CrossMark

Learning from others: conditioning versus averaging

Richard Bradley[1] (ID)

Abstract How should we revise our beliefs in response to the expressed probabilistic opinions of experts on some proposition when these experts are in disagreement? In this paper I examine the suggestion that in such circumstances we should adopt a linear average of the experts' opinions and consider whether such a belief revision policy is compatible with Bayesian conditionalisation. By looking at situations in which full or partial deference to the expressed opinions of others is required by Bayesianism I show that only in trivial circumstances are the requirements imposed by linear averaging compatible with it.

1 Introduction

When others have information or judgemental capabilities that we lack, then their opinions are a resource that we can and should exploit for the purposes of forming or revising our own opinions. But how should we do this? In this paper I will compare two types of answer to this question—Bayesian conditioning and opinion pooling—and ask whether they are compatible. Some interesting work on the question suggests a positive answer under various conditions, see for instance the studies by Genest and Schervish (1985), Bonnay and Cozic (submitted) and Romeijn and Roy (submitted). But the question remains of how restrictive these conditions are.

✉ Richard Bradley
r.bradley@lse.ac.uk

1 Department of Philosophy, Logic and Scientific Method, London School of Economics,
Houghton Street, London WC2A 2AE, UK

A Bayesian treats the expressed opinions of others as evidence for and against the truth of the claim under consideration, evidence whose relevance is captured by an assignment of a conditional probability for the claim, given each possible combination of others' opinions. She responds to this evidence by conditioning on it, i.e. by adopting as her revised opinion her conditional degrees of belief given the expressed opinions. The opinion pooler, on the other hand, adopts as her new opinion an aggregate of the expressed opinions of others (and perhaps her own), an aggregate that in some way reflects the epistemic value that she attaches to each of the expressed opinions.

I shall assume here that Bayesianism provides the gold standard for coherent revision of belief in the kinds of situation in which it applies, namely when we have prior probabilities for not only the hypotheses of ultimate interest but also for all possible combinations of evidence (in this case, the expressions of opinion) that either confirm or disconfirm these hypotheses, and when everything that we learn is representable by one such possible combination. The problem is that it is not always easy to apply the Bayesian theory. In circumstances in which the evidence takes a 'non-standard' form, such as when it is imprecise or conditional in form, we can turn to other forms of belief revision, such as Jeffrey conditioning or Adams conditioning.[1] But when we are unable to assign prior probabilities to the possible evidence propositions, or determine associated likelihoods, then no conditioning method at all may be applicable. This can happen because there are simply too many possibilities for us to process them all, or because we do not have enough information to assign a precise probability with any confidence.

These difficulties are acutely pertinent to the question of how to exploit the information taking the form of expert opinion reports. Consider, for instance, someone who claims to be an expert on wines. What is the probability that they will make any particular judgement about any particular wine? If you do not know much about wine, it will be hard to say. In the statistics literature, agents who change their beliefs by conditionalising on the testimonial evidence of experts are known as supra-Bayesians [see, for instance, Morris (1974) and French (1981)]. Supra-Bayesians must have priors over the opinion states of all those whose opinions count as evidence for them with regard to some proposition. But opinions about opinions might be evidence too, and opinions about opinions about opinions. And so on. It would be fair to say that supra-Bayesians are required to be cognitive super-Humans.

A Bayesian with more limited cognitive resources has two reasons for taking an interest in opinion pooling. First, it might help her in thinking about how to assign the probabilities to the hypotheses, conditional on combinations of opinion, that she needs to revise her own opinions by conditionalisation when she gets information of this kind. Second, in circumstances in which she cannot conditionalise on expressions of opinion because she lacks the requisite conditional probabilities for the hypotheses that interest her, she might adopt opinion pooling as an alternative method of belief revision. In both cases, the diligent Bayesian will want to know whether a particular rule for opinion pooling is compatible with her commitment to conditionalisation. This is true not just when she uses opinion pooling as a guide to making conditional

---

[1] See Bradley (2007) for a discussion.

probability judgements, but also in the case when she uses it as an alternative to conditionalisation. For in this latter case, she will want to know that there is some way of assigning prior probabilities to combinations of expressed opinion and to the hypotheses that are evidentially dependent on them, such that the pooled opinion is what she would have obtained by conditionalisation if these had been her prior probabilities.

In this paper I will address the question of whether revision of belief in response to testimonial evidence by application of a well-known form of opinion pooling—linear averaging—is compatible with Bayesian norms of belief change together with some minimal conditions on appropriate respect for expert judgement. I will begin by defining this form of opinion pooling and making more precise what is required for it to be compatible with Bayesian revision. In subsequent sections, I investigate what Bayesianism requires of agents in some rather simple situations; in particular, ones which mandate deference to the opinions of one or more experts on some specific proposition. Finally I consider the consistency of these requirements with those implied by linear averaging, drawing particularly on Dawid et al. (1995). The conclusion is somewhat surprising. Linear averaging, even when applied to a single proposition, is not (non-trivially) compatible with Bayesian conditionalisation in situations involving more than one source of expert testimony to which deference is mandated.

