

# Group Epistemology Essay

Adam Chalmers

November 5, 2016

## 1 Group Epistemology

Epistemology is the study and philosophy of learning, belief and knowledge. Traditionally, epistemology is centered around the individual — how one person should learn about the world around them, what knowledge a person has, and what justifies a certain level of belief in a proposition. Group epistemology asks these questions not of individual agents, but *group* agents.

It often makes sense to ask what a group believes. For example, does a board of investors anticipate their company becoming profitable? Does a jury believe a defendant is innocent? What is the consensus opinion of a panel of climate scientists? What does a team of pundits jointly anticipate the Liberal Party's primary vote percentage to be? In each of these cases, we're not interested in the opinions of any individuals in the group. We're interested in the group's overall opinion, because the group itself has some sort of authority over and above the individuals who comprise it.

Although group and individual epistemology share the goals and concerns, group epistemology involves several problems which individual epistemology does not face. For example, how do we aggregate the beliefs of many individuals into one group belief? How many individuals in a group are required to possess some piece of evidence before we can safely say the group as a whole possesses it? How do we resolve contradictions between the beliefs of different group members? These issues simply don't present themselves when considering the beliefs, knowledge and reasoning of individuals. This paper aims to examine the specific question of *group credence*: how do we determine the credence<sup>1</sup> a group has in a given proposition?

---

<sup>1</sup>In this paper I will use 'belief' to designate a binary propositional attitude (agents do or don't believe a certain proposition) and credences as probability assignments (an agent may assign probability 0.3 to a proposition being true).

One intuitive solution to the group credence problem is known as *unweighted linear averaging*. In this scheme, groups should just average every individual's credence in a proposition to get the overall group credence. Unfortunately this approach yields poor results in a number of cases, which I shall examine below. A variety of more complicated procedures have been proposed which aim to succeed where unweighted linear averaging (or other approaches) fail. In this paper, I will examine a recently-published theory, Dietrich & List (2013)'s *Probabilistic Opinion Pooling* (hereafter POP) and evaluate its success as a solution to the group credence problem.

## 2 Evaluating group credence procedures

To demonstrate the complications inherent in group credence functions, I will examine scenarios where the unweighted linear averaging rule yields poor outcomes. Later in this paper I will use these same scenarios to evaluate how successful POP is as a group credence procedure.

In order to outline unweighted linear averaging I must first define some terminology. Call a *credence function* a function which maps propositions to probabilities (between zero and one). Given a group  $G$  of  $n$  individuals and a list of their credence functions  $CR = \langle cr_1, cr_2, \dots, cr_n \rangle$ , unweighted linear averaging says that the credence function for the entire group  $G$  is

$$cr_G(P, CR) = \frac{1}{n} \times \sum_{cr \in CR} cr(P)$$

Russell et al. (2015) outlines a number of properties which are *prima facie* desirable in group credence functions. Unweighted linear averaging has a number of these properties, such as

- **Anonymity:** if  $CR'$  is a permutation of  $CR$ ,  $cr_G(P, CR) = cr_G(P, CR')$  (i.e. the identity of each individual doesn't matter, only their credence).
- **Systematicity:** the credence of a specific proposition  $P$  is a pure function of each individual's credence in  $P$ , and nothing else — neither credences for other propositions, nor outside information beyond credences will affect the output of  $CR$ .
- **Unanimity:** if each individual in the group has the same credence for  $P$ , then the group should have that credence too.

Unweighted linear averaging has **anonymity** because it treats every individual in the group the exact same way. Anonymity is motivated by considerations of fairness — a group should use information from all its members equally, and not privilege one member above others. There shouldn't be a small minority responsible for the group's decisions, and there certainly shouldn't be a *dictator* individual whose credences determine the entire group's.

The procedure meets **systematicity** because it calculates the group credence for a proposition using only the individual credences for that proposition (averaging them). Systematicity is desirable because otherwise group members could manipulate the group's decision by introducing other propositions or outside information (sometimes called 'irrelevant alternatives')

**Unanimity** is a natural consequence of using averaging: if all the numbers being averaged are equal, then the average will be that value too. Unanimity has an intuitive appeal — if every member in the group agrees, then there's no dispute to even resolve, and thus no need to consult a group credence function.

