Forthcoming in *Mind*. Penultimate version.

Scoring rules and epistemic compromise

Sarah Moss
ssmoss@umich.edu

Formal models of epistemic compromise have several fundamental applications. Disagreeing agents may construct a compromise of their opinions to guide their collective action, to give a collective opinion to a third party, or to determine how they should update their individual credences. Recent literature on disagreement has focused on certain questions about epistemic compromise: when you find yourself disagreeing with an epistemic peer, under what circumstances, and to what degree, should you change your credence in the disputed proposition? Elga (2006) and Christensen (2007) say you should compromise often and deeply; Kelly (2010) disagrees. But these authors leave open another question: what constitutes a perfect compromise of opinion?

In the disagreement literature, it is sometimes assumed that if we assign different credences c_a and c_b to a proposition p, we reach a perfect compromise by *splitting the difference* in our credences. In other words: to adopt a perfect compromise of our opinions is to assign $(c_a+c_b)/2$ credence to p. For instance, Kelly (2010) says that when peers assign 0.2 and 0.8 to a proposition, to adopt a compromise is to split the difference with one's peer and believe the hypothesis to degree 0.5.[2] But why does 0.5 constitute a perfect compromise? Of course, 0.5 is the arithmetic mean of 0.2 and 0.8. But why must a compromise of agents' opinions always be the arithmetic mean of their prior credences? In other cases of compromise, we do not simply take it for granted that the outcome that constitutes a perfect compromise is determined by the

1. Thanks to Allan Gibbard, Bob Stalnaker, Eric Swanson, Roger White, the 2008 MITing of the Minds Conference, the 2008 Formal Epistemology Workshop, and the 2009 Formal Epistemology Festival for helpful discussion of this paper.

2. See Christensen 2007, Joyce 2007, and Weatherson 2011 for further recent references to compromising by splitting the difference.

arithmetic mean of quantities that reflect what individual agents most prefer. Suppose we are running partners, and I want to run one mile, while you want to run seven. Running four miles may not be a perfect compromise, especially if I strongly prefer running one mile rather than four, while you only slightly prefer running farther.

The moral of this paper is that the same sort of situation may arise in purely epistemic cases of compromise, cases in which each agent prefers having certain credences over others, where this preference is grounded in purely epistemic concerns. Suppose I strongly prefer assigning 0.1 credence to a disputed proposition, while you weakly prefer credences around 0.7. In this kind of case, we may reach a perfect compromise by converging on some shared credence *lower than* 0.4. Splitting the difference may not constitute a perfect compromise when agents who have different credences also have different epistemic values. To make this moral precise, we must say how an agent may value certain credences over others, in a purely epistemic sense. I take an agent's *scoring rule* to measure how much she epistemically values various alternative credences she might assign a given proposition.[3] Using scoring rules, we can develop a natural alternative to the strategy of compromising by splitting the difference: agents may compromise by coordinating on the credences that they collectively most prefer, given their epistemic values.

I have two main aims in this paper: to develop this alternative strategy, and to argue that this strategy governs how agents should compromise. In section 1, I define the notion of a scoring rule and introduce relevant properties of these epistemic value functions. In section 2, I develop the alternative strategy for compromising that I defend. In section 3, I compare my alternative strategy with the traditional strategy of splitting the difference. I characterize the situations in which the two strategies coincide, and those in which they differ. In section 4, I argue that where the strategies do yield different recommendations, compromising by maximizing epistemic value is a reasonable strategy.

1 Scoring rules

In assigning a particular credence to a proposition, you are estimating its truth value. In a purely epistemic sense, closer estimates are better. Having 0.8 credence in a truth is better than having 0.7 credence, and having 0.9 credence is better still.

How much better? Different agents may value closer estimates differently. For instance, suppose you care a lot about your credences as they approach certainty. In

3. For recent literature using scoring rules as measures of epistemic value, see Joyce 1998, Percival 2002, Gibbard 2007, and Joyce 2009.

this case, you may value having 0.9 credence in a truth much more than having 0.8 credence, without valuing 0.6 much more than 0.5. Or suppose you do not particularly care about having credences that approach certainty, as long as your beliefs are on the right track. In that case, you may equally value 0.9 and 0.8 credence in a truth, or equally value 0.2 and 0.3 credence in a falsehood.