2 Linear averaging and Bayes compatibility

When an agent revises her opinions by linear averaging she responds to the reported opinions of others by adopting a new opinion that is a weighted average of the reported opinions and her own initial one. When these weights are allowed to vary with the proposition under consideration, such a revision rule will be said to exhibit proposition dependence. More precisely, an agent with probabilistic degrees of belief P on a Boolean algebra of propositions revises her opinions on some subset Ω of it by proposition-dependent linear averaging in response to the testimonial evidence of n others just in case, for any proposition $X \in \Omega$ and profile $x_1, ..., x_n$ of reported probabilistic opinions on X, she adopts new probabilistic degrees of belief Q such that

$$Q(X) = \left(\sum_{i=1}^{n} x_i \cdot \alpha_i^X \right) + \left(1 - \sum_i \alpha_i^X \right) \cdot P(X) \tag{1}$$

where the α_i^X are non-negative weights such that $\sum_i \alpha_i^X \leq 1$, canonically interpreted as measures of the (relative) judgemental competence of individuals. Note that these weights (potentially) depend on X but not on what the others report on X.

There is a large statistics literature, and a modest philosophical one, on proposition-*independent* linear averaging and its properties, in which the weights are taken to be constant across propositions, see Genest and Zidek (1986) and Dietrich and List (2016) for good surveys.[2] Much of this literature is focused on the problem of how a group

---

[2] Because this literature largely considers aggregation only on full Boolean algebras, proposition-dependent weights are ruled out by the requirement that aggregate opinions be probabilities. See Dietrich and List (2017) for a more general discussion.

should form a consensual opinion on some issue, a question that is not directly related to that of how individuals should improve their own opinions by taking into account the information provided by the opinions of others. These two issues can come together, as they do in the theory of Lehrer and Wagner (1981) and Wagner (1982), if the way that the consensus is achieved is by individuals revising their beliefs rationally in response to the expressed opinions of others. But here we only concerned with the question of whether revising one's beliefs in this way is epistemically rational.

This is important because, whether or not proposition-independent linear averaging has merit as a means of forming a consensual or group probability from a diverse set of individual ones, it is seriously flawed as a method of belief revision where epistemic considerations are paramount [contrary to Lehrer's (1976) claim that rationality requires that we revise our beliefs in this way]. This is so for two important reasons. First, linear averaging is not sensitive to the proposition-specific competencies of individuals. But intuitively individuals have different domains of expertise: I do not take my plumber's pronouncements about health risks very seriously, for instance, but equally I would not get my doctor to tell me the cause of the drainage problems in my house. Second, proposition-independent linear averaging is insensitive to whether the opinions expressed by different individuals on the same proposition are independent or not. But the evidential significance of individuals' reported probability for some proposition is quite different when their judgements are arrived at independently than, say, when they are based on exactly the same information.

Neither of these problems affect the proposition-dependent version of linear averaging under consideration here, which leaves the learning agent free to assign weights to others which are sensitive to their competence on the specific proposition in question and any probabilistic dependencies between their reports (though not to tailor the weights to what they report). Likewise although it is well known that proposition-independent linear averaging is not generally consistent with Bayesian norms of revision [see for instance Bradley (2006)], the question of whether proposition-independent averaging is so remains open. To explore it, I will work within a rather restricted setting. Our protagonist will be an agent (You) who learns the opinions of n truthful experts on some particular contingent proposition X and who revises her own opinions on X in the light of this testimonial evidence. You and the n experts are assumed to be Bayesian reasoners in the sense of having probabilistic degrees of belief for all propositions of interest, which are revised by conditionalising on any evidence that is acquired. To keep things simple, I shall assume that all have probabilistic beliefs regarding the elements of a background Boolean algebra of propositions $\mathcal{S}$ that is sufficiently rich as to contain all propositions of interest: including not only X itself, but propositions expressing evidence relevant to X and, in particular, the reports of the experts on it.

Let Pr be a (given) prior probability measure on $\mathcal{S}$ measuring the initial degrees of belief regarding its elements of an agent (You). Let E be your total evidence at the time when You receive the experts' reports on X and let $P(\cdot) := \Pr(\cdot|E)$. So P is a constant in what follows. To avoid problems with zero probabilities when conditionalising, I assume that Pr is regular, i.e. that it assigns non-zero probability to all contingent propositions in $\mathcal{S}$. I shall say that an event (or a value of a random variable or a combination of values of random variables) is "possible" when it is

compatible with the background information E, so that in some world both the event (or value or combination of values) and E obtain. Given the regularity assumption on Pr, it follows that anything that is possible has positive probability given E: in particular, any possible expert report, any possible profile of expert reports, and any possible expert posterior (this allows us to meaningfully conditionalise on them).

Let $\mathcal{R}_i(X)$ be a random variable ranging over the possible probabilistic opinions on X that expert i will report. Note that the regularity assumption implies that only countably many expert reports are possible (otherwise, there would exist uncountably many mutually exclusive events of positive probability). Now revising your beliefs by linearly averaging reported opinions on X is compatible with Bayesian conditionalisation, given P, *only if* it satisfies the following constraint:[3]

(LAC) There exist non-negative weights $\alpha_0, \alpha_1, \ldots, \alpha_n$ such that $\sum_{i=0}^{n} \alpha_i = 1$ and such that for any possible profile, $x_1, ..., x_n$, of reported probabilistic opinions on X:

$$P(X|\mathcal{R}_1(X) = x_1, ..., \mathcal{R}_n(X) = x_n) = \left(\sum_{i=1}^{n} \alpha_i \cdot x_i\right) + \alpha_0 \cdot P(X)$$

We now want to know what restrictions LAC places on P—in particular, on your conditional probabilities for X given the experts' reports—with an eye to establishing whether these restrictions are consistent with reasonable responses to expert opinion. I will tackle the issue in the reverse direction, showing in the next section what Bayesianism requires us to do in response to the experts' reports of their probabilistic opinions before testing, in subsequent sections, for the compatibility of these requirements with LAC.