## 2.1 Failure to respect conditionalisation

Unfortunately, unweighted linear averaging fails to meet several other important criteria of a good credence aggregation function — for example, it does not *commute with conditionalisation*. As Russell et al. (2015, pg. 1290) puts it, “if every individual updates on new information by conditionalizing, the resulting average credence won't be the same as what they would get from first averaging their unconditional credences in each proposition, and then conditionalizing on the new evidence.” Unweighted linear averaging doesn't respect conditionalisation (Loewer & Laddaga, 1985, pg. 87).

This is a serious problem because the group can now assign incoherent probabilities. Consider this example based on Russell et al. (2015, pg. 1288). A group of two members M1 and M2 disagree over whether anvils or balloons are a good investment. Let A be the proposition that anvils are a good investment and B the proposition that balloons are. The two members assign credences like so:

- M1 assigns  $P(A) = 2/3$  and  $P(B) = 1/3$
- M2 assigns  $P(A) = 1/3$  and  $P(B) = 2/3$

Given these assignments, both members calculate the probability of exactly one of balloons or anvils being a good investment, i.e.  $P(A \text{ xor } B)$ , at  $5/9$ .

If the group they jointly form assigns its group-level credences via unweighted linear averaging, it will assign  $P(A) = \frac{1}{2} \times (\frac{1}{3} + \frac{2}{3}) = \frac{1}{2} = 0.5$ . The same reasoning yields 0.5 as group credence in B. Averaging 5/9 and 5/9 yields 5/9, so the group credence in  $P(A \text{ xor } B)$  is also 5/9. However, this is inconsistent with the laws of probability. If  $P(A) = 0.5$  and  $P(B) = 0.5$ , then  $P(A \text{ xor } B)$  should also equal 0.5. Therefore, the group has reached probabilistically inconsistent credences by following unweighted linear averaging.

This means the group can now be tricked into buying a Dutch Book — “a set of bets bought or sold at such prices as to guarantee a net loss” (Hájek et al., 2008). No agent with rational probability assignments would ever buy a bet which is guaranteed to lose them money. However, if an agent’s probabilities are inconsistent (like the M1 + M2 group agent above), it will approve of this irrational, guaranteed-loss purchase. Therefore, any group whose credences do not commute with conditionalisation opens themselves up to irrational behaviour. Unweighted linear averaging is therefore a poor choice of group credence function, because groups which use it will endorse irrational decisions which are guaranteed to lose them money.

## 2.2 Failure at the Elephant Test

Unweighted linear averaging also fails at what I call the **Elephant Test**. In the ancient fable, three blind men come across an elephant and each touches a different part of it. One man touches its trunk, and concludes the animal is a snake. One man touches the tail and concludes the animal is a horse. One man touches the leg and concludes it is a rhinoceros<sup>2</sup>. Let  $H_{elephant}$  be the hypothesis that the three men have encountered an elephant.

No individual in this group has sufficient evidence to believe  $H_{elephant}$ , and therefore they would each assign low credence to it. The first individual would give high credence to  $H_{snake}$ , the second would give high credence to  $H_{horse}$  and the third would give high credence to  $H_{rhino}$ . This means that when the group’s unweighted linear average credence is taken, the group will jointly assign low credence to  $H_{elephant}$  (because the average of three small numbers is itself a small number).

Notice, however, that between the three members the group does possess enough evidence to jointly confirm  $H_{elephant}$ . An elephant is the only animal

---

<sup>2</sup>Estlund (2009, pg. 229) uses this fable as an example of voters who do not have accuracy above 50%. I attribute my recognition of the fable’s philosophical significance to him, however he uses the fable for a completely different purpose to me.

with the sort of trunk, legs and tail the men encountered. A good group credence procedure, therefore, should be justified in assigning high credence to  $H_{elephant}$ . This demonstrates a general failing of unweighted linear averaging — it fails to make proper use of the evidence available to the group.

Passing the Elephant Test is absolutely vital for a group credence procedure. Failure indicates that the group is not making proper use of all the information (and/or evidence) available to its members. I believe misidentifying the elephant in this case demonstrates irrationality at a group level.

Of course, not all mistakes at a group level demonstrate irrationality. Some mistakes just demonstrate lack of information: for example, if the group lacks evidence for a proposition, then it is in fact *rational* to assign low credence to that proposition. This does not hold in the Elephant Test — the group jointly possesses evidence necessary to justify high credence.

Mistakes based on probabilistic outcomes are also not evidence of irrationality: if a group assigns credence 0.5 to a fair coin landing heads, and it lands tails, this is not a failing of the group. It is merely an inevitable consequence of judgement under uncertainty.

