Jun 3rd, 10:00 AM - 11:11 AM

Confidence in Arguments in Dialogues for Practical Reasoning

Waleed Mebane
University of Windsor

Confidence in Arguments in Dialogues for Practical Reasoning

WALEED MEBANE
Argumentation Studies
University of Windsor
401 Sunset Ave., Windor, Ontario
Canada
mebane@uwindsor.ca

Abstract: For the context of practical reasoning, this paper suggests a method of assessing the level of confidence we should rationally have in arguments. It draws from dialectic, which induces the elaboration of reasons for a position, and on auditors' prior knowledge. Accurate assessment depends on evidential standards, on selecting dialogue moves according to their practical and epistemic importance, and on selecting auditors according to their competence and diversity of relevant knowledge.

Keywords: Argument strength, Bayesian argumentation, deliberation, dialectical argumentation, dialogue, public policy, social epistemology

1. Introduction

This paper is concerned with arguments that could be used as a part of public (policy) deliberation. In such deliberations, some small group deliberates about what actions to take (or policy that indirectly specifies actions to take) to respond to some issue of public concern. The decisions taken by the smaller group, e.g., members of a representative assembly, will bind a larger group, a whole political community. The smaller group should be able to publicly justify their decision in terms that are acceptable to all rational (and perhaps reasonable) members of the community.

One way of justifying a decision could be with arguments for the claim that the decision is the best decision for the community with regard to the issue of concern. Arguments in this domain generally involve predicting future states of affairs and weighing expected positive and negative consequences of actions. These are things that cannot be known for certain. A dialogue evaluating proposals for action and resulting in a decision to accept one such proposal together with a justifying case in favor of that proposal will contain uncertain arguments. However, it will not only be the predictions that contain uncertainty. Just as with most arguments, we will not be able to say for certain that the generalizations warranting inferences from premises to conclusion are without exception, and we will not generally be able to say that the premises are certainly true (or certainly false). The arguments will be defeasible in one way or another. In the course of the dialogue, we would like to reduce the uncertainty. At some point a deliberation dialogue must come to an end for practical reasons. Action becomes expedient or other issues plausibly come to have higher priority for our attention. At that time, if we have reduced the uncertainty as much as we could have, then we can claim that our decision is rational on the grounds that we could not have done better.

The view of argument taken in this paper is of a set of premises and a conclusion such that the truth of the conclusion can be rationally judged as more likely when we know that the premises are true than otherwise. Getting at the truth of some claim(s) will be considered an important purpose of the dialogue-type examined here. This is as compared with reaching a consensus on the acceptability of a claim, which might coincide with the truth-seeking goal depending on the procedures we adopt.

Now public justification of a decision does not necessarily require (i.e., under all plausible theories of public justification) predicted close proximity to truth. It seems possible, even if difficult, simply to make out a case that would be acceptable to all members of the larger group. For example,

suppose the small group has been intentionally constituted of representatives from all stakeholding social groups[1] in the community at-large. Next, the small group has a dialogue in which they seek a fair exchange of valued 'goods' (predicted states of affairs that a representative takes to be good) so that they put together a final proposal that maximizes the good for each representative up to the point that none are made worse off in consideration of the goods they value. This is a method of fair bargaining. It is plausible, then, if the community has been well represented by the small group, that each member of the community could accept the final decision in virtue of the fact that it gives them as much of the goods they value and requires them to give up as little of what they value (in proportion as they value them) as could obtain under fair conditions. The method only misses opportunities to have one's mind changed about what one ought to value or to change the minds of others.[2] As long as no one is too concerned about that, the proceedings, containing steps of preference aggregation and agreement on their appropriateness, could be an adequate justification. The dialogue does not have to contain arguments for the fairness and equitability of the distribution, but if it does, those arguments only need to correspond with the preferences of the participants aggregated in a way they find acceptable.

The acceptability of such a justification could be questioned on practical grounds: would every member of the community really be satisfied if some believed they could have obtained more of their favored goods by persuading others to change their value priorities? But what can be questioned from a theoretical perspective is the rationality of the procedure: whether every member of the community ought to accept the decision. So truth-seeking procedures are taken to be interesting even though they might not turn out to produce different results than very good consensus-seeking procedures.