3 Deference to experts

As many authors have observed, we might defer to someone's opinion because we think that they hold information that we do not or because we believe them to have skills which make them better at judging the significance of the information that we both hold (see for instance, Joyce 2007; Elgar 2007). To investigate the former cases in isolation from the latter, let us assume that the experts share with You prior Pr over the background domain $\mathcal{S}$. In this case, because You and the experts are Bayesian reasoners, the posterior beliefs of the experts are just the beliefs that You would have adopted had You acquired the same evidence as them. Furthermore, any situation in which (You know that) they have acquired at least as much evidence as You, You should be disposed to adopt whatever degrees of belief they have acquired on the basis of this evidence. In short, Bayesianism requires that You should defer to the opinions of any of the experts who acquire more information than You.

To establish this claim more formally, let $\mathcal{E}_i$ be a (proposition-valued) random variable ranging over the possible total evidences of expert i and let $\mathcal{R}_i$ be a random

---

[3] I make no claim for LAC being a sufficient condition for Bayes compatibility, only that it is necessary. Presumably consistency across a sequence of revisions would also be required for instance.

variable ranging over i's possible posterior probabilistic belief states on $\mathcal{S}$ after conditionalisation on their total information. Note that $\mathcal{R}_i(\cdot) = \Pr(\cdot|\mathcal{E}_i)$ and that, for all possible belief states of expert i, $\mathcal{R}_i = R_i$ is equivalent to $\mathcal{E}_i = E_i$, with E_i being the logically strongest proposition in $\mathcal{S}$ to which R_i assigns probability one. Now circumstances in which the expert holds (weakly) more information than You are ones in which their total evidence E_i implies your evidence E. Given the common prior assumption, these are just the circumstances in which $R_i(E) = 1$. So it is a consequence of the assumption that You and the experts are Bayesian agents that:

Information Deference For all of expert i's possible belief states R_i on $\mathcal{S}$, if $R_i(E) = 1$ (i.e. if expert i has no less information than You), then:

$$P(\cdot|\mathcal{R}_i = R_i) = R_i(\cdot)$$

Proof Let E_i be the support of R_i. Then $P(\cdot|\mathcal{R}_i = R_i) = P(\cdot|\mathcal{E}_i = E_i) = \Pr(\cdot|\mathcal{E}_i = E_i, E)$. Now E is implied by E_i which in turn is equivalent to $\mathcal{E}_i = E_i$ since in the event of the expert learning any proposition they also learn that they learn it. It follows that $\Pr(\cdot|\mathcal{E}_i = E_i, E) = \Pr(\cdot|E_i) = R_i(\cdot)$. $\qquad\qquad\square$

As an aside, note that when You regard your future opinions as a source of expertise because based on more information then Information Deference implies Van Fraassen's (1984) well-known Reflection Principle. For let $P_0 = \Pr(\cdot|E)$ be your current degrees of belief and $\mathcal{P}_t$ be a random variable ranging over your possible posterior probabilistic belief states on $\mathcal{S}$ at future time t obtained by conditionalisation on any additional information that You have acquired by that time. Then, as You will be better informed at t than You are now, Information Deference requires that your conditional degrees of belief should satisfy $P_0(\cdot|\mathcal{P}_t = P_t) = P_t(\cdot)$.

3.1 Revising by deferring

It follows immediately from Information Deference that in situations in which You know that expert i has obtained more information than You, your degrees of belief should equal the expected posterior probabilities of expert i, conditional on their information E_i, i.e. that whenever $P(\mathcal{R}_i(E) = 1) = 1$:[4]

$$P = \mathbb{E}(\mathcal{R}_i) \qquad\qquad\qquad (2)$$

It follows that You expect the expert to believe what You currently do; justifiably so, because what You believe is based on all the evidence that You have at that time. Or to put it the other way round, if You believed the expert to hold evidence regarding the truth of some proposition that makes it more probable than You judge it to be, then You should immediately adjust your belief in that proposition. Evidence of evidence for a proposition is evidence for that proposition.

---

4 See [Joyce (2007), p. 191] for a generalisation of this claim.

Testimony from experts provides precisely this kind of evidence of evidence and hence should induce revision of your opinions. In particular, when a single expert's testimony regarding some proposition X is the only information that You acquire, then You should be disposed to adopt whatever opinion the expert reports on it. For in virtue of the expert's truthfulness, what she reports is her posterior degree of belief in X which, in virtue of the assumption of a common prior, is just your probability for X given the evidence obtained by her.