Facts like these are traditionally modelled by your *scoring rule*, a record of how much you value various estimates of truth values.[4] The simplest sort of scoring rule models how much you value having a certain credence in a true or false proposition. On this simple model, a scoring rule f is a pair of functions f_1, f_0 from $[0, 1]$ to $\mathbb{R}$. Intuitively, the first function $f_1(x)$ measures how much you value having credence x in a proposition that turns out to be true. For instance, if you value having 0.9 credence in a true proposition much more than having 0.8 credence, without valuing 0.6 much more than 0.5, then the first function f_1 of your scoring rule will reflect this preference:

$$\big(f_1(0.9) - f_1(0.8)\big) > \big(f_1(0.6) - f_1(0.5)\big)$$

The second function $f_0(x)$ measures the value of having credence x in a proposition that turns out to be false. For instance, if it makes no difference to you whether you have 0.2 or 0.3 credence in a falsehood, then the second function f_0 of your scoring rule will have equal values on these credences:

$$f_0(0.2) = f_0(0.3)$$

Given this simple model of your epistemic values, we can formally model the statement that closer estimates of truth value are better—that is, that they are more valuable to you. Closer estimates of the truth value of propositions that turn out to be true are more valuable to you just in case your scoring rule $f_1(x)$ is strictly increasing. Closer estimates of the truth value of falsehoods are more valuable just in case $f_0(x)$ is strictly decreasing.

Scoring rules can measure more than just the value of having a certain credence in a single proposition. In practice, a more useful kind of scoring rule measures how much you value having a certain *credence distribution* over a partition of propositions.[5]

4. Scoring rules were independently developed by Brier (1950) and Good (1952) to measure the accuracy of probabilistic weather forecasts. Savage (1971) uses scoring rules to assess forecasts of random variables, and treats assignments of credence to particular events as a special case of such forecasts. See Percival 2002 for references to recent literature using scoring rules as measures of epistemic value.

5. See Oddie 1997, Joyce 1998, Greaves & Wallace 2006, Fallis 2007, and Joyce 2009 for examples of scoring rules that score entire credence distributions.

In what follows, I take a scoring rule to assign a value to pairs of credence distributions and truth value assignments. Let P be a finite partition consisting of propositions $p_1, p_2, \ldots, p_N$. Suppose these are the propositions relevant to some dispute between you and another agent. Intuitively speaking, your scoring rule measures the value of your having the credences that you do, if it turns out that the actual world is in some particular element of the partition P. Formally, your scoring rule is a function from $\mathcal{C}_P \times \mathcal{U}$ to $\mathbb{R}$, where $\mathcal{C}_P$ contains credence distributions over P and $\mathcal{U}$ contains unit vectors of length N. The unit vector u^i says that p_i is the true member of the partition P. And the score $S(X, u)$ measures how valuable it is to have credence distribution X over P if it turns out that the truth values of the P propositions are just as u says.

Taken together, your scoring rule and your actual credence distribution determine the *expected epistemic value* of your having a particular credence distribution. Suppose you actually have credence distribution A over a partition P of propositions. Then your expected epistemic value of having some other credence distribution X over P is just a weighted sum of the values you assign to having credence distribution X when particular members of P are true:

$$\textsc{ev}(X, A, S) =_{df} \sum_{i=1}^{N} A(p_i) \cdot S(X, u^i)$$

For instance, in the case where P contains just a proposition p and its negation, your expected value of having credence distribution X is the weighted sum of two values: the value of your having credence distribution X if p is true, weighted by how likely you think it is that p is in fact true, and the value of your having credence distribution X if p is false, weighted by how likely you think it is that p is in fact false.

Even if you and another agent have the same scoring rule, you may well assign your own credence distribution a greater expected epistemic value than you assign her credence distribution. You may even assign your own credences a greater expected value than you assign *any* alternative credences. If this holds no matter what credences you have, then your scoring rule is *credence-eliciting*. In other words, your scoring rule is credence-eliciting when no matter what credences you have, having those credences maximizes the expected value of your credence distribution. In other words, you always do the best you can by your own lights by having the credences that you do.