2. Using probability

A property of arguments that can be compared to determine which of two conflicting ones should be accepted has been called 'strength' (Hahn & Oaksford, 2006b; Prakken, 2010, p. 169). Prakken observes that there are three ways to attack an argument: on its conclusion, on its premises, and on the inferential link between premises and conclusion (Prakken, 2010, p. 169). Deductive arguments have inference rules that are certain and cannot be defeated; furthermore, if a deductive argument's conclusion is false, it must be because one or more of its premises is false. So effective attacks against deductive arguments are limited to attacks against their premises. A deductive argument is strong whenever it uses a valid inference rule and its premises are all true. The argument strength of interest in this paper is the corresponding property in non-deductive arguments. Inference warrants in such cases may be generalizations admitting exceptions. Premises may be likely but uncertain. Argument strength is a measure of the amount of support given to the conclusion by its premises and the inferential link

1 By this I mean, all groups which would be affected by actions resulting from a policy decision on the issue.

2 In theory, there might be a case in which each group is worse off in consideration of the goods they value, but the community is somehow better off as a whole. I cannot think of an example of this, though, unless it comes about because people do not initially value what they ought to value and what they ought to value would emerge through conversation.

between its premises and its conclusion.[3] Furthermore, the premises of an argument can be supported by further arguments, and the strengths of those arguments should influence the strength of the original argument under consideration, the premises of which are being supported. That is because the argument's strength is the strength of its ability to justify its conclusion. The likelihood of the truth of each conclusion should be proportional to the (aggregated) strength of the arguments for that conclusion.[4]

Various ways of comparing conflicting arguments have been proposed. Argument analysts could propose a *labelling* of arguments, as 'in' (accepted) or 'out' (not accepted) (Caminada, 2006). Once some key arguments are labelled it would be possible to determine which arguments are defeated by arguments labelled 'in' and then determine the status of the remaining arguments. Decision between conflicting arguments based on an ordering of preference between arguments has been proposed (Amgoud & Cayrol, 1998). Thomas Gordon has proposed (and implemented a formal system for) pro and con "balancing" argumentation by applying weights to considerations; practical reasoning arguments can be distinguished as better (or worse) by whether their recommended actions achieve more positively weighted goals (Gordon, 2018). Probabilistic argumentation has been proposed that labels arguments and attacks between arguments with the probability values (Li, Oren, & Norman, 2011). *Possibility* calculation has been proposed for argumentation for decision making (Amgoud & Prade, 2004). And there are other proposals besides these.

Hahn and Oaksford have suggested that Bayesian probability might provide a measure of argument strength (Hahn & Oaksford, 2006b) and have developed the idea in several papers (Hahn & Oaksford, 2006a; Hahn, Oaksford, & Corner, 2005; Hahn, Oaksford, & Harris, 2013; Oaksford & Hahn, 2013). Their approach develops separate Bayesian models for each argument scheme that provides a unique characterization of details of the inference. A Bayesian calculation of argument strength would rely on eliciting (or otherwise discovering) some subjective probability values from discussion participants and in some cases also obtaining some non-subjective empirical data; whereas most of the aforementioned approaches only use one value per argument. But the Bayesian calculus has some properties that make it especially interesting. For the purposes of this paper, two will be noted. The first is that the prescribed way in which a Bayesian reasoner should update their beliefs is conservative. If a large amount of weighty evidence has been accumulated for a conclusion, new evidence generally does not reverse the conclusion (although it can if the new evidence is, what I am calling, for now, sufficiently weighty). The second interesting property is the empirical evidence that Bayesian rationality is a good model of human rationality (see, e.g., Oaksford & Charter, 2007, Chapter 6; Oaksford & Hahn, 2013).

This paper will take a variation of Bayesian belief updating as its model of rationality. The potential difficulty of creating Bayesian models and calculating argument strength values from them in

3 About ways to attack arguments, Prakken elaborates that there are two ways to defeat an argument, by removing support for a claim, undercut, and by contradicting a claim, rebuttal. Undercut can happen, for example, when an exception is found to a defeasible inference rule (Prakken, 2010, p. 169). Exceptions do not contradict the inference rule because it still stands in the general case; they only provide a reason why the inference rule is not applicable in the current case. A measure of argument strength is only needed in the case of rebuttals. A rebuttal could target any of premises, inference warrant, or conclusion, even though the inference warrant is commonly attacked by presenting an exception. Inference warrants not supported by evidence (what Toulmin calls 'backing') would not by themselves arguments. These details are not needed in this paper. It is enough to know that the use to be made of an argument strength measure is to resolve a conflict between arguments having inconsistent conclusions.

4 Belief in the truth of some premises might be justified independently of any supporting argument, for example, by being sufficiently plausible.

the general case will be set aside and not directly addressed in this paper. Bayesian belief updating is based on Bayes' Theorem, shown as Equation 1.