However, failing to identify the elephant is unlike these failures. Unweighted linear averaging groups fail to identify the elephant not because they lack information or evidence, not because of probabilistic failure, but simply because they *do not use the information available to them*. They ignore relevant evidence, which is surely a sign of irrationality.

To summarise: although unweighted linear averaging has a number of desirable properties (anonymity, systemacity, unanimity) it fails to commute with conditionalisation. Therefore, groups whose credences are chosen with this mechanism will be vulnerable to Dutch Books and can be easily exploited. It also fails the Elephant Test, which indicates it ignores relevant evidence possessed by group members. These are both serious problems and demonstrate the need for a more sophisticated group credence procedure, such as POP.

### 3 Probabilistic Opinion Pooling

#### 3.1 What is POP?

Dietrich & List (2013) proposes POP as a way to assign group credences to propositions in the group's *agenda* — the set of propositions which the group is considering. POP involves dividing the agenda into two sets (premises and

conclusions) and uses different credence-calculation rules for each set.

Consider this example. Say a group is deciding whether or not to buy a beachfront property. Their agenda might include three propositions (and their negations):

- T: There will be a tsunami on this coastline in the near future
- L: Land taxation rates will not be increased in the near future
- B: Beachfront housing is a good short-term investment

Intuitively, the likelihood of T and L should be independent of each other. It is difficult to see a way they could be correlated. However, the likelihood of B depends on both T and L — if either a tsunami or a land tax increase occurred, the investment value of beachfront housing would diminish. It seems rational for a group to determine their confidence in T and L independently, consulting only facts about tsunamis for T and facts about politics and economics for L. Once the group has credences for T and L, they can then use them (perhaps alongside other considerations) to determine their credence in B.

In short, some propositions in the group’s agenda are independent relative to each other; others are probabilistically dependent on each other. Call these ‘independent’ propositions ‘premises’. It seems rational to calculate premises independently, then use the group’s credences in each premise to calculate conclusion credences.

This is the idea behind POP. Groups should “first aggregate individual probabilities for basic events [premises] and then let the resulting collective probabilities constrain the collective probabilities of all other events” (Dietrich & List, 2013, pg. 6).

The group’s credence in a premise comes from simply pooling each individual’s credence in that premise. This pooling function could be an arithmetic average (just like in unweighted linear averaging does), but other functions are possible too. Importantly, premises are *independent* of each other. A premise P’s credence comes *only* from individual credences in P — not from any individual or group credence in any other proposition on the agenda.

Unlike premises (which are required to be independent of each other), the credence of each conclusion will depend on the credences assigned to premises. Once the group has assigned credences to all premises, it can then use them to calculate the credence for each conclusion. The premises *constrain* the group’s credences in the conclusions. It would be probabilistically incoherent for the

group to believe both that there will *definitely* be a tsunami next year and also that beachfront housing is a good short-term investment. Therefore, the need to maintain probabilistic coherence means *premise credences constrain conclusion credences*.

POP is based on List & Pettit (2011)’s Premise-Based Procedure, which can be viewed as a special case of POP which deals with beliefs rather than credences (i.e. binary do/don’t beliefs rather than probabilistic degrees of belief). In the Premise-Based Procedure, propositions are divided into premises and conclusions in a way such that premises are logically independent of each other, and conclusions are logically related to premises. Individual beliefs about premises are aggregated to group beliefs about premises.<sup>3</sup> The group then uses the logical implications of the premises which it believes to determine which conclusions the group believes.

## 3.2 Evaluating POP

Does POP succeed as a theory of group credence? I will begin by noting that POP is underspecified in two ways. Firstly, it admits a (limited) range of pooling functions.<sup>4</sup> Secondly, it doesn’t specify how the group should generate conclusion credences from premise credences. If probabilistic coherence is to be maintained then the probability of each premise constrains the probability of each conclusion — but which probability function(s) should the group use to generate these conclusion credences?

### 3.2.1 Anonymity

POP’s anonymity depends on the anonymity of the pooling function; Dietrich & List (2013, pg. 21) specifically considers dictatorship by one individual as a valid pooling function. This is desirable: it is easy to determine whether a given POP procedure is anonymous, and it also allows the theory to accommodate dictatorship should dictatorship be a good choice for a specific group or agenda.

---

<sup>3</sup>Aggregating binary beliefs requires different methods to aggregating credences. Common examples are majority rule (the group believes P if a majority of members believe P), unanimity (the group believes P only if all members believe P), dictatorship (the group believes P if a specific member believes it) and supermajority (the group believes P if a certain percentage of members over 50% believes it).