As long as your scoring rule is credence-eliciting, you prefer your credences to any others. That is, as long as your rule is credence-eliciting, your scoring rule will never dictate that you should change your credence distribution to another with greater expected epistemic value. It is widely accepted that this means that if you are rational, you must have a credence-eliciting scoring rule.[6] Otherwise, your scoring rule could motivate you to raise or lower your credences *ex nihilo,* in the absence of any new evidence whatsoever. Hence the epistemic utility function of a rational agent is analogous to the practical utility function of an agent who has everything she most desires. Her preferences will be reflected in her judgements about the relative well-being of other agents and in her preferences regarding future and counterfactual situations that are less desirable than her own.

2 Compromise beyond splitting the difference

The epistemic value of having accurate credences is often taken to be a degreed analog of the more familiar epistemic value of having true beliefs.[7] Following in this tradition, I take it that agents who value having accurate rather than inaccurate credences are just like agents who simply value having true rather than false beliefs. To take a realistic example inspired by Cohen 2000: suppose that you are deciding where to go to graduate school. For years, you have been studying vagueness, and you are currently on the fence about which theory to endorse. Over the course of your graduate studies, you want to develop more accurate opinions about which theory is correct. Of course, you are aware that your choice of department is likely to affect what evidence and arguments you get in favor of each theory. Suppose you have an informed and reasonable belief that if you go to Cambridge, your credence in epistemicism will decrease from 0.55 to 0.3, whereas if you go to Oxford, your credence in epistemicism will increase to 0.98. Given that you want to have more accurate opinions about vagueness, which graduate program should you choose? Does your current slight affinity for epistemicism mitigate the risk of more decisively endorsing an incorrect theory? The answer depends on—and hence reveals—your epistemic values. Oxford is the correct choice if and only if 0.98 has a higher expected value than 0.3, given your epistemic values and your current credence of 0.55. In so far as you value truth for its own sake, you will not willingly trade in your current credences for others. But in deciding what sort of evidence to collect, you could make a number of other principled decisions, each revealing various epistemic priorities.[8]

Cases of compromise between agents provide another practical application for

6. See Oddie 1997, Joyce 1998, and Gibbard 2007 for further discussion.

7. See Gibbard 2007 and Swanson 2007 for discussion of the relationship between 'valuing the truth' and valuing accurate credences.

8. Others have used different realistic examples to illustrate similar points. For instance, some have argued that scientists evince their epistemic values in deciding which experiments to perform. See Fallis 2007 for an overview of the relevant literature.

the notion of a scoring rule. Scoring rules are practically relevant when agents have to assess the epistemic value of alternative credences. Cases of compromise call for exactly this. Compromising agents must pick credences to coordinate on. In doing so, they have to assess the epistemic value of alternative credences, if they are to determine which shared credences they epistemically ought to pick. This points the way to an alternative strategy for constructing a compromise of agents' opinions: *maximizing expected epistemic value.* Suppose we assign different credences to some proposition p, but we must construct a credence distribution that constitutes the compromise of our opinions. The following seems like a natural strategy: choose our consensus credence in p to maximize the average of the expected values that we each assign to alternative credence distributions over the algebra with atomic propositions p and $\neg p$.

For instance, suppose you have credence distribution A and I have credence distribution B, and suppose that you score credences with S, and I score them with T. Then we should compromise by choosing the credence distribution X that maximizes:

$$\textsc{aev}(X, A, S, B, T) =_{df} \frac{\textsc{ev}(X, A, S) + \textsc{ev}(X, B, T)}{2}$$

This is our alternative compromise strategy: rather than just splitting the difference between our credences, we may compromise by maximizing the average of the expected values we give to our consensus credence distribution. Suppose that you and a fellow scientist disagree about which theory is correct but must collaborate in order to perform more experiments. If you must construct a compromise of your current credences in the midst of such deliberations, then it is natural to think that your epistemic values should affect whatever credence distribution is meant to represent your collective opinion.