Equation 1
Bayes' Theorem

$$P(H|E) = P(H) \times \frac{P(E|H)}{P(E)}$$

The letters H and E were used in this rendering of Bayes' Theorem to indicate hypothesis and evidence, respectively. Interpretation of the theorem as a belief updating function, takes the probability of the hypothesis given the evidence, P(H|E), to be the *posterior* probability of the hypothesis. That is, the degree of belief one should have in the hypothesis after the evidence has been observed. In the case of an argument, it will become the degree of belief one should have in the argument's conclusion. The term, P(H), is interpreted as the *prior* probability of the hypothesis, or the degree of belief one had in the hypothesis prior to observing the evidence. The final term is the probability that the evidence would be observed if the hypothesis were true divided by the probability of observing the evidence whether or not the hypothesis is true. We can express the denominator in terms of whether or not the hypothesis is true by applying the law of total probability, Equation 2.

Equation 2
Law of Total Probability

$$P(A) = \sum_{n} P(A|B_n) \times P(B_n)$$

In Equation 2, the B_n are all distinct, and in our Bayesian belief updating case they represent all alternative hypotheses, the probabilities of which sum to one. In the simplest case we have:

Equation 3
Law of Total Probability for two hypotheses

$$P(A) = P(A|B) \times P(B) + P(A|{\sim}B) \times P({\sim}B)$$

Now substituting this in for the denominator of Bayes' Theorem, we get:

Equation 4
Second form of Bayes' Theorem

$$P(H|E) = P(H) \times \frac{P(E|H)}{P(E|H) \times P(H) + P(E|{\sim}H) \times P({\sim}H)}$$

The probability of P(H) + P(~H) is 1. These are our prior beliefs about the probabilities of the truth of the hypothesis and the complement, that the hypothesis is false. It is also the case that evidence that increases the posterior probability of the hypothesis, decreases the posterior probability of its complement. Suppose we know of an additional hypothesis, H_2, the denominator becomes $P(E|H) \times P(H) + P(E|H_2) \times P(H_2) + P(E|\sim H\&\sim H_2) \times P(\sim H\&\sim H_2)$. Notice that when considering how to design a test to disconfirm H, a good use of resources might be to try to find evidence for H_2 in the case that $P(H_2) > P(\sim H\&\sim H_2)$ or for $\sim H\&\sim H_2$ in the case that $P(\sim H\&\sim H_2) > P(H_2)$. In other words, work on the hypotheses in which we have the strongest beliefs. In argumentation the hypothesis is the conclusion of an argument; the evidence is the reasons for that conclusion. In actual cases, it might not be obvious how to find values for terms like P(E|H) and P(E|~H). These values often come from empirical observations. For some common argument schemes, Hahn and Oaksford have presented Bayesian update formulas specific to those schemes (see Hahn & Oaksford, 2006b); that is, they have modelled the relationships between terms in the argument schemes and terms in Bayes' Theorem according to a subjective Bayesian interpretation of probability.

These update formulas make the assumption that we are certain of the evidence; we are only measuring the impact of the evidence on our belief in the hypothesis. That is sensible when updating beliefs on evidence that is the observed occurrence of an event. We would only doubt it if we doubt our instruments of perception or our ability to interpret our perceptions accurately. But reasons for conclusions in arguments are generally not certain. In order to account for that, Godden and Zenker have proposed modelling update for arguments using *Jeffrey Conditionalization*, with the new update formula shown in Equation 5 (Godden & Zenker, 2018, sec. 3.8).

Equation 5
Belief update with Jeffrey Conditionalization (Godden & Zenker, 2018, fig. 16)

$$P_f(C) = P(C) \times \left[\frac{P(R/C)}{P(R)} \times P_f(R) + \frac{P(\sim R/C)}{P(\sim R)} \times P_f(\sim R)\right]$$

Subscript f now indicates a posterior probability, C refers to the argument's conclusion and replaces H from previous equations, and R refers to the argument's reasons for C and replaces E. If $P_f(R)$ is 1, then $P_f(\sim R)$ is 0, and we have Bayes' Theorem as shown in Equation 1. Otherwise, we use the terms $P_f(R)$ and $P_f(\sim R)$ to indicate our subjective level of belief in the reasons. This formulation is the model of rationality for updating beliefs based on arguments which is to be understood through the rest of this paper. The part in square brackets is called the *impact term* by Godden and Zenker, according to the name given to it by Korb (Godden & Zenker, 2018, pp. 1723–1727; Korb, 2004, p. 45). The impact term tells us the contribution of the probability of the reasons and our inference warrant, the likelihood that the conclusion is indicated in the presence of the reason[5], on our final belief in the conclusion, leaving out the influence of our prior belief in the conclusion.

5 This warrants our inference because of its conformity to the laws of probability, if we accept findings that probabilistic measures of certainty or uncertainty model normative rationality and if we accept that we ought to update our beliefs by *conditionalizing* on the evidence, that is, by taking the evidence for granted as true (or by using Jeffrey Conditionalization, giving the evidence partial credence).