More formally, let $\mathcal{R}_i$ (as before) be a random variable ranging over the possible posterior probabilistic belief states on $\mathcal{S}$ of expert i after they have conditionalised on their total information and let $\mathcal{Q}$ be a random variable representing your possible probabilistic belief states after hearing expert i's reported opinion on X. Suppose that $P(\mathcal{R}_i(E) = 1) = 1$ and that the expert reports a probability of x for X. Then as a Bayesian agent You must revise your degrees of belief in accordance with:

Deference-based Revision If $P(\mathcal{R}_i(E) = 1) = 1$ then:

$$\mathcal{Q}(X) = \mathcal{R}_i(X)$$

Proof Bayesian conditionalisation requires that $\mathcal{Q}(X) = P(X|\mathcal{R}_i(X))$. By equation 2, $P(X) = \mathbb{E}(\mathcal{R}_i(X))$ and $P(X|\mathcal{R}_i(X)) = \mathbb{E}(\mathcal{R}_i(X)|\mathcal{R}_i(X)) = \mathcal{R}_i(X)$. So $\mathcal{Q}(X) = \mathcal{R}_i(X)$. $\square$

Two remarks. First, Deference-based revision requires that You adopt whatever opinions on X that an expert reports in circumstances in which the expert holds strictly more information. Revision of this kind is consistent with linear averaging if and only if full weight is given to the expert's opinion. For in the envisaged circumstances your revised probabilities, $\mathcal{Q}$, after hearing expert i's report on X, are required to satisfy, for $0 \leq \alpha_i \leq 1$, both:

$$\mathcal{Q}(X) = \mathcal{R}_i(X)$$
$$\mathcal{Q}(X) = \alpha_i \mathcal{R}(X) + (1 - \alpha_i)P(X)$$

and this implies that $\alpha_i = 1$ (or that they always report what you already believe). This is, of course, just what one expects when the expert has strictly greater information than You. So in situations in which full deference to a single expert is appropriate, supra-Bayesian conditionalisation is (trivially) consistent with proposition-dependent linear averaging.

Second, conformity to Deference-based Revision does not determine your entire posterior belief state; just a fragment of it. To propagate the implications of deferring to the expert's reported opinion on X to the rest of one's beliefs, one might reasonably follow Steele's (2012) recommendation to Jeffrey conditionalise on the partition $\{X, \neg X\}$ taking as inputs one's newly revised degrees of belief for its elements. This yields, for all propositions $Y \in \mathcal{S}$:

$$\mathcal{Q}(Y) = P(Y|X) \cdot \mathcal{Q}(X) + P(Y|\neg X) \cdot \mathcal{Q}(\neg X) \tag{3}$$

Revising your beliefs in this fashion is demonstrably the rational thing to do just in case You take the evidential significance for Y of the expert's report on X to be screened out by the truth of X or $\neg X$, i.e. just in case for any reported probability of x for X, $P(Y|X, \mathcal{R}_i(X) = x) = P(Y|X)$ and $P(Y|\neg X, \mathcal{R}_i(X) = x) = P(Y|\neg X)$. Then,

$$\begin{aligned}
\mathcal{Q}(Y) &= P(Y|\mathcal{R}_i(X) = x) \\
&= P(Y|X, \mathcal{R}_i(X) = x) \cdot P(X|\mathcal{R}_i(X) = x) \\
&\quad + P(Y|\neg X, \mathcal{R}_i(X) = x) \cdot P(\neg X|\mathcal{R}_i(X) = x) \\
&= P(Y|X) \cdot \mathcal{Q}(X) + P(Y|\neg X) \cdot \mathcal{Q}(\neg X)
\end{aligned}$$

Such screening out need not always occur, for there may be occasions on which the expert's report on a proposition is informative on more than the truth of just that proposition, such as when the content of a report reveals something about the conditions under which it is obtained. For instance, a report that the music is loud in some location may reveal that the reporter has been there if there is no other way that they could have learnt this. But screening-out is, I speculate, normally the case, for we are usually interested in someone's opinion on some proposition in virtue of the information it provides about just that proposition.

A couple of cautionary points. First, it must be emphasised that revising one's beliefs by application of Deference-based Revision and Jeffrey conditionalisation does not guarantee that one's posterior beliefs will be the same as the expert's, i.e. $\mathcal{Q} = \mathcal{R}_i$. (In the absence of further information, however, one will expect them to be so.) Second, one cannot mechanically apply this type of revision to a sequence of reports by the expert. Suppose, for instance, having deferred to the expert's opinion on X and then Jeffrey conditioned on the partition $\{X, \neg X\}$, the expert reports her opinion on Y. It is clear that You should defer to this opinion as well, since the expert is still strictly better informed than You. But if You attempt once again to revise your other beliefs by Jeffrey conditioning, this time on the partition $\{Y, \neg Y\}$, using your newly acquired opinions on its members as inputs, You will be led to revise your opinion on X (except, of course, when X and Y are probabilistically independent). And this would conflict with your commitment to adopting the expert's opinion on X as your own. What You should do is revise your other beliefs subject to the *dual* constraint on the partition $\{XY, X\neg Y, \neg XY, \neg X\neg Y\}$ implied by your deference to the expert's reported opinions on both X and Y. However, this constraint does not determine a unique redistribution of probability across the relevant partition. So we have not settled the question of how, in general, one should respond to multiple reports by a single expert.

3.2 Different priors

Let us temporarily drop the assumption of common priors and look at cases in which the expert has special skills rather than additional information. Suppose for example that an investor has doubts about the financial viability of a company in which she holds shares. The investor gets access to the company's accounts but, lacking expertise in accounting, has the books examined by an accountant. The accountant looks them over

and declares a probability that the return on the investment will be at least as high as required by the investor. Although the investor may completely trust the accountant's judgement of the evidence provided by the company's accounts, she may nonetheless not be willing simply to adopt the accountant's posterior probabilistic judgement on the returns to an investment, because she suspects that the accountant's prior is less reasonable than her own.