Our alternative compromise strategy yields genuinely alternative recommendations. Disagreeing agents who compromise by maximizing epistemic value rarely end up simply splitting the difference in their credences. Suppose you meet two agents with exactly the same credences, but with different scoring rules. If you were to compromise with each agent individually, you may end up constructing different consensus opinions with each of them, even if your scoring rule is credence-eliciting. In what follows, I characterize the cases in which compromising by maximizing expected epistemic value comes apart from the traditional strategy of compromising by splitting the difference.

3 Comparing compromise strategies

In some cases, our alternative strategy yields a traditional recommendation. In particular, when agents *share a credence-eliciting scoring rule,* they maximize the average expected value of their compromise credences by simply splitting the difference between their current credence distributions. For example, suppose we value credences using the *Brier score.*[9] This scoring rule is given by the following function:[10]

$$f(X,u) =_{df} \frac{-\sum_{i=1}^{N} \big(X(p_i) - u_i\big)^2}{N}$$

Suppose you have 0.7 credence in p and I have 0.1 credence in p. The Brier score is a credence-eliciting scoring rule. Hence to maximize the average of our expected values for prospective shared credence distributions, we should compromise by giving 0.4 credence to p.[11]

Sometimes agents pursue a perfect compromise, one that is as fair and even as possible. But agents may also pursue an imperfect compromise, one that favors some agents more than others. For example, when an expert and an amateur disagree, they may elect to compromise in a way that favors the expert's original opinion. One traditional strategy for generating an imperfect compromise is to take a weighted arithmetic mean of agents' original opinions.[12] For example, an amateur may give the prior opinion of an expert four times as much weight as his own. Then if the expert initially has 0.1 credence in p and the amateur has 0.7 credence, they may compromise at $0.8(0.1) + 0.2(0.7) = 0.22$. Our alternative compromise strategy can also be extended to cases of imperfect compromise. In a case of imperfect compromise, agents may maximize a weighted arithmetic mean of the expected values they give to their consensus credence distribution. Furthermore, the above equivalence result extends to cases of imperfect compromise. In the general case: when agents share a credence-eliciting rule, they maximize the weighted average of the expected values they give to

9. This scoring rule was first proposed by Brier 1950, and it is still used by meteorologists to measure the accuracy of probabilistic weather forecasts. See section 12 of Joyce (2009), 'Homage to the Brier Score', for an overview of reasons why the rule is well-suited for assessing the accuracy of credences.

10. I use superscripts to distinguish unit vectors and subscripts to distinguish coordinate values. For example, u_j is the truth value that the unit vector u assigns to the proposition p_j. Strictly speaking, traditional scoring rules measure the *inaccuracy* of a credence, and hence the *disvalue* of having that credence. For simplicity, I follow Gibbard (2007) in using versions of scoring rules that measure positive epistemic value.

11. See Example 1 of Appendix for details.

12. For instance, see the discussion of epistemic deference in Joyce 2007.

their consensus credence distribution when the consensus is the *same weighted average* of their prior credences. It follows that when agents share a credence-eliciting scoring rule, they maximize the exact average of their expected epistemic values by exactly splitting the difference between their prior credences.

Hence in some cases, including cases of imperfect compromise, splitting the difference and maximizing expected value coincide. But it is not hard to see that in principle, these strategies could yield different results. Sometimes agents who disagree about what is *practically* valuable do not maximize their expected utility by splitting the difference between their most preferred outcomes, as in the case where I want to run one mile and you want to run seven. Similarly, agents who disagree about what is epistemically valuable do not always maximize expected epistemic value by splitting the difference in their credences. For instance, agents with different scoring rules may not maximize expected epistemic value by splitting the difference. Suppose you value credences using the Brier score, and I value them using the following credence-eliciting rule:

$$g(X,u) =_{df} \frac{\sum_{i=1}^{N} \ln\left(|1 - u_i - X(p_i)|\right)}{N}$$

Suppose you have 0.7 credence in p and I have 0.1 credence in p. Even though our scoring rules are each credence-eliciting, our perfect compromise is asymmetric: in constructing a compromise, we maximize our expected epistemic value by choosing a consensus credence of approximately 0.27, not by splitting the difference.[13]

