mathbf{Y}$ the vector of counts from the next n rolls, and $\mathbf{Z}$ the vector of counts from some sequence of $\alpha(m+n)$ rolls. If P and Q have overlapping supports on the partition of possible biases $\mathcal{H}$, then*

$$M^\alpha_\mathcal{H}(P(- \mid \mathbf{X} = \mathbf{k}), Q(- \mid \mathbf{Y} = \mathbf{l})) = M^\alpha_\mathcal{H}(P, Q)(- \mid \mathbf{Z} = \alpha(\mathbf{k} + \mathbf{l})).$$

Informally put, taking the multiplicative$^\alpha$ pool of your posterior with your fellow's has the same result as conditionalizing the multiplicative$^\alpha$ pool of your respective priors on a sample of $\alpha(m+n)$ rolls, where the observed counts are given by $\alpha(\mathbf{k} + \mathbf{l})$.

6 Beyond conditionalization: Douven on abduction

In this section we put Proposition 4 to work. We'll show that it explains the success of certain social-epistemic strategies observed in recent computer simulations. And it identifies an even better strategy, one that is provably optimal in an important sense.

Douven (2019, 2022) uses simulations to defend alternative update rules to conditionalization (see also Douven and Wenmackers, 2017). In these simulations, agents gather evidence about the bias of a coin by flipping it privately, updating on the result using either conditionalization or some alternative. Then they report their opinions about the possible biases to one another, and update on this social evidence using geometric pooling.[11]

Interestingly, the communities that do best in terms of accuracy[12] in these simulations tend not to use conditionalization. Douven studies a range of alternative update rules, drawing inspiration from the idea of inference to the best explanation. Different rules do better under different circumstances, but for our purposes we can focus just on the simplest one, which Douven dubs EXPL.

The EXPL rule is very similar to Bayes' rule, but with greater emphasis on fit-to-the-data. When updating on a body of evidence E, EXPL adds a bonus quantity c to the hypothesis with the highest likelihood, $P(E \mid H_i)$. The new probability of hypothesis H_i, call it $P'(H_i)$, is

$$P'(H_i) = \frac{P(H_i)P(E \mid H_i) + c_i}{\sum_i [P(H_i)P(E \mid H_i) + c_i]},$$

where $c_i = c$ if H_i maximizes $P(E \mid H_i)$, and $c_i = 0$ otherwise.

The value we choose for c determines how much our simulated agents emphasize fit-to-the-data. At the start of each simulation, we'll fix a value of c and keep it constant throughout. But we'll experiment with different values in different simulations, to see what works best. Following Douven, we'll try values from 0 to 1, in 0.1 increments. Notice that, when $c = 0$, the EXPL formula just is Bayes' theorem, so EXPL has conditionalization as a special case.

In addition to c, two other parameters also need to be chosen at the start of each simulation. One is the actual bias of the coin, p, which is what the agents are trying to discover. Here again we'll experiment with values in the range from 0 to 1 in increments of 0.1. Thus there are 11 hypotheses for our simulated agents to consider: H_i is the hypothesis that the true bias is $p = i/10$, where $i \in \{0, 1, \ldots, 10\}$.

[11] Douven also experiments with linear and harmonic pooling, but finds that geometric pooling performs best in terms of accuracy. So we'll focus on just geometric pooling here.

[12] In addition to accuracy, Douven also considers the goal of speed, understood as how soon a majority of the community becomes confident in the true hypothesis. We will focus exclusive on accuracy here, to keep the discussion manageable.

Finally, a third parameter, ϵ, controls how "open-minded" agents are. In Douven's simulations, agents only pool with those whose opinions are within a certain distance ϵ of their own. The distance between opinions is measured by the sum of absolute differences:[13]

$$\sum_{i=0}^{10} |P(H_i) - Q(H_i)|.$$

Since the maximum possible distance is 2, we will consider values of ϵ in the range from 0 to 2, in increments of 0.1. (Douven only considers values up to 1, but we'll see that this omits important results.)