3. Incorporating Bayesian updating into dialogue

Arguments for and against proposals are being developed over the course of a dialogue; they are not static. When an argument is attacked, a new argument may be put forth to counter the attacking argument. This new argument contains considerations that shore up the original argument. The original argument plus these new considerations is itself an argument, which I will usually refer to as the 'case' being made out in support of the proposal. The rest of this paper is concerned with the argument quality of arguments in two argument categories: the quality of smaller arguments used in the dialogue to attack or support individual claims; and the quality of the cases for (or against) proposals. Both categories are of arguments, but since a case is being developed through the dialogue, the goodness of the dialogue procedure will influence how strong that case may become (relative to some ideal case). We should expect in general that a case that is ultimately justified will have been weaker at the start of a dialogue than at the end. The word 'strength' seems to suit the kind of quality I am interested in isolating in an argument; however, it has sometimes been used to refer only to the inferential link between premises and conclusions. I also wish to refer to levels of confidence in the truth of the premises.

In ordinary language, 'strength' is a property of an object which property does not depend on the observer. A strong object will be strong no matter who is looking. Also, an object that is strong at one time will not be weak at another time, unless the object itself has undergone a change. It is not dependent on change of setting. If we call it strong at one time and weak at another, it ordinarily must be because we were mistaken at one of those times.

Therefore, the property of 'strength' that we should like to have is as follows. It should be unlikely that our judgement that an argument is 'strong' will be reversed by new evidence and counterarguments. It would be an embarrassment if an argument we called strong were defeated in the next dialogue move. If the argument was wrong, how could it have been strong? But since the arguments in our dialogues are defeasible, it is all but inevitable that arguments will be defeated given the limits of individual knowledge and enough time. What we are looking for is a kind of *prediction*, a prediction of whether the argument will turn out to have true premises and a robust warrant of inference. This prediction of the truth-likeness of an argument, I propose to call 'confidence' rather than strength. With the idea of a prediction, we can get close in defeasible argumentation to the notion of 'strength' we inherit from ordinary language.[6] Unlike strength, our prediction is inherently prone to change as new information comes in, but we will make it as stable as we can.

Let us suppose, for now, that each dialogue has exactly two participants.[7] A straightforward way to use Bayesian calculations of degrees of belief (given an adequate Bayesian model) to evaluate arguments in a dialogue would be for each of the two participants to make their own subjective probability calculations for the arguments of their opponent as well as for their own arguments. A participant determines that the opponent's conflicting argument has defeated their own when the

6 One might think that it is not unlike what we call 'strength'; instead, it is as though we were observing something from a distance. Something that looks strong might turn out to be weak on closer inspection, such as a block of iron that turns out to be rusted through, and therefore brittle. The intention in using the word 'prediction' is to use methods that are most likely to result in a stable assessment, methods that are selective and efficient with regard to the information gathered and the way in which beliefs are formed from it, rather than gathering information by haphazard means and then taking as our assessment a snapshot of the current state of our knowledge about the observed thing.

7 It is commonplace for deliberations to include more than two participants and for models of deliberation dialogue to accommodate that (see, e.g., Kok, Meyer, Prakken, & Vreeswijk, 2011, p. 34; McBurney, Hitchcock, & Parsons, 2007). However, it will be easier to start by considering the case of two participants.

posterior probability of the opponent's conclusion is greater than that of their own conclusion. (Since they are conflicting arguments, the two conclusions are inconsistent.)

This means that participants must use (or attempt to use, to the best of their understanding of their opponent) premises and inference warrants that are acceptable to their opponent. A participant has a good reason to accept that their opponent's rebutting argument defeats their own when it is composed of premises that they find acceptable and a warrant of inference that they find acceptable and to a greater degree in aggregate than that which they attribute to their own argument including their prior belief in the truth of their own conclusion. So, they have a reason to retract their claim unless they can marshal evidence for a new argument that successfully attacks (defeats) their opponent's defeating argument (defeater). This way, burden of proof is modelled as a rationality requirement on the individual participant. They must accept rational conclusions, including those that defeat their previously accepted conclusions, by having their degree of belief updated to an amount greater than those of their previously accepted inconsistent conclusions. The dialogue comes to a close when both participants accept a claim, either that the initial claim was correct or the counterclaim, the initial claim is contradicted; or when the participant who put forth the initial claim accepts their claim as defeated, being unable to put forth any further arguments sufficient either to shore up the claim or to defeat the last defeater by their opponent. Therefore, the dialogue ends with agreement. The strategy that participants must adopt in order to advocate their positions is to advance arguments that they expect to be assessed as strong by their respective opponent.