Even when agents share a scoring rule, they may not maximize expected epistemic value by splitting the difference in their credences. For example, suppose we both value credences using the following 'Brier cubed' score:

$$h(X,u) =_{df} \frac{-\sum_{i=1}^{N} |X(p_i) - u_i|^3}{N}$$

Suppose you have 0.7 credence in p and I have 0.1 credence in p. Then our perfect compromise is again asymmetric: in constructing a compromise, we maximize our expected epistemic value by choosing a consensus credence of approximately 0.45.[14]

This value has an interesting property: it is exactly what an agent using the Brier cubed score would assign to p to maximize the expected value of her new credence

13. See Example 2 of Appendix for details.

14. See Example 3 of Appendix for details.

distribution, if she started with precisely 0.4 credence in p.[15] This is no coincidence: whenever agents share a scoring rule, they maximize the average of their expected values for prospective credence distributions by picking the distribution that has maximal expected value for an agent with their same scoring rule and the average of their credences. Furthermore, this claim extends to a result about weighted averages of agents' expected epistemic values. That is, agents who share a scoring rule can maximize a given weighted arithmetic mean of their expected values for consensus credence distributions by following a straightforward rule; namely, they should choose the credence distribution that a hypothetical agent would most prefer, if she shared their scoring rule, and if her credence in p were just that same weighted arithmetic mean of their actual credences.

Once we see that this general result is what underwrites the others, the proof of the result is straightforward. Suppose we have k disagreeing agents with shared scoring rule S and credence distributions $A_1, A_2, \dots, A_k$ over a finite partition consisting of propositions $p_1, p_2, \dots, p_N$. Suppose these k credence distributions should be counted according to weights $\delta_1, \delta_2, \dots, \delta_k$, respectively. Then the weighted average of the agents' epistemic values for an alternative credence distribution X is as follows, where B is the desired weighted average of the A_i credence distributions:

$$\begin{aligned}\sum_{i=1}^{k} \delta_i \cdot \mathrm{EV}(X, A_i, S) &= \sum_{i=1}^{k} \delta_i \Big(\sum_{j=1}^{N} A_i(p_j) \cdot S(X, u^j) \Big) \\ &= \sum_{j=1}^{N} \Big(\sum_{i=1}^{k} \delta_i \cdot A_i(p_j) \Big) \cdot S(X, u^j) \\ &= \sum_{i=1}^{k} \mathrm{EV}(X, B, S)\end{aligned}$$

This result applies in cases where exactly one proposition is under dispute, but also in any case where some finite number of agents assign different credences to propositions in some finite partition. Furthermore, the result incorporates several results discussed above. In the special case where compromising agents share a *credence-eliciting* rule, any hypothetical agent with that rule would most prefer her own credences. So agents sharing a credence-eliciting rule should compromise at the weighted average of their actual credence distributions. If the compromising agents pursue a perfect compromise, they maximize their expected epistemic value by splitting the difference in their credences. But in a number of cases, our alternative compromise strategy

15. See Example 4 of Appendix for details.

comes apart from splitting the difference. It just remains to be argued that in such cases, our alternative compromise strategy is a reasonable one.

4 Norms governing compromise

There are several properties we might wish for in a procedure for aggregating credences. It is notoriously impossible to find a procedure that is satisfactory in all respects. For instance, splitting the difference in agents' credences can fail to preserve unanimous judgements of probabilistic independence. The same goes for the compromise procedure I defend. But this is not yet reason to reject either of these procedures, since any procedure with this virtue will necessarily lack others. In what follows, I will compare compromise strategies on more holistic grounds.[16]

In order to understand how compromising agents should be influenced by their epistemic value functions, we must first understand how a single agent should be influenced by her own epistemic values. In the literature on scoring rules, several theorists have addressed the latter question. Our aim will be to generalize their suggestions to norms governing multiple agents at once.