So here's how each simulation works. At the start, we pick values for each of the three parameters c, p, and ϵ. Then we create 50 agents, all with a uniform prior over the possible biases. Each agent performs one flip of the coin privately, and updates on the result using the EXPL rule with the chosen value of c. Then they pool geometrically with everyone within distance ϵ of their own opinion. This flip-update-pool cycle then repeats, for a total of 500 cycles.

At the end of each cyle, after pooling, we gauge the community's accuracy. More precisely, each individual agent's inaccuracy is evaluated using the Brier score,

$$\sum_{i=0}^{10} (P(H_i) - V(H_i))^2,$$

where $V(H_i) = 1$ if H_i is true, and 0 otherwise. The average Brier score of all 50 agents is then calculated, and at the end of all 500 cycles these averages are summed to generate an overall score for that simulation.

Figure 5 shows the results.[14] Each square represents a choice of values for c, p, and ϵ. The square's colour shows the expected inaccuracy for a community using that combination of values, based on 50 simulations averaged together. Lower scores are better, and the best ones in each panel are marked with an asterisk.

The conditionalizers are the bottom row of each panel, where $c = 0$. But the communities with the best scores are consistently those with high values of c, and also high values of ϵ. That is, the most accurate communities have agents who strongly favour hypotheses that fit their private data, but are also very "open-minded" in that they pool even with those whose opinions differ greatly from their own.

[13] Douven considers a variety of distance measures, but again we'll focus on just one for simplicity.

[14] Compare Figure 3 in Douven (2019) and Figure 7.3 in Douven (2022). Note that our results look slightly different, for two reasons. First, we explore the full range of possible values for ϵ. Second, our simulations use the definition of EXPL Douven states in the text, which differs from the implementation in the accompanying code. Nothing we will say hangs on this second difference; either way, the results are the same in the respects that matter here.

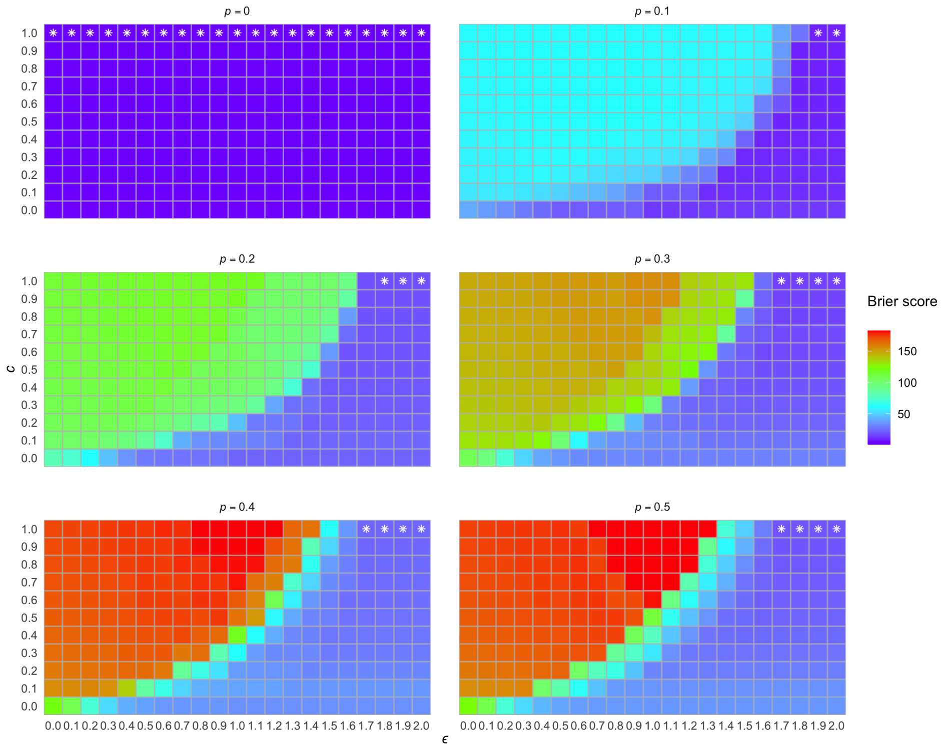

Figure 5: Simulation results for communities using EXPL and geometric pooling. Each square shows the expected Brier score for a choice of c, p, and ϵ, based on 50 *runs averaged together. The white asterisks mark the best scores in each panel.*