This means that they do not necessarily advance the arguments that they believe to be the strongest, which is to say that they believe best predict that their claim is true. This might not sound like a problem since the fact that the opponent does not agree (or is not expected to agree) is plausibly evidence that the participant's belief is incorrect. The procedure described induces each participant to account for the opposed viewpoint, as best they understand it, and results in a combined judgement at the end. However, suppose one participant has more relevant correct beliefs than the other. For example, Participant 1 is an expert in a field of the claim at issue, and Participant 2 is a layperson. Correct beliefs of Participant 1 can only be made to prevail if Participant 1 is capable of expressing them in terms Participant 2 can rationally accept from her or his epistemic position. That makes the (likely) proximity to truth of the finally developed case dependent not on the likelihood of the truth of the premises and the reliability of the warrant, but on the rhetorical competence[8] of Participant 1. A knowledgeable, but rhetorically incompetent Participant 1 cannot get their case accepted.[9] Or, alternatively, Participant 1 makes a flawed case that caters to the faulty beliefs of Participant 2. Or, Participant 1 advances less bold claims. If we are evaluating claims in order to show that some proposal for action is best with regard to addressing an issue of concern, this might lead to taking a decision that is less advantageous but more initially acceptable.

---

8 By rhetorical competence, I mean the ability of the participant to guess, intuit, or otherwise discover the kinds of reasons and warrants of inference that would be acceptable to their opponent. Likelihood of truth and acceptability to one's opponent are probably not entirely distinct or independent.

9 Bayesian belief updating has been shown to lead persons' beliefs to converge eventually regardless of their initial degrees of belief in the claim, as long as accurate evidence continues to be presented. In the particular task of estimating the mean of a population from samples, for example, convergence happens quickly (Howson & Urbach, 2006, pp. 239–241 and 245–147). So we might think that all the more knowledgeable person must do is make one argument after the other, based on all of the relevant facts she or he knows. However, the findings of convergence were based on certainty about the evidence. An interesting question for future research is whether there is a not too demanding set of conditions under which participants could converge quickly on the solution to typical decision problems of the sort usually resolved through deliberation. My intuition is that evidence and inference in such argumentation contains significantly more uncertainty than observations of sample data and that participants' beliefs would not converge quickly when priors are far apart.

The situation is similar in a case of 2 participants both having little relevant knowledge. In that case, if they have the time and access to appropriate resources and they are aware of their lack of knowledge, their best rational course of action is to attempt to learn more about the subject.[10] To the best of their existing beliefs, the rational participants should arrange things so that they gain maximum benefit from additional information and knowledgeable sources to improve their own relevant knowledge. In case either is a little more knowledgeable or resourceful than the other on some dimension of the issue, that might mean accepting the rulings of a more competent authority, a judge.

In Nicolas Rescher's model of *disputation* described in *Dialectics*, the participant to advance an initial claim is called the *proponent*. A second participant has the role of attempting to cast sufficient doubt on the initial claim so that it is shown that the proponent has not met his or her dialectical obligations. The second participant is called the *opponent*. A third party judges the outcome and is called the *determiner* (Rescher, 1977, p. 4). In order to meet dialectical obligations and be an effective advocate for an initial claim, a proponent disputing with a sufficiently good opponent is forced to put forth the best case within her or his ability. Or, at least, the case needs to be sufficient to persuade the determiner, in the face of the doubts and/or opposing considerations advanced by the opponent. In other respects, the dispute proceeds much like the dialogue described above. The participants take turns advancing arguments, either to attack some previous argument by the other participant or to shore up one of their own previous arguments (Rescher, 1977, pp. 23–24). In this case, however, participants do not need to use reasons acceptable to their interlocutor. Instead, they should attempt to use reasons acceptable to the determiner. In the case of a disputation modelling a legal proceeding, the determiner is a magistrate, and there are legal rules known to all of the parties that the magistrate applies instead of or in addition to her or his own judgement. Now, if we have a determiner who is very knowledgeable on topics relevant to the issue at hand (the issue which the initial claim is about) or who can become knowledgeable through research and by examining experts, then we can expect a judgement that predicts the truth-likeness of the case supporting the claim even when one of the proponent or opponent is less knowledgeable or less rhetorically skilled. The knowledgeable proponent, for example, attempts to use reasons they expect to be acceptable to the determiner (or acceptable given a set of rules, such as legal rules, that encapsulate prior good judgements based on knowledge and normative theory) rather than reasons they expect to be acceptable to a less knowledgeable opponent. A final judgement can be rendered even when proponent and opponent remain in disagreement at the end. The problem then becomes how to select such a good judge. Furthermore, if the opponent is not knowledgeable and/or skilled at argument, or if neither proponent nor opponent is, then even though a good judgement might still be possible at the end, the case made out might not provide an adequate public justification for the initial claim. In the judicial setting, a judicial opinion might substitute, but if the magistrate, though knowledgeable, is depending on the advocates to raise the relevant issues that should be considered when making a judgement, even the judicial opinion might turn out to be a poor justification. Ideally we need all participants to play a role in making the case as good as it can be.