It is generally accepted that on pain of irrationality, a single agent must aim to maximize the expected epistemic value of her credences.[17] For example, Percival compares epistemic value with practical utility:

> Cognitive decision theory is a cognitive analog of practical decision theory ... Bayesian decision theory holds that a rational agent ... maximise[s] expected utility. Similarly, Bayesian cognitive decision theory holds that a rational cogniser ... maximise[s] expected cognitive utility. (Percival 2002, p. 126)

In the same vein, Oddie says that scoring rules measure 'a (pure) cognitive value which it is the aim of a rational agent, qua pure inquirer, to maximize' (Oddie 1997, p. 535).

How can we extend this condition to a norm that applies to compromising agents? It is useful to first consider the following single agent case: suppose an evil scientist tells you he is going to perform an operation to change your credence in a certain

16. See French 1985 for a discussion of impossibility results, and Genest & Zidek 1986 for a canonical overview and bibliography of the consensus literature. See Shogenji 2007 and Jehle & Fitelson 2009 for discussion of how results about judgement aggregation present challenges for the strategy of splitting the difference.

17. Or at least she must maximize expected epistemic value, given that her epistemic values are themselves rationally permissible—for instance, that they are credence-eliciting. See Percival 2002 for further discussion.

proposition p, which you currently believe to degree 0.7. The scientist gives you a choice: after the operation, you may have either 0.6 or 0.8 credence in p. On pain of irrationality, you should choose whichever credence distribution has the greater expected epistemic value. In general, if you are forced to adjust your credence in p, so that your new credences satisfy certain constraints, the epistemically best credences are those with the greatest expected epistemic value *for you.* Cases of compromise are relevantly similar. In order to adopt a compromise of their prior opinions, agents are forced to adjust their credences in p so that they satisfy certain constraints. Only in cases of compromise, these constraints are defined extrinsically rather than intrinsically; namely, their adjusted credences must be equal to each other. In this situation, agents should choose the alternative credences with the greatest possible expected epistemic value *for them.*

In saying precisely how agents should construct a compromise of their opinions, it is useful to see that we face an analogous question when we extend norms governing expected practical utility to cases of practical compromise. Suppose we are deciding where to go for dinner. Based on my credences and practical utilities, I slightly prefer Chinese. Based on your credences and practical utilities, you strongly prefer Indian. Our natural response to the situation is that your stronger preference matters more: we should maximize the average of our expected utilities by choosing Indian. Roughly the same principle governs many democratic voting systems. Every voter decides which outcome he thinks is most likely to make him the most satisfied. His vote reflects his expected practical utilities, and the election outcome reflects the average of many agents' expected practical utilities.

Of course, maximizing average expected utility may not be an ideal decision procedure. But in many situations, maximizing average expected utility is an intuitively reasonable method of deciding what to do when we have different practical utility functions and want to choose a maximally fair action. Epistemic value is a direct cognitive analog of practical utility, and so it is intuitively reasonable to use the same method when deciding what to believe, when we have different epistemic utility functions and want to choose a maximally fair credence. In other words, it is reasonable to aggregate epistemic preferences in the way we generally aggregate practical preferences. It is not as if we must weigh practical and epistemic preferences in deciding what credence to assign: in epistemology contexts, epistemic values—encoded in scoring rules—are the only values at issue.[18]

---

18. Since I do not have space to compare various epistemic analogs of practical compromise strategies, I have focused on one natural representative of the class of such strategies here. But it is not difficult to see that other reasonable procedures for aggregating practical preferences have epistemic analogs, some of which may be further reasonable strategies for epistemic compromise.