A natural thing to wonder about this result is: how is it possible? According to an influential theorem of Greaves & Wallace (2006), conditionalization minimizes expected Brier score. So how can conditionalization be dominated by an alternative rule, as it is here?

The answer is that the $c = 0$ agents only conditionalize on their private evidence. They do not conditionalize on the opinions of their fellows; instead they use geometric pooling. And Proposition 4 tells us they suffer a kind of "data loss" as a result. With two agents, geometric pooling effectively cuts each agent's sample in half. When 50 agents pool geometrically, their samples are effectively scaled down by 1/50. So the "conditionalizers" in these simulations are actually leaving a lot of evidence on the table.

This analysis suggests a better way for these agents to update on their private evidence. Instead of using EXPL, they should conditionalize on their data but scaled up by a factor of 50. That is, if an agent observes 1 head, they should conditionalize on the proposition that they observed 50 heads instead. Ditto for tails. If we then set $\epsilon = 2$, so that all 50 agents always pool with each other, the result will be the same as if they had all conditionalized on the aggregate evidence, by Proposition 4.

We could test this method by running new simulations, but we don't have to. By Greaves & Wallace's theorem, conditionalizing on the aggregate evidence is the optimal strategy for minimizing expected Brier score.[15] Since our proposal is equivalent, it too is optimal.[16] It won't be optimal within every panel, *i.e.* for every value of p. But no feasible rule can be. It does best on average though, since it has the lowest expected Brier score relative to a uniform prior over p. If we do repeated simulations, picking a random value for p each time, no procedure can do as well on average.

So we agree with Douven that, when conditionalizing on your social evidence is unavailable, there is a way to compensate for the loss of expected accuracy that results: change how you update on your private evidence. But we disagree on how best to do that. Douven says you should use a rule like EXPL with a high value of c.[17] We say you should conditionalize on a scaled up version of the sample you in fact witnessed. This proposal is provably better in expectation.

Proposition 4 also explains why a high-c, high-ϵ strategy did best in Figure 5. Updating on private evidence using EXPL with a high c approximates our proposal of conditionalizing on a scaled up sample. Both are ways of "over-fitting" your private evidence—of giving special favour to the hypotheses that best fit your actual private evidence. Coupled with a high value of ϵ and geometric pooling then, a high c approximates conditionalization on the aggregate evidence.

One consequence of this is that, as the number of other agents in the group decreases, so does the optimal value of c. This is because the optimal way to update, if you're then going to pool the results geometrically, is to update on your private evidence scaled up by the size of the group. And as the group size diminishes, using EXPL with lower values of c will approximate that optimal solution more closely. Indeed, as we see in Figure 6, if there are just five agents then the optimal value of c is zero, which is just conditionalization.

[15] Strictly speaking, their result only addresses a single agent's Brier score. But it extends readily to a group's average Brier score, when its members share a common prior, as here. And indeed it applies not only to the Brier score but to any strictly proper scoring rule.

[16] Note that, in Douven's set up, we only measure the accuracy of our agents' probabilities after they've updated on both the private and social evidence from a given round; we do not measure their accuracy after updating on the private evidence and then measure it again after they've updated on the social evidence. If we were to do that, neither our proposal nor Douven's would be optimal; if we were to do that, the only optimal approach would be to update by Bayes' Rule on both private and social evidence. By measuring accuracy only after both private and social updates have taken place, we can get the accuracy benefits of Bayes' Rule by using a strategy for private update and a strategy for social update that, when combined, match Bayes' Rule perfectly.