In Bayesian statistics the support that some piece of evidence gives to one hypothesis relative to another is often measured, in a prior-free way, by its *Bayes factor*.[5] We arrive at this factor in the following way. Note that for any probability P, propositions X and Y and evidence proposition E, such that $P(Y|E) > 0$:

$$\frac{P(X|E)}{P(Y|E)} = \frac{P(E|X)}{P(E|Y)} \cdot \frac{P(X)}{P(Y)}$$

So let the associated Bayes factor, $\mathbb{B}(E, X, Y)$, on E, for X relative Y, be defined by:

$$\mathbb{B}(E, X, Y) := \frac{P(E|X)}{P(E|Y)}$$

Then it follows that:

$$\frac{P(X|E)}{P(Y|E)} = \mathbb{B}(E, X, Y) \cdot \frac{P(X)}{P(Y)} \tag{4}$$

Now in situations in which one wishes to defer to an expert's judgement on the significance of some evidence E, but not to their prior opinions, one can use the decomposition given by Eq. 4 to determine one's posterior probabilities from the expert's Bayes factors and one's own prior opinion. To illustrate, as before let your and the expert's prior and posterior probabilities following receipt of E, respectively, be $\Pr$ and $\Pr_i$ and P and R_i. Suppose that the expert reports both $\Pr_i(X)$ and $R_i(X)$. From this their Bayes factor on E for X relative to $\neg X$ can be determined by:

$$\begin{aligned} \mathbb{B}(E, X, \neg X) &= \frac{\Pr_i(E|X)}{\Pr_i(E|\neg X)} \\ &= \frac{\Pr_i(X|E)}{\Pr_i(\neg X|E)} \cdot \frac{\Pr_i(\neg X)}{\Pr_i(X)} \\ &= \frac{R_i(X)}{1 - R_i(X)} \cdot \frac{1 - \Pr_i(X)}{\Pr_i(X)} \end{aligned}$$

Let $\mathbb{B}_X^E$ be the Bayes factors on E for X relative to its negation that is implicitly reported by the expert. Then if You defer to the expert's judgement (only) on the significance of the evidence E for X in the sense of being disposed to adopt whatever Bayes factor that they report, but also to retain your own prior probabilistic judgement on X, then your degrees of belief, $\mathcal{Q}$, after hearing the expert's report, should satisfy:

---

[5] See Jeffreys (1961) for the classic statement of the case for this measure.

Bayes Deference

$$Q(X) = \frac{\mathbb{B}_X^E \cdot P(X)}{\mathbb{B}_X^E \cdot P(X) + P(\neg X)}$$

Rationale Since You are disposed to adopt the expert's Bayes factor on E for X, but not their prior probability for X, on learning their Bayes factor you should set your new conditional probabilities for X and $\neg X$, given E, in accordance with Eq. 4, i.e. by:

$$\frac{Q(X|E)}{Q(\neg X|E)} = \mathbb{B}_X^E \cdot \frac{P(X)}{P(\neg X)}$$

But since You are a Bayesian agent and You know that E, $Q(X|E) = Q(X)$ and $Q(\neg X|E) = Q(\neg X) = 1 - Q(X)$. Hence,

$$\frac{Q(X)}{1 - Q(X)} = \mathbb{B}_X^E \cdot \frac{P(X)}{P(\neg X)}$$

Bayes deference follows by rearrangement.

Two remarks. First, Bayes Deference, unlike Information Deference, is not implied by Bayesianism. Indeed Q is not a Bayesian posterior in the usual sense because the Bayes factors of the expert(s) are not assumed to be propositions in $\mathcal{S}$. Instead Bayes Deference expresses the implications for a Bayesian agent of adopting the relevant Bayes factors of the expert(s) for some proposition but not their prior probability for it. Second, revision of belief by application of Bayes Deference to a single proposition does not determine the impact of the expert's reports on your other degrees of belief. But, with the same caveats as before, we may again combine it with Jeffrey conditionalisation as in Eq. 3, taking as inputs the values $Q(X)$ and $Q(\neg X)$, to propagate the change to all other propositions.

4 Partial deference and Bayes compatibility

4.1 Single expert

As we have seen, Information Deference implies a disposition on the part of a Bayesian agent to adopt whatever opinions on X that an expert reports, when that expert holds strictly more information than her. Situations in which full deference is appropriate are obviously quite special, however. Other situations call for only *partial* deference: for instance when You and an expert hold different information relevant to X or when different experts, although all holding strictly more information than You, hold different information. In such cases, complete deferral to the opinion of the expert or one of the experts would be tantamount to discarding some relevant information.

Let us consider the first of these cases and leave the second to the next subsection. Suppose that You and expert i have, respectively, acquired new information E and E_i such that $\Pr(E, E_i) > 0$ and adopted posterior beliefs $P = \Pr(\cdot|E)$ and $R_i(\cdot) = \Pr(\cdot|E_i)$. In these circumstances You should defer, not to the expert's posterior opinions, but to her posterior *conditional* opinions, given the information E that

You have acquired; not to what she believes but to what she would believe were she to hold all the information You do. For although R_i is not based on more information than You hold, $R_i(\cdot|E)$ is. So the principle of deference to expertise based on greater information implies:

Conditional Deference For all possible probability functions R_i on $\mathcal{S}$:

$$P(\cdot|\mathcal{R}_i = R_i) = R_i(\cdot|E) \tag{5}$$

Proof As before, let $\mathcal{E}_i$ be a random variable ranging over the possible total evidences of expert i. Then $P(\cdot|\mathcal{R}_i = R_i) = P(\cdot|\mathcal{E}_i = E_i) = \Pr(\cdot|\mathcal{E}_i = E_i, E)$. But since E_i is equivalent to $\mathcal{E}_i = E_i$, it follows that $\Pr(\cdot|\mathcal{E}_i = E_i, E) = \Pr(\cdot|E_i, E) = R_i(\cdot|E)$. □