4. A role for diversity

On many complicated issues, such as many issues of public policy, it might not be plausible for one or two persons to possess all of the relevant knowledge between them. In that case, we could try to gather

10 The McBurney, Hitchcock, and Parsons (MHP) dialogue model for deliberation recognizes this need by including a dialogue phase called "Inform", and Walton, Toniolo, and Norman (2014) describe deliberation dialogue models, such as the previously mentioned MHP model, as containing embedded information-seeking dialogues.

together several people who possess relevant knowledge. Then we could proceed as in the first style of dialogue, modified to admit more than two participants. Participants will take turns contributing proposals, positive or negative evaluations of proposals, or arguments for or against some aspect of the case(s) for or against some proposal. If there is no disagreement, the dialogue can be ended and the agreed upon proposal accepted; otherwise, the dialogue continues through the turns of each of the participants.[11] Suppose that one participant is an advocate for a particular proposal, to be called Participant 1 and Proposal 1, respectively, and three other participants have contributed opposing arguments (or negative evaluations). Participant 1 has two strategies for defending their proposal. Participant 1 may contribute an argument attacking (or shoring up their own case against) an opposing argument in terms acceptable to the participant who contributed that opposing argument, or Participant 1 may contribute an argument attacking (or shoring up against) opposing arguments by two or more participants, in terms acceptable to all of those participants.[12] Suppose that, just as when employing the previous two dialogue models, some dialogues under the model contain conflicting arguments, which we must compare if we are to determine which of the arguments ought to prevail. Now that we have multiple participants whose subjective judgements about the truth of the premises and the reliability of the inference warrants should be considered, we could aggregate those judgements using a voting rule. For example, a unanimous voting rule requires that for a new argument to succeed (at attack or at shoring up against opposing arguments) it be such that it causes all participants to update their degrees of belief in a claim to above the threshold for acceptance; a majority rule requires that the new argument thus impacts a majority of the participants. The other choice we need to make is whether, under a less than unanimity voting-rule, we need for the plurality of dissenting participants to accept the claim or only the plurality of all participants. The first option follows the intuition that we need to persuade our critics, while the second follows the intuition that the whole group's judgements are better than those of individual members or of subgroups. The second option is in line with the concerns of this paper; it makes best use of the pool of knowledge available in the group. In fact, instead of voting, we could simply average the values of the subjective terms in the update function: $P(C)$, $P_f(R)$, and any subjective terms in the model of the inference and its reliability. This is appropriate because we are interested not in what each individual would accept but what each ought to accept, which, under the assumption that the group is more likely to be correct than any of its individual members, is what the group (rationally) accepts. (Voting would be appropriate if we are not able to determine the member's

11 This procedure is similar to the procedure described by McBurney, Hitchcock, and Parsons. But McBurney *et al.* have a rule that before participants may assert a preference for one action option over another (i.e., may vote), evaluations must have been asserted for each of the action options (McBurney, Hitchcock, & Parsons, 2007, pp. 103–104). They write, "This combination rule ensures that preferences expressed in the dialogue are grounded in an evaluation of each action-option according to some proposed goal, constraint, or perspective, and thus contestable (McBurney et al., 2007, p. 104)." In other words, there is required to be a minimal public justification for the final expression of preferences.

12 These two strategies were experimented with by Gregor Betz in computer simulations and the results described in his book, *Debate Dynamics*. In both of these cases, participants eventually discover and agree upon an arbitrarily chosen position dubbed 'truth' unless their dialogue rule specifies that an established consensus cannot be further criticized and they first come to agree upon some other position, a non-truth consensus. A non-truth consensus is not necessarily a bad position; it is just not the best (since there is only one 'truth' in each simulation). In *Debate Dynamics*, only deductive arguments were modelled. (Betz offers a convincing argument that non-deductive arguments can be represented as deductive arguments with premises accounting for their inference warrants and exceptions/*ceteris paribus* conditions; however, if we model all arguments that way, we would not gain the benefit of quantifying our uncertainty, and the only available attacks would be ones that show logical inconsistencies between propositions a participant regards as true. For example, a conclusion could be denied by a participant on the basis an argument composed of only premises regarded as true by an opponent, causing the opponent to conclude that one or more of her or his own premises must have been false.) Betz's results suggest that we should be optimistic about the epistemic benefits of dialectical argumentation.

values for the terms in the update function. In that case, if we assume that they are rational, then the differences in their votes are due to differences in the values of the respective subjective terms they use when they update their beliefs. And the vote indirectly measures the group's degree of belief in the claim.)