[17] Actually, Douven finds that another rule he calls "Popper's Rule" outperforms EXPL. But the same conclusion applies: our proposal is provably better in expectation, because it is equivalent to conditionalizing on the aggregate evidence.

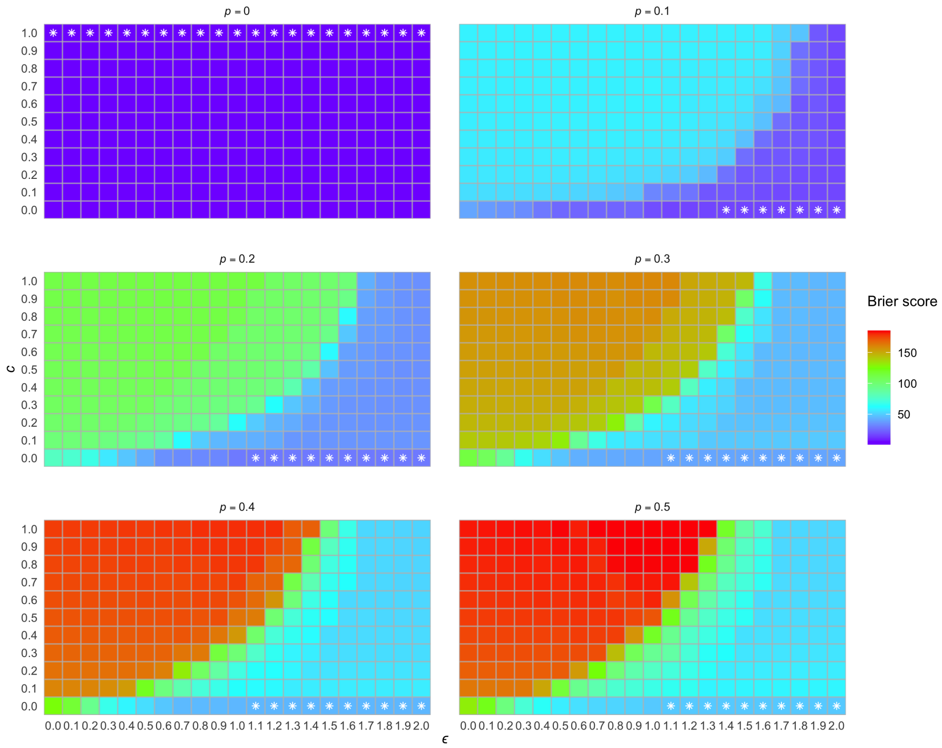

Figure 6: The same setup as Figure 5, but with 5 *agents rather than* 50.

7 Conclusion

How well, then, does geometric pooling serve as a means to combine your probabilities with your fellow's, when you wish to gain the benefit of their evidence?

If your fellow shares their probabilities concerning the next coin toss, geometric pooling does not serve you well. But arithmetic pooling does, so long as you don't use Jeffrey conditionalization to extend the result to other propositions.

If your fellow shares their probabilities in the possible sequences of coin tosses, geometric pooling does serve you well, as long as you shared a prior with your fellow or you wish to aggregate your fellow's prior with your own. But many pooling operators do this, including harmonic pooling. And if you do not wish to aggregate your prior, there are alternatives that are better. What's more, it would be easier simply to share the evidence itself than to share the credences in the sequences to which it has given rise.

Finally, if your fellow shares their probabilities in the chance hypotheses, geometric pooling leads to a posterior that points in the same direction as the aggregate evidence, but with less conviction than the pooled evidence warrants. It's as if you've updated on half the evidence. And this observation allows us to see why the abductive updating rules studied by Douven seem to do well as a private updating rule, when paired with geometric pooling as your social updating rule: abductive updating approximates the posterior you'd obtain by conditionalizing on an appropriately scaled up version of your private evidence. But it also allows us to see a private updating rule that will do better on average than Douven's.