Conditional Deference expresses a general requirement on all Bayesian agents of partial deference to expertise in circumstances of shared priors. Information Deference follows from it, expressing the special case in which the expert acquires no less information than You (when E_i implies E and hence $R_i(\cdot|E) = R(\cdot)$). Note that it also follows from Conditional Deference that your posterior degrees of belief should equal the expected conditional degrees of belief of expert i given that E, i.e. $P = \mathbb{E}(\mathcal{R}_i|E)$. So it tells You to set your opinions to what You expect the expert would believe were they to learn your information. This does not of course suffice to determine how You should revise your beliefs in response to the expert's report on X, since it is her conditional opinion on X given that E that You wish to adopt, not her expressed unconditional opinion on X. So there is now an additional motive to ask whether the kind of sensitivity to the expert's opinion implied by linear averaging can further constrain your opinion in a Bayes-compatible way. That is, we want to know whether You can assign a weight α to the expert's opinion on X, independent of what she reports, such that for any reported opinion x in the range of $\mathcal{R}_i(X)$:

$$P(X|\mathcal{R}_i(X) = x, E) = \alpha x + (1 - \alpha)P(X|E) = R_i(X|E)$$

Some simple examples suffice to reveal the obstacles to a positive response. Suppose the expert reports a probability of zero for X. Now if $R_i(X) = x = 0$ then $R_i(X|E) = 0$ as well. So $(1 - \alpha)P(X|E) = 0$ and hence either $P(X|E) = 0$ or $\alpha = 1$, i.e. either the expert reports what You already believe on the basis of your evidence or You defer completely to their opinion. Similarly, if the expert reports a probability of one for X. For if $R_i(X) = 1$ then $R_i(X|E) = 1$ as well. And so either $P(X|E) = 1$ or $\alpha = 0$, i.e. either the expert reports what You already believe on the basis of your evidence or You attach no weight at all to their opinion. But complete deference to the expert's opinion is inappropriate in these circumstances because You hold information that she does not.[6] So we must conclude that linear averaging is Bayes compatible only if the expert cannot learn the truth or falsity of X, unless You do too. In fact this conclusion extends to cases in which the expert reports an intermediate probability for X. But instead of showing this directly, I want to turn our attention to a case that is formally analogous

---

[6] Indeed complete deference to the expert is inconsistent with the Reflection Principle, which requires P to defer to $P(\cdot|E)$.

to the one just discussed, namely in which two experts both hold more information than You, and show that the corresponding claim holds in this case.

4.2 Multiple experts

Suppose that two experts, again sharing prior Pr with You have, respectively, acquired new information E_1 and E_2 and adopted posterior beliefs R_1 and R_2, but that You have acquired no information (hence $P = \text{Pr}$). In this case, in virtue of the shared prior, Information Deference tells us that:

$$P(\cdot|\mathcal{R}_1 = R_1) = R_1(\cdot) = \text{Pr}(\cdot|E_1) \tag{6}$$

$$P(\cdot|\mathcal{R}_2 = R_2) = R_2(\cdot) = \text{Pr}(\cdot|E_2) \tag{7}$$

And from these two equations it follows that:

$$P(\cdot|\mathcal{R}_1 = R_1, \mathcal{R}_2 = R_2) = \text{Pr}(\cdot|E_1, E_2) \tag{8}$$

Proof By 7, $P(\cdot|\mathcal{R}_1 = R_1, \mathcal{R}_2 = R_2) = \text{Pr}(\cdot|\mathcal{R}_1 = R_1, E_2) = \text{Pr}(\cdot|E_1, E_2)$ by 6. □

Since the experts' reports do not allow You to infer what the propositions E_1 and E_2 are that the experts have learnt, 8 is not enough to determine how You should revise your beliefs in X. Nonetheless, we can now ask whether linear averaging (and specifically LAC) is consistent with Eqs. 6, 7, and 8, i.e. whether there exists non-negative weights α_0, α_1 and α_2 such that $\alpha_0 + \alpha_1 + \alpha_2 = 1$ and such that for all possible profiles of reported probabilistic opinions x_1, x_2 on X:

$$P(X|\mathcal{R}_1(X) = x_1, \mathcal{R}_2(X) = x_2) = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_0 P(X) \tag{9}$$

where, as indicated before, the α_i can depend on X but not on what the experts report about X.

The answer is that no such weights exist unless the experts always make the same report. The reason for this is quite simple: without this restriction, Deference-based Revision together with Eq. 9 requires that the weight You put on each expert's report *does* depend on what they report. To see this consider the following example. Suppose that $R_2(X) = 0$ but $R_1(X) = x_1 > 0$. Then by Information Deference $P(X|\mathcal{R}_2(X) = x_2) = 0$. It follows that:

$$P(X|\mathcal{R}_1(X) = x_1, \mathcal{R}_2(X) = 0) = 0$$

So if You conditionalise on the experts' reports on X then your posterior beliefs are such that $\mathcal{Q}(X) = 0$. But then by linear averaging:

$$\mathcal{Q}(X) = \alpha_1 \mathcal{R}_1(X) + \alpha_2 \mathcal{R}_2(X) + \alpha_3 P(X)$$
$$= \alpha_1 x_1 + 0 + \alpha_3 P(X) = 0$$

and this implies that $\alpha_1 = 0 = \alpha_3$. But by the same argument if $R_1(X) = 0$ but $R_2(X) = x_2 > 0$ then Bayes compatibility of proposition-dependent linear averaging requires that $\alpha_2 = 0 = \alpha_3$ and hence that $\alpha_1 = 1$. So the weights on the experts' reports on X are *not* independent of what they report.