So far, I have been claiming that the judgement of the group should be expected to be better than that of any individual member or of a subgroup, so now I will explain why. I base the claim on empirical findings of phenomena known as *wisdom of crowds* and on theoretical findings of *jury theorems* (List & Goodin, 2001; Surowiecki, 2004). In his book, *The Wisdom of Crowds*, Surowiecki begins with an anecdote of an early 20th century scientist, Francis Galton, statistically analyzing the guesses of fairgoers about the weight of an ox. The competition had been designed to award the one participant whose guess was closest to the correct weight, but as part of his statistical analysis, Galton obtained the mean of the participants' guesses. The average guess of 1,198 pounds was unexpectedly very close to the correct weight of 1,197 pounds (Surowiecki, 2004, Chapter 1).

Exemplary of this type of "wisdom of crowds", addressing problems Surowiecki calls "cognition problems", are prediction markets. Beginning in the latter 20th century prediction markets have since had remarkable success predicting many kinds of events. Participants in prediction markets place bets on some future event by buying and selling shares up until the time of the event. Prediction markets for predicting election outcomes outperformed opinion polls in the U.S. between 1988 and 2004, and they also had a tendency to outperform polls in Australia, Germany, and Canada (Bragues, 2009, p. 98). Prediction markets also performed well at predicting the outcomes of football matches, Hollywood box office receipts, and developments in science and technology (Bragues, 2009, p. 98). Presenting the *efficient markets hypothesis* that attempts to explain their success, Bragues writes:

> Only a segment of the total investor population need be rational, at least enough to quickly bring prices back into line when there is a deviation from correct values. And if we make the assumption that the mistakes of other investors follow the normal distribution seen in many natural and human phenomena, then the opinions of the non-rational will cancel each other out. (Bragues, 2009, p. 96)

The Condorcet Jury Theorem is a result proved using probability by the Marquis de Condorcet in 1785. Under the assumption that jurors make their decisions independently, and for a two option question with one correct answer, Condorcet showed that if all jurors are at least 50% likely to be correct, then the majority verdict will be correct with at least 50% probability, increasing with the size of the jury. Contemporary jury theorems have extended Condorcet's results (List & Goodin, 2001).

We might not be able to identify the set of people who have the most relevant knowledge regarding a particular issue and that between them have some knowledge that is relevant to all of those who might be affected by a final policy proposal. But the jury theorems and the efficient markets hypothesis give us good reason to believe that we can get away with having a large number of people who collectively know, even when the average participant is not especially knowledgeable. So, in that case, we can try to make sure we have a genuinely diverse group in terms of all factors we can identity that influence knowledge. We might try to select people with different educational backgrounds and different sets of academic or profession skills or trades; people from different locations and with different ancestry; people of different races or ethnicities; people of different age, gender; etc. If we expand our knowledge pool broadly enough, we can hope to find within it, the knowledge we seek.

5. Dialectical conduct of a truth-seeking dialogue

Earlier, candidates for truth-seeking deliberation dialogues were presented in the context of dialectical obligations that have been described for persuasion dialogues, but other rules have been proposed. From the perspective of this paper, what we would like are dialectical practices that tend to expose weak proposals and arguments as weak quickly and to shore up strong (or most promising) proposals and arguments in as robust ways as possible as early as possible. In other words, participants should be efficient in their deliberations so that by the time the discussion is ended, our prediction of what is the best proposal is plausibly as accurate as it could have been. Alvin Goldman has presented some such dialectical rules in his book, *Knowledge in a Social World* (1999, pp. 139–143, "Dialogical argumentation"). This paper will not take up that area of inquiry, but future research in that area, specific to practical reasoning and deliberation dialogue, is intended.

6. Publicly acceptable evidence

A public justification of a proposal for action cannot rest on premises (or warrants of inference) that are not acceptable to members of the public. An important question is how much of the public must find the justification acceptable. A unanimity rule would be ideal, but impractical. A single dissent could derail an agreement and become leverage for the dissenter to use to bargain for more of their preferred goods. (If the dissenter is deliberating in bad faith) such bargaining would subvert the goals of the deliberation. That is a question that will not be given further consideration here. Instead we will assume that some proportion of the whole can be found that is reasonably close to unanimity and/or that participants somehow can be held to a good faith participation standard and/or that evidence for acceptability of premises independent of self-report by participants can be found such that we can speak intelligibly about acceptability of premises to the whole of the public or of generally acceptable premises. Presumably, for a public justification to be acceptable, it will ultimately rest on generally acceptable premises (*endoxa*) or factual evidence from generally trusted sources, and inference will be supported by generally acceptable warrants of inference.