So, in the end, it's a mixed bag. Geometric pooling has some features that make it attractive when your purpose is to extract evidence from your fellow's probabilities. And it outperforms arithmetic pooling in some cases, though not all. But it must be handled with care.

8 Appendix

In this Appendix, we give formal statements and proofs of the results in the main text. We first prove Proposition 3, which establishes Proposition 1 as a special case; then we prove Proposition 2. Next we prove Theorem 6, which has Proposition 4 as a special case, before finally proving Proposition 5. This covers all six results stated in the paper.

We begin by formally defining the family of multiplicative pooling rules, of which geometric pooling is a special case.

Definition 1 (Multiplicative pool; discrete case)**.** *Let P and Q be probability functions. And let $\mathcal{A} = \{A_i\}$ be a countable partition on which P and Q are both defined. Suppose $P(A_i), Q(A_i) > 0$ for at least one A_i. Then,*

$$M^{\alpha}_{\mathcal{A}}(P,Q)(A_i) = \frac{(P(A_i)Q(A_i))^{\alpha}}{\sum_{A_j \in \mathcal{A}}(P(A_j)Q(A_j))^{\alpha}}.$$

We call $M^{\alpha}_{\mathcal{A}}(P,Q)$ the multiplicative$^{\alpha}$ pool of P and Q over $\mathcal{A}$.

Note that $M^{\frac{1}{2}}_{\mathcal{A}}(P,Q)$ is the geometric pool of P and Q, and we write $G_{\mathcal{A}}(P,Q)$.

Definition 2 (Multiplicative pool; continuous case)**.** *Let f and g be probability density functions on $\mathbb{R}^n$ with overlapping supports. Let D be that overlap, and fix $\alpha \geq 0$. Then for $\mathbf{x} \in \mathbb{R}^n$ we define*

$$M^{\alpha}(f,g)(\mathbf{x}) = \frac{(f(\mathbf{x})g(\mathbf{x}))^{\alpha}}{\int_D (f(\mathbf{x})g(\mathbf{x}))^{\alpha}}.$$

We call $M^{\alpha}(f,g)$ the multiplicative$^{\alpha}$ pool of f and g.

Again $M^{\frac{1}{2}}(f,g)$ is the geometric pool of f and g, which we write $G(f,g)$.

Now we prove a general version of Proposition 3. Since the result holds not only for geometric pooling ($\alpha = 1/2$), but for multiplicative$^{\alpha}$ pooling in general, we state and prove this more general version.

Proposition 7. *Let E and F be subsets of $\mathcal{S}$ such that $P(- \mid E)$ and $Q(- \mid F)$ have overlapping support on $\mathcal{S}$. Then for any state S in $\mathcal{S}$,*

$$\mathcal{M}^{\alpha}_{\mathcal{S}}(P(- \mid E), Q(- \mid F))(S) = \mathcal{M}^{\alpha}_{\mathcal{S}}(P, Q)(S \mid EF).$$

Proof. First we analyze $\mathcal{M}^{\alpha}_{\mathcal{S}}(P, Q)(S \mid EF)$. By definition,

$$\mathcal{M}^{\alpha}_{\mathcal{S}}(P, Q)(S) = \frac{(P(S)Q(S))^{\alpha}}{\sum_{S \in \mathcal{S}} (P(S)Q(S))^{\alpha}}.$$

So for $S \in EF$,

$$\mathcal{M}^{\alpha}_{\mathcal{S}}(P, Q)(S \mid EF) = \frac{(P(S)Q(S))^{\alpha}}{\sum_{S \in EF} (P(S)Q(S))^{\alpha}}, \tag{1}$$

with $\mathcal{M}^{\alpha}_{\mathcal{S}}(P, Q)(S \mid EF) = 0$ for $S \notin EF$.