We can put the point slightly differently. If the weights on the experts' reports on X are independent of what they report, then we can infer that they always make the same report. In our example, we assumed that the experts made different reports to infer weight variability. But if weights cannot vary then the experts cannot have made different reports. So it cannot be possible for one expert to learn the truth or falsity of X without the other doing so.

The conclusion holds more generally: so long as You put positive weight on both experts' opinions (which is mandatory, given that they hold more information than You) then the requirement that averaging weights on expert opinion be independent of what they report implies that they always make the same report, even when they report intermediate probabilities for X. For Information Deference implies that You expect, conditional on expert 1 reporting a probability of x for X, that your posterior degree of belief in X, after hearing *both* experts' reports, to be just x, i.e.

$$\mathbb{E}(\mathcal{Q}(X)|\mathcal{R}_1(X) = x) = x \tag{10}$$

(This is proved as Theorem 1 in the "Appendix"). But Eq. 10, together with assumption that your posterior beliefs are a linear average of the experts', implies that your conditional expectation for expert 2's report on X, given that expert 1 reports a probability of x for X, equals x as well (Theorem 2 in the "Appendix"), and vice versa. By Lemma 1 this can be the case only if the random variables corresponding to each of the expert's opinion reports are the same, i.e. if $\mathcal{R}_1 = \mathcal{R}_2$. So, on pain of contradiction, your posterior opinion on X can be both a linear average of the expert's reported opinions and defer appropriately to these reports only if they always report the same opinions.

4.3 Concluding remarks

We have shown that revising your beliefs on some proposition X by taking the linear average of reported expert opinion on it is consistent with Bayesian conditioning only in the trivial case when the experts always make the same report (with probability one). This I take to rule linear averaging out as a rational response to disagreement in expert opinion. It also, therefore, rules out linear averaging as a response to disagreement with your epistemic peers, construed as others who share your priors but hold different information. Or to disagreement between You and an expert when neither of you holds strictly more information than the other. For these cases are formally analogous to that of disagreement amongst experts with different information.

A couple of cautionary notes about the scope of these conclusions. First, nothing has been said in this paper about other forms of averaging or indeed forms of opinion pooling that do not involve averaging. Considerable guidance on this question can be found in Dawid et al. (1995), where a much more extensive set of formal results on

the Bayes compatibility of opinion pooling is proved. The upshot of these results is far from settled, however, and the philosophical status of other pooling rules deserve further exploration.

Second, it should also be emphasised that this study leaves open the question of whether linear averaging is the appropriate response to situations in which you find yourself in disagreement with peers who hold the same information as you and are as a good at judging its significance. In the philosophical literature, the view that one should respond to such disagreements by taking an equal-weighted average of your opinions has been hotly debated. But nothing presented here militates either for or against this view.

Appendix: Proofs

Lemma 1 *For all bounded random variables Y and Z, if $\mathbb{E}(X|Y) = Y$ and $\mathbb{E}(Y|X) = X$, then $X = Y$.*

Proof Note that by the law of iterated expectations, $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$. Now since $(Y|X = x) = x$, it follows by another application of the law of iterated expectations that:

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(\mathbb{E}(XY|X))$$
$$= \mathbb{E}(X \cdot \mathbb{E}(Y|X))$$
$$= \mathbb{E}(X^2)$$

It follows that:

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)]$$
$$= \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$
$$= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

since $\mathbb{E}(X \cdot Y) = \mathbb{E}(X^2)$ and $\mathbb{E}(Y) = \mathbb{E}(X)$. So $\text{cov}(X, Y) = \text{var}(X) = \text{var}(Y)$. Hence, $\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y) = 0$. So $X = Y$ with probability 1. □

Theorem 1 *Suppose that P satisfies Information Deference with respect to $\mathcal{R}_1$ and $\mathcal{R}_2$. Let χ_i be the range of $\mathcal{R}_i(X)$ and let Q over the possible posterior probabilities obtained by conditionalising P on the reports of the experts. Then for any possible value $\bar{x}_1$ of $\mathcal{R}_1(X)$:*

$$\mathbb{E}(\mathcal{Q}(X)|\mathcal{R}_1(X) = \bar{x}_1) = \bar{x}_1$$

Proof By Information Deference, $P(X|\mathcal{R}_1(X) = \bar{x}_1) = \bar{x}_1$; and

$$P(X|\mathcal{R}_1(X) = \bar{x}_1) = \sum_{x_2 \in \chi_2} P(X, \mathcal{R}_2(X) = x_2|\mathcal{R}_1(X) = \bar{x}_1))$$

$$= \sum_{x_2 \in \chi_2} P(X|\mathcal{R}_2(X) = x_2, \mathcal{R}_1(X) = \bar{x}_1)) \cdot P(\mathcal{R}_2(X)$$

$$= x_2|\mathcal{R}_1(X) = \bar{x}_1)$$

$$= \mathbb{E}[\mathcal{Q}(X)|\mathcal{R}_1(X) = \bar{x}_1]$$

Hence, $\mathbb{E}(\mathcal{Q}(X)|\mathcal{R}_1(X) = \bar{x}_1) = \bar{x}_1$. $\qquad\qquad\qquad\qquad\qquad\qquad \square$