<sup>4</sup>Specifically, “given certain logical connections between the premises, independence on premises, together with a unanimity-preservation requirement, implies that the collective probability for each premise is a (possibly weighted) linear average of the individual probabilities for that premise” (Dietrich & List, 2013).

Dictatorship in a group is not always bad: for example, if one group member is an expert on a particular topic, it may be reasonable to allow their opinion to determine the group opinion. Anonymous pooling functions allow for anonymous group credences, so POP does as well as unweighted linear averaging here.

### 3.2.2 Systematicity

To begin, note that if some two propositions aren't probabilistically independent in a group credence function, then that group credence function violates systematicity. This demonstrates that POP breaks systematicity: by design, some propositions (conclusions) are probabilistically dependent on other statements (relevant premises). Dietrich & List (2013, pg. 6) believe systematicity is "normatively unattractive": as they write, "it seems implausible to apply independence [and therefore systematicity] to composite events such as 'the economy will grow or atmospheric CO2 causes global warming,' since this would prevent us from using the probabilities of each of the constituent events in determining the overall probability." This reasoning demonstrates that while POP breaks systematicity, it's actually a strength of the procedure, not a weakness. Composite propositions shouldn't be independent, and therefore systematicity is actually often an undesirable property of a group credence procedure.

### 3.2.3 Unanimity

Dietrich & List (2013) distinguishes two different kinds of partial unanimity:

- **Consensus preservation:** if every individual assigns credence 1 to a proposition, the group will assign credence 1 to a proposition.<sup>5</sup>
- **Conditional consensus preservation:** for all premises A and B, if all group members think the credence of A given B is 1, then the group will too.

In POP, both these properties hold for premises, but not conclusions. Why does POP make such weak guarantees about unanimity? It turns out that stronger forms of unanimity are often undesirable.

The group often benefits from keeping unanimity on premises but breaking unanimity for conclusions. Consider this example scenario from (Pettigrew, forthcoming):

---

<sup>5</sup>This implies the preservation of credence 0 unanimity as well.



Suppose two historians, Jonathan and Josie, are researching the same question, but in two different archives. Both know that there may be a pair of documents, one in each archive, whose joint existence would establish a controversial theory beyond doubt. Jonathan finds the relevant document in his archive, but doesn't know whether Josie has found hers; and Josie finds the relevant document in her archive, but doesn't know whether Jonathan has found his. Indeed, each assigns a very low credence to the other finding their document; as a result, both have a very low credence in the controversial theory.

We can summarise each historian's credences as follows:

	Jonathan	Josie
Document A exists	1	0.1
Document B exists	0.1	1
Controversial theory	0.05	0.05

The existence of each document should be considered a premise (in the POP sense); the controversial theory should be a conclusion. In this example, both individuals unanimously assign credence 0.05 to the theory, but the group jointly possesses enough evidence to confirm the theory with certainty. This demonstrates why unanimity-preservation for conclusions is undesirable: as Dietrich & List (2013) says, “a consensus on a non-basic event could be ‘spurious’ in the sense that there might not be any agreement on its basis.” If each individual justifies their credence for a conclusion based on differing evidence, the joint sum of their evidence may justify a different conclusion credence. In this way, unanimity for conclusions is actually often an undesirable property of a group credence procedure.

[TODO: elaborate on this?]

### 3.2.4 Conditionalisation

I previously demonstrated that group credence functions should commute with conditionalisation, otherwise they will not assign their probabilities in coherent ways and open themselves up to irrational purchasing behaviour and exploitation via Dutch Books. POP happens to commute with conditionalisation, which is a marked improvement over many other credence procedures which use pooling.

How does POP protect groups from Dutch Books? Consider again the earlier example from page [TODO: LOOK UP PAGE NUMBER]. The group's individuals assigned these credences:

- **M1:**  $P(A) = 2/3$ ,  $P(B) = 1/3$ ,  $P(A \text{ xor } B) = 5/9$
- **M2:**  $P(A) = 1/3$ ,  $P(B) = 2/3$ ,  $P(A \text{ xor } B) = 5/9$

Therefore, if group credences are the average of individual opinions, the group will believe  $P(A) = P(B) = 1/2$  and  $P(A \text{ xor } B) = 5/9$ . This is probabilistically incoherent, because the laws of probability imply that for all and any propositions  $X$  and  $Y$ , if  $P(X) = P(Y) = 1/2$  then  $P(X \text{ xor } Y)$  should be  $1/2$ , not  $5/9$ .