Now consider $\mathcal{M}^{\alpha}_{\mathcal{S}}(P(- \mid E), Q(- \mid F))$. For $S \in E$,

$$P(S \mid E) = \frac{P(S)}{\sum_{S \in E} P(S)},$$

with $P(S \mid E) = 0$ for $S \notin E$. Similarly, for $S \in F$,

$$Q(S \mid F) = \frac{Q(S)}{\sum_{S \in F} Q(S)},$$

with $Q(S \mid F) = 0$ for $S \notin F$. Thus for $S \in EF$, where c is the requisite normalizing constant,

$$\begin{aligned} \mathcal{M}^{\alpha}_{\mathcal{S}}(P(- \mid E), Q(- \mid F))(S) &= \frac{1}{c} \left(\frac{P(S)}{\sum_{S \in E} P(S)} \frac{Q(S)}{\sum_{S \in F} Q(S)} \right)^{\alpha} \\ &= \frac{(P(S)Q(S))^{\alpha}}{c(\sum_{S \in E} P(S) \sum_{S \in F} Q(S))^{\alpha}}. \end{aligned} \tag{2}$$

While for $S \notin EF$ we have $\mathcal{M}^{\alpha}_{\mathcal{S}}(P(- \mid E), Q(- \mid F))(S) = 0$.

Now observe that the numerators in equations (1) and (2) are the same. For both distributions, the nonzero probability masses are proportional to $(P(S)Q(S))^{\alpha}$ when $S \in EF$, while for $S \notin EF$ both assign mass 0. Hence these must actually be the same distribution. ∎

This establishes Proposition 3, which is the special case where $\alpha = 1/2$. Proposition 1 then follows as the further special case where $P = Q$.

Next, we prove Proposition 2, which we restate here for convenience.

Proposition 2 (restated). *If E and F are subsets of $\mathcal{S}$ such that $P(EF) > 0$, then for any state S in $\mathcal{S}$,*

$$H_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S) = P(S \mid EF).$$

Proof. First consider the case where $S \notin EF$. Then either $P(S \mid E) = 0$ or $P(S \mid F) = 0$, so $H_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S) = 0$ by definition.

Now suppose $S \in EF$. Then by the definition of harmonic pooling, where c is an appropriate normalizing constant:

$$\begin{aligned}
H_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S) &= c\frac{P(S \mid E)P(S \mid F)}{P(S \mid E) + P(S \mid F)} \\
&= c\frac{P(S)^2/[P(E)P(F)]}{P(S)/P(E) + P(S)/P(F)} \\
&= c\frac{P(S)/[P(E)P(F)]}{1/P(E) + 1/P(F)} \\
&= c\frac{P(S)}{P(F) + P(E)}.
\end{aligned}$$

Thus $H_{\mathcal{S}}(P(- \mid E), P(- \mid F))(S)$ is proportional to $P(S)$ for $S \in EF$, and is zero otherwise. This implies $H(P(- \mid E), P(- \mid F))(S) = P(S \mid EF)$, as desired. ■

That establishes the three results of Section 3, where the topic was pooling over sequences. Next we turn to the results for pooling over chance hypotheses, which were the topic of Sections 4 and 5.

We first give a more formal statement of Theorem 6. Here we consider only the continuous case, as the discrete case runs closely parallel.

Theorem 6 (formal)**.** *Let $\mathbf{A}_1, \ldots, \mathbf{A}_{m+n}$ be categorical random vectors of length s. Fix a value $0 \leq \alpha \leq 1$ such that $\alpha(m+n)$ is an integer, and let*

$$\mathbf{X} = \sum_{i=1}^{m} \mathbf{A}_i, \quad \mathbf{Y} = \sum_{i=m+1}^{n} \mathbf{A}_i, \quad \mathbf{Z} = \sum_{i=1}^{\alpha(m+n)} \mathbf{A}_i.$$