Recognizing the value of alternative compromise strategies does not settle how disagreeing peers should update. Suppose that we find out that we are peers and that we disagree. Let us grant that in light of this information, we should adjust our beliefs. Suppose that our collective opinion is represented by whatever credence distribution has greatest epistemic value for us. It is still a further question whether what you should come to believe—or what I should come to believe—is what we collectively believe. For those who oppose radical compromise among disagreeing agents, it is important to note that cases of compromise are not limited to cases in which disagreeing agents trade in their prior credences for matching ones. For instance, an agent may also use a compromise strategy when updating his credence distribution in light of probabilistic evidence, in the sense of Jeffrey 1968. Suppose a certain screeching bird looks 0.8 likely to be a bald eagle, but sounds only 0.6 likely to be a bald eagle. If you had only seen the bird, it would have been clear how to respond: give 0.8 credence to the proposition that it is a bald eagle. But what credence should you assign in the face of conflicting evidence? In cases where your visual and aural evidence conflict, your scoring rule may determine how you should update: suppose there are two agents with exactly your credences, except that one agent updates only on your visual evidence, and the other only on your aural evidence, and then update by accepting a compromise of their opinions as your own.

In addition, even if disagreeing agents may rationally retain their individual credences in a proposition, such agents may still need to act based on a collective opinion. Disagreeing gamblers may need to decide how much of their shared income to bet on a certain horse. Disagreeing weather forecasters may need to report a single opinion to their employers. Suppose an evil scientist tells us he is going to perform an operation to force us to have the same credence in a certain proposition, and that we can choose only what our new credence will be. Recall that an individual should ask an evil scientist for whatever alternative credence has the greatest expected epistemic value. Similarly, we should ask for whatever alternative credence maximizes the average of our expected epistemic values. A final benefit of the compromise strategy I have defended is that agents who compromise by maximizing the average of their expected epistemic values will never both prefer an alternative shared credence. Other compromise strategies—including most strategies that make no reference to epistemic value—could in principle lead agents to compromise on one credence when there is an alternative that both agents would independently prefer. And that kind of prescription would sit uncomfortably with the norm that every agent should independently aim to maximize her expected epistemic value. If epistemic values should influence an individual in isolation, they should continue to influence her when she compromises with others. The scoring rules strategy for compromising is a natural way of extending accepted norms governing single agents to norms governing compromise.

Appendix

Example 1. It may be instructive to give an instance of the proof in section 3, demonstrating the result that if we both assess credences using the Brier score, and you have 0.7 credence in p and I have 0.1 credence in p, then the alternative compromise strategy recommends that we compromise by giving 0.4 credence to p.[19]

Let B be the Brier score. Let X range over credence distributions defined over the partition consisting of p and its negation, and let A_1 be your credence distribution and let A_2 be my credence distribution over this partition. Let the unit vector u^1 say that p is true, and let u^2 say that its negation is true. The alternative compromise strategy recommends that we adopt the credence distribution X that maximizes the following:

$$\begin{aligned}
\textsc{aev}(X, A_1, B, A_2, B) &= 0.5\big(\textsc{ev}(X, A_1, B) + \textsc{ev}(X, A_2, B)\big) \\
&= 0.5\big(0.7B(X, u^1) + 0.3B(X, u^2) + 0.1B(X, u^1) + 0.9B(X, u^2)\big) \\
&= 0.5\big(0.8B(X, u^1) + 1.2B(X, u^2)\big) \\
&= 0.4B(X, u^1) + 0.6B(X, u^2)
\end{aligned}$$

Since the Brier score is credence-eliciting, this value is maximized when $X(p) = 0.4$.

Example 2. Running the following *Mathematica* notebook verifies that agents with different credence-eliciting scoring rules may not maximize the average of their expected epistemic values in an alternative credence distribution by splitting the difference in their credences.

```
b[x_, v_] :=  - (Abs[v - x])^2
l[x_, v_] :=  Log[Abs[1 - v - x]]
ev[x_, m_, f_] := (m*f[x, 1]) + (1 - m)*f[x, 0]
aev[x_, m1_, s1_, m2_, s2_] := 0.5 (ev[x, m1, s1] + ev[x, m2, s2])
Maximize[{aev[y, 0.7, b, 0.1, l], 0 <= y <= 1}, y]
```

For an agent with the Brier score and 0.7 credence in p, and an agent with the logarithmic scoring rule and 0.1 credence in p, having approximately 0.269505 credence in p maximizes expected epistemic value:

`{-0.202268, {y -> 0.269505}}`.