Under POP, propositions are classified as either premises or conclusions, with the requirement that all premises must be probabilistically independent. Therefore, if  $A$  and  $B$  are both premises,  $(A \text{ xor } B)$  must be a conclusion. This means  $(A \text{ xor } B)$  is not chosen by averaging individual credences for  $(A \text{ xor } B)$ , but rather by applying probability laws to the group credence in the premises.

Group credence for premises  $A$  and  $B$  are determined by averaging, so the group has credence  $1/2$  in each. The group can then determine its credence in  $(A \text{ xor } B)$  by applying probability theory and noting that  $P(A \text{ xor } B) = P(A \text{ and not } B) + P(B \text{ and not } A) = 1/2$ . Thus, POP yields a probabilistically consistent credence assignment in this example — a marked improvement over unweighted linear averaging.

### 3.2.5 Elephant Test

POP has difficulty passing the Elephant Test. Part of the difficulty is that the Elephant Test can be formulated into an agenda of propositions in many different ways. The quality of POP's answer will vary depending on which precise formulation is being used. POP also tends to dilute evidence: if one group member has significant evidence that could greatly alter the group's opinion, increasing group size will drown this evidence out.

Here's one way of formulating the Elephant Test as an agenda. Consider the following credence assignments of the three people in this group, M1, M2 and M3.

	Snake	Horse	Hippo	Elephant	Other
M1	0.6	0.01	0.01	0.02	0.36
M2	0.01	0.6	0.01	0.02	0.36
M3	0.01	0.01	0.6	0.02	0.36

Note that Other is equivalent to “not Snake, Horse, Hippo or Elephant”. One could therefore treat Snake, Horse, Hippo and Elephant as premises and Other as a conclusion. Averaging individual credences for premises and assigning a coherent credence for the conclusion yields these group credences:

	Snake	Horse	Hippo	Elephant	Other
Group	0.21	0.21	0.21	0.02	0.35

It seems POP fails the Elephant Test on this formulation. If the elephant proposition is a premise, then its group credence will simply be the average of individual credences. Just like in unweighted linear averaging, this yields poor results. If no individual personally possesses enough evidence to believe the elephant proposition, then the group will distrust the elephant proposition too.

However, more sophisticated formulations of the problem may help. What if the group considered three additional propositions: that the animal has a prehensile appendage, is a quadruped, and has a tail. These are all premises, and the remaining propositions (is snake, is horse, is hippo, is elephant, is other) are all conclusions.

	Prehensile appendage	Quadruped	Tail
M1	0.9	0.01	0.01
M2	0.01	0.9	0.01
M3	0.01	0.01	0.9
Group average	0.31	0.31	0.31

Given these premise credences, it’s implausible that the group will determine the animal is an elephant. After all, the group doesn’t even have significant confidence that the animal has a tail, or is a quadruped. It seems that even this more sophisticated formulation fails the Elephant Test.

These examples demonstrate the aforementioned dilution of evidence. In the second formulation, the group only has 1/3rd confidence in each piece of evidence, despite each individual being fully confident in their individual evidence. POP fails the Elephant Test because it makes poor use of evidence. In particular, POP has no concept of *group level evidence*, only individual evidence. This

makes it poorly suited to single individuals possess highly significant information, such as the Elephant Test.

### 3.3 Telling premises from conclusions

We have seen that POP improves over unweighted linear averaging in a number of ways. It disrespects unanimity and systematicity in situations where it makes sense to do so; it commutes with conditionalisation. Neither system passes the Elephant Test. So far, POP does at least as well as unweighted linear averaging in all areas, and it appears to be a highly valuable (although imperfect) group credence procedure.

However, I have some additional concerns with POP. My first is that the POP methodology doesn't properly specify how groups should distinguish premises from conclusions.

## References

- Dietrich, F., & List, C. (2013). Probabilistic opinion pooling generalized Part two: The premise-based approach.
- Estlund, D. M. (2009). *Democratic authority: A philosophical framework*. Princeton University Press.
- Hájek, A., et al. (2008). Dutch book arguments. *The Oxford handbook of rational and social choice*, (pp. 173–195).
- List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press.
- Loewer, B., & Laddaga, R. (1985). Destroying the consensus. *Synthese*, 62(1), 79–95.
- Pettigrew, R. (forthcoming). On the accuracy of group credences. *Oxford studies in epistemology*, 6, to appear.
- Russell, J. S., Hawthorne, J., & Buchak, L. (2015). Groupthink. *Philosophical studies*, 172(5), 1287–1309.