Let f, g, and h be probability density functions such that the $\mathbf{A}_i$ are i.i.d. with parameter vector $\mathbf{T} = (T_1, \ldots, T_s)$. Write $f_{\mathbf{T}}$ for the marginal distribution of f over $\mathbf{T}$, and similarly for $g_{\mathbf{T}}$ and $h_{\mathbf{T}}$. If $h_{\mathbf{T}} = M^{\alpha}(f_{\mathbf{T}}, g_{\mathbf{T}})$, then

$$M^{\alpha}\left(f_{\mathbf{T}|\mathbf{X}}(- \mid \mathbf{k}), g_{\mathbf{T}|\mathbf{Y}}(- \mid \mathbf{l})\right)(\mathbf{t}) = h_{\mathbf{T}|\mathbf{Z}}(\mathbf{t} \mid \alpha(\mathbf{k} + \mathbf{l})). \tag{3}$$

Proof. We first analyze the left-hand side of equation (3). By Bayes' theorem and our i.i.d. assumption,

$$\begin{aligned}
f_{\mathbf{T}|\mathbf{X}}(\mathbf{t} \mid \mathbf{k}) &= c_1 f(\mathbf{t}) \prod_i t_i^{k_i}, \\
g_{\mathbf{T}|\mathbf{Y}}(\mathbf{t} \mid \mathbf{l}) &= c_2 g(\mathbf{t}) \prod_i t_i^{l_i},
\end{aligned}$$

where c_1 and c_2 are appropriate normalizing constants. By the definition of M^α, where c_3 is another normalizing constant:

$$\begin{aligned} M^\alpha\left(f_{\mathbf{T}|\mathbf{X}}(- \mid \mathbf{k}), f_{\mathbf{T}|\mathbf{Y}}(- \mid \mathbf{l})\right)(\mathbf{t}) &= c_3\left(c_1 f(\mathbf{t})\prod_i t_i^{k_i}\, c_2\, g(\mathbf{t})\prod_i t_i^{l_i}\right)^\alpha \\ &= (c_1c_2)^\alpha c_3\left(f(\mathbf{t})g(\mathbf{t})\right)^\alpha \prod_i t_i^{\alpha(k_i+l_i)}. \end{aligned} \tag{4}$$

Now we analyze the right-hand side of equation (3). By hypothesis,

$$h_{\mathbf{T}}(\mathbf{t}) = c_4\left(f(\mathbf{t})g(\mathbf{t})\right)^\alpha,$$

where c_4 is the appropriate normalizing constant. So, by Bayes' theorem and the i.i.d. assumption,

$$h_{\mathbf{T}|\mathbf{Z}}\left(\mathbf{t} \mid \alpha(\mathbf{k}+\mathbf{l})\right) = c_5\left(f(\mathbf{t})g(\mathbf{t})\right)^\alpha \prod_i t_i^{\alpha(k_i+l_i)}, \tag{5}$$

where c_5 is again an appropriate normalizing constant. Since equations (4) and (5) are proportional, these must actually be the same distribution. ■

Observe that, in the special case of Theorem 6 where $\alpha = 1/2$ and $s = 2$, equation (3) becomes

$$G\left(f_{T|X}(- \mid k), g_{T|Y}(- \mid l)\right)(t) = h_{T|Z}(t \mid (k+l)/2).$$

This is the statement of Proposition 4 in the continuous case.

Finally, we prove Proposition 5, whose formal statement is as follows.

Proposition 5 (formal). *Let $A_1, \ldots, A_{m+n}$ and H be Bernoulli random variables, with X, Y, and Z as in Theorem 6. Let P be a probability function such that the A_i and H are i.i.d. with shared parameter T, with f the associated p.d.f. And let R be a probability function with h its associated p.d.f., such that $H \sim \mathrm{Bern}(T)$ and*

$$h_T = G\left(f_{T|X}(- \mid k), f_{T|Y}(- \mid l)\right).$$

If $m = n$ and f_T is $\mathrm{Beta}(a, b)$, then

$$R(H) = \frac{P(H \mid X = k) + P(H \mid Y = l)}{2}. \tag{6}$$

Proof. We begin by analyzing the right-hand side of equation (6). Because of the conjugate relationship between the Beta and binomial distributions, $f_{T|X}(- \mid k)$ and $f_{T|Y}(- \mid l)$ have the following Beta distributions:

$$\begin{aligned} T \mid X = k &\sim \mathrm{Beta}(a+k, b+m-k), \\ T \mid Y = l &\sim \mathrm{Beta}(a+l, b+m-l). \end{aligned}$$