Theorem 2 *Suppose that prior P satisfies Information Deference with respect to $\mathcal{R}_1$ and $\mathcal{R}_2$ and that $\mathcal{Q}(X) = \alpha_1\mathcal{R}_1(X) + \alpha_2\mathcal{R}_2(X) + a_0 P(X)$, where $a_0 = 1 - (\alpha_1 + \alpha_2)$, $0 < \alpha_1, \alpha_2$ and $0 \le \alpha_0$. Then*

$$\mathbb{E}[\mathcal{R}_1(X)|\mathcal{R}_2(X) = \bar{x}_2] = \bar{x}_2$$
$$\mathbb{E}[\mathcal{R}_2(X)|\mathcal{R}_1(X) = \bar{x}_1] = \bar{x}_1$$

Proof By Theorem 1, $\mathbb{E}(\mathcal{Q}(X)|\mathcal{R}_1(X) = \bar{x}_1) = \bar{x}_1$. Hence,

$$\mathbb{E}[\alpha_1\mathcal{R}_1(X) + \alpha_2\mathcal{R}_2(X) + a_3 P(X)|\mathcal{R}_1(X) = \bar{x}_1] = \bar{x}_1$$

But by Information Deference, $P(X|\mathcal{R}_1(X) = \bar{x}_1) = \bar{x}_1$. So by the Linearity property of expectations:

$$\alpha_1\mathbb{E}[\mathcal{R}_1(X)|\mathcal{R}_1(X) = \bar{x}_1] + \alpha_2\mathbb{E}[\mathcal{R}_2(X)|\mathcal{R}_1(X) = \bar{x}_1] + \alpha_3 P(X|\mathcal{R}_1(X) = \bar{x}_1)$$
$$= \alpha\bar{x}_1 + \alpha_2\mathbb{E}(\mathcal{R}_2(X)|\mathcal{R}_1(X) = \bar{x}_1) + \alpha_3\bar{x}_1$$
$$= \bar{x}_1$$

But this can be the case iff $\mathbb{E}(\mathcal{R}_2(X)|\mathcal{R}_1(X) = \bar{x}_1) = \bar{x}_1$. By the same argument $\mathbb{E}(\mathcal{R}_1(X)|\mathcal{R}_2(X) = \bar{x}_2) = \bar{x}_2$. $\qquad\qquad \square$

Corollary 1 *Under the assumptions of Theorem 2 and Lemma 1, $\mathcal{R}_1 = \mathcal{R}_2$*

Proof Follows from Theorem 2 and Lemma 1. $\qquad\qquad\qquad\qquad\qquad \square$

References

Bradley, R. (2006). Taking advantage of difference of opinion. *Episteme*, *3*, 141–155.

Bradley, R. (2007). Reaching a consensus. *Social Choice and Welfare*, *29*, 609–632.

Bonnay, D. & Cozic, M. (2017). Weighted averaging, Jeffrey conditioning and invariance, *Theory and Decision* (submitted)

Bradley, R. (2007). The kinematics of belief and desire. *Synthese*, *156*, 513–535.

Dawid, A. P., DeGroot, M. H., & Mortera, J. (1995). Coherent combination of experts' opinions. *Test*, *4*, 263–313.

Dietrich, F. & C. List (2017) Probabilistic opinion pooling generalized - Part one: general agendas. *Social Choice and Welfare 48*, 747–786.

Dietrich, F., & List, C. (2016). Probabilistic opinion pooling. In C. Hitchcock & A. Hajek (Eds.), *Oxford Handbook of Probability and Philosophy*. Oxford: Oxford University Press.

Elgar, A. (2007). Reflection and disagreement. *Noûs*, *41*(3), 478–502.

French, S. (1981). Consensus of opinion. *European Journal of Operations Research*, *7*, 332–340.

Genest, C., & Schervish, M. J. (1985) Modeling Expert Judgments for Bayesian Updating. *Annals of Statistics*, *13*(3), 1198–1212.

Genest, C., & Zidek, J. V. (1986). Combining probability distributions: A critique and annotated bibliography. *Statistical Science*, *1*, 113–135.

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford: Oxford University Press.

Joyce, J. (2007). Epistemic deference: The case of chance. *Proceedings of the Aristotelian Society*, *107*, 1–20.

Lehrer, K. (1976). When rational disagreement is impossible. *Noûs*, *10*, 327–332.

Lehrer, K., & Wagner, C. (1981). *Rational consensus in science and society*. Dordrecht: Reidel.

Morris, P. A. (1974). Decision analysis expert use. *Management Science*, *20*, 1233–1241.

Romeijn, J. W., & Roy, O. (2017). All Agreed: Aumann meets DeGroot, *Theory and Decision (submitted)*

Steele, K. (2012). Testimony as evidence. *Journal of Philosophical Logic*, *41*, 983–999.

Van Fraassen, B. (1984). Belief and the will. *The Journal of Philosophy*, *81*, 235–256.

Wagner, C. (1982). Allocation, lehrer models, and the consensus of probabilities. *Theory and Decision*, *14*, 207–220.

Wagner, C. (1985). On the formal properties of weighted averaging as a method of aggregation. *Synthese*, *62*, 97–108.