Example 3. Running the previous notebook with the following additions verifies that agents with non-credence-eliciting scoring rules may not maximize the average of

19. I am grateful to the editors for suggesting that I include the present illustration.

their expected epistemic values in an alternative credence distribution by splitting the difference in their credences.

```
c[x_, v_] :=   - (Abs[v - x])^3
Maximize[{aev[y, 0.7, c, 0.1, c], 0 <= y <= 1}, y]
```

For agents who share the Brier cubed score and have 0.7 credence and 0.1 credence in p, having approximately 0.45 credence in p maximizes expected epistemic value:

```
{-0.121225, {y -> 0.44949}}.
```

Example 4. Running the previous notebook with the following addition verifies that the credence distribution preferred by agents sharing the Brier cubed score is that which a hypothetical agent would prefer, if she shared the same scoring rule and had the arithmetic mean of their credences:

```
Maximize[{ev[y, 0.4, c], 0 <= y <= 1}, y]
```

For an agent with the Brier cubed score and 0.4 credence in p, having approximately 0.44949 credence in p maximizes expected epistemic value:

```
{-0.121225, {y -> 0.44949}}.
```

References

Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review*, vol. 78 (1): 1–3.

Christensen, David. 2007. "Epistemology of Disagreement: the Good News." *Philosophical Review*, vol. 116 (2): 187–217.

Cohen, Gerald A. 2000. *If You're an Egalitarian, How Come You're So Rich?* Harvard University Press, Cambridge.

Elga, Adam. 2006. "Reflection and Disagreement." *Noûs*, vol. 41 (3): 478–502.

Fallis, Don. 2007. "Attitudes Toward Epistemic Risk and the Value of Experiments." *Studia Logica*, vol. 86 (2): 215–246.

French, S. 1985. "Group Consensus Probability Distributions: A Critical Survey." In *Bayesian statistics 2*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith, editors, 183–197. North-Holland, Amsterdam.

Gendler, Tamar Szabó & John Hawthorne, editors. 2007. *Oxford Studies in Epistemology*, vol. 2. Oxford University Press, Oxford.

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

Gibbard, Allan. 2007. "Rational Credence and the Value of Truth." In Gendler & Hawthorne (2007).

Good, I. J. 1952. "Rational Decisions." *Journal of the Royal Statistical Society, Series B*, vol. 14 (1): 107–114.

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

Jeffrey, Richard C. 1968. "Probable Knowledge." In *Probability and the Art of Judgment*, 30–43. Cambridge University Press, Cambridge. Originally published in *The Problem of Inductive Logic*, I. Lakatos (ed.), North-Holland Publishing Company, Amsterdam, 1968, pp. 166–180.

Jehle, David & Brandon Fitelson. 2009. "What is the 'Equal Weight View'?" *Episteme*, vol. 6 (3): 280–293.

Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science*, vol. 65 (4): 575–603.

—. 2007. "Epistemic Deference: The Case of Chance." *Proceedings of the Aristotelian Society*, vol. 107 (1): 187–206.

—. 2009. "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief." In *Degrees of Belief*, Franz Huber & Christoph Schmidt-Petri, editors, 263–300. Springer, Dordrecht.

Kelly, Tom. 2010. "Peer Disagreement and Higher Order Evidence." In *Disagreement*, Richard Feldman & Ted Warfield, editors. Oxford University Press, New York.

Oddie, Graham. 1997. "Conditionalization, Cogency, and Cognitive Value." *British Journal for the Philosophy of Science*, vol. 48 (4): 533–41.

Percival, Philip. 2002. "Epistemic Consequentialism." *Proceedings of the Aristotelian Society*, vol. 76 (1): 121–51.

Savage, Leonard. 1971. "Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association*, vol. 66: 783–801.

Shogenji, Tomoji. 2007. "A Conundrum in Bayesian Epistemology of Disagreement." Ms., Rhode Island College.

Swanson, Eric. 2007. "Rational Credence and the Value of Truth." In Gendler & Hawthorne (2007).

Weatherson, Brian. 2011. "Disagreements, Philosophical and Otherwise." Ms., Department of Philosophy, Rutgers University.