Now, by the law of total probability, the probability of a Bernoulli random variable like H is the expected value of T. Since the expected value of a $\text{Beta}(x, y)$ distribution is $x/(x+y)$, this gives us

$$P(H \mid X = k) = \frac{a+k}{a+b+m},$$
$$P(H \mid Y = l) = \frac{a+l}{a+b+m}.$$

Taking the arithmetic average yields

$$\frac{P(H \mid X = k) + P(H \mid Y = l)}{2} = \frac{a + (k+l)/2}{a+b+m}. \tag{7}$$

Now we consider equation (6)'s left-hand side. By assumption,

$$h_T = G\left(f_{T|X}(- \mid k), f_{T|Y}(- \mid l)\right),$$

which we can rewrite using Theorem 6 with $\alpha = 1/2$ as:

$$h_T = f_{T|Z}(- \mid (k+l)/2).$$

The right-hand side here is another Beta distribution:

$$T \mid Z = (k+l)/2 \quad \sim \quad \text{Beta}(a + (k+l)/2, b + m - (k+l)/2).$$

Since its expected value is $R(H)$, we have

$$R(H) = \frac{a + (k+l)/2}{a + (k+l)/2 + b + m - (k+l)/2} = \frac{a + (k+l)/2}{a+b+m}. \tag{8}$$

As equations (7) and (8) are equal, this completes the proof. ■

References

Baccelli, Jean and Rush T. Stewart. 2023. "Support for Geometric Pooling." *The Review of Symbolic Logic* 16(1):298–337.

Briggs, R. A. and Richard Pettigrew. 2018. "An Accuracy-Dominance Argument for Conditionalization." *Noûs* 54(1):162–181.

Brown, Peter M. 1976. "Conditionalization and Expected Utility." *Philosophical Studies* 43(3):415–419.

Dietrich, Franz. 2010. "Bayesian Group Belief." *Social Choice and Welfare* 35(4):595–626.

Dietrich, Franz and Christian List. 2016. Probabilistic Opinion Pooling. In *Oxford Handbook of Philosophy and Probability*, ed. Alan Hàjek and Christopher Hitchcock. Oxford University Press pp. 519–42.

Douven, Igor. 2019. "Optimizing Group Learning: An Evolutionary Computing Approach." *Artificial Intelligence* 275:235–251.

Douven, Igor. 2022. *The Art of Abduction*. Cambridge: The MIT Press.

Douven, Igor and Sylvia Wenmackers. 2017. "Inference to the Best Explanation versus Bayes's Rule in a Social Setting." *British Journal for the Philosophy of Science* 68(2):535–70.

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.

Gallow, J. Dmitri. 2019. "Learning and Value Change." *Philosophers' Imprint* 19(29):1–22.

Genest, Christian and James V. Zidek. 1986. "Combining Probability Distributions: A Critique and an Annotated Bibliography." *Statistical Science* 1(1):114–135.

Greaves, Hilary and David Wallace. 2006. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind* 115(459):607–632.

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

Jeffrey, Richard C. 1965. *The Logic of Decision*. University of Chicago Press.

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.

Lewis, David. 1999. Why Conditionalize? In *Papers in Metaphysics and Epistemology*. Cambridge University Press pp. 403–7.

Madansky, Albert. 1964. Externally Bayesian Groups. Technical Report RM-141-PR The Rand Corporation.

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

Oddie, Graham. 1997. "Conditionalization, Cogency, and Cognitive Value." *British Journal for the Philosophy of Science* 48(4):533–541.

Russell, Jeffrey Sanford, John Hawthorne and Lara Buchak. 2015. "Groupthink." *Philosophical Studies* 172(5):1287–1309.

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