It might be the case, that in our small group of deliberators, we find that if we average the subjective terms a probabilistic model, our posterior probability that some proposal is better than all others passes some threshold, and we are therefore justified in believing it to be the case. But in the larger group of all members of the community, we might find that a majority (or other plurality) are unconvinced by the case that is supposed to justify the action.

This might happen if the community members ignore the case and make their judgement on the basis of their prior degrees of belief in the conclusion. This is contrary to our model of rationality. Let us suppose that is not the case. Instead anything that looks like that is only due to the community members having priors much further away from the conclusion drawn by the small group than members of the small group initially did. This means that the members of the small group have adduced enough evidence to justify the case amongst themselves but not enough to justify it to the larger community.

In order to obtain an acceptable public justification at the end, the acceptance of the proposal cannot be based too much on subjective degrees of beliefs of the (small-group) participants. While it might be rational to accept their judgement based on the evidence that the small-group participants were knowledgeable and that models such as the jury theorems predict that their consensus will come close to the truth, a justification such as that is too opaque to inspire confidence in a doubter. And we cannot readily assess the likelihood that any error will disproportionately accrue to the disadvantage of some members of the community rather than others. In fact, if the small group is actually insensitive to

such differences between the small group and the community as a whole, it is even possible that the small group, for that reason, misidentified the proposal that is best for the whole community. In other words, an expanded justification might produce evidence of error that would be evident to some members of the larger community.

We should like our public justification to depend less on confidence in the participants and the procedure than on confidence in the evidence (other than evidence of properties of the participants and the procedure). So the goal of the procedure is not only to be a good procedure for identifying truth-like conclusions, but also to produce independently justifying cases supporting those conclusions. And, therefore, small-group participants must, in many cases, go further in adducing evidence than is necessary to achieve consensus among themselves (especially if they are groups of experts). The participants should attempt to identify generally acceptable premises that can serve as their stopping points and should attempt to use inference warrants that can be explained in an acceptable way to the community at-large (even if the inference warrant itself is somewhat abstruse). The participants attempt to put themselves in the shoes of other members of the community, even those who are not present at the deliberations.

7. Conclusion

This paper has presented an approach to producing a stable prediction of whether an argument's conclusion will turn out to be true in virtue of its premises and its inference warrant. I do not claim that it is a best approach but rather that it is an example of approaches one could take to produce such a stable prediction. In summary, it involves selecting a competent and diverse group of dialogue participants; quantifying, as well as we can, our uncertainties about the truth of the premises and the reliability of the inference and using that information to determine argument defeat and to guide our search for better arguments; and adducing sufficient evidence to justify the conclusions to the whole community. It brings together content-oriented considerations pertaining to confidence in arguments and procedure-oriented considerations pertaining to the building of robust justifying cases in support of action or policy proposals. While limited detail has been given, the outline and reasons for the approach have been made clear. Inspiration has come from probabilistic approaches to measuring argument strength or quality and from dialectical approaches to collaboratively developing robust justifications.

There are many areas for further research before a practical implementation of an approach like that described here could be achieved. Probabilistic models are not necessarily easy to construct, and it is not trivial to elicit participants' degrees of belief in terms of probability values for each subjective term in their update functions. This is a potential objection to attempting to use probability. The approach described so far might seem impractical for that reason, but I believe that it holds promise toward addressing a key concern in argumentation: assessment of the quality of arguments and the conduct of dialogues, especially ones having the practical goal of determining justifiable best course of action.

References

Amgoud, L., & Cayrol, C. (1998). On acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI1998)* (pp. 1–6).

Amgoud, L., & Prade, H. (2004). Using arguments for making decisions: a possibilistic logic approach. In *Proceedings of the twentieth conference on uncertainty in artificial intelligence (UAI2004)* (pp. 10–15).

Bragues, G. (2009). Prediction markets: The practical and normative possibilities for the social production of knowledge. *Episteme*, *6*(1), 91–106.

Caminada, M. (2006). On the issue of reinstatement in argumentation. In M. Fisher, W. van der Hoek, B. Konev, & A. Lisitsa (Eds.), *Logics in Artificial Intelligence* (pp. 111–123). Liverpool, U.K.: Springer.

Godden, D., & Zenker, F. (2018). A probabilistic analysis of argument cogency. *Synthese*, *195*(4), 1715–1740. https://doi.org/10.1007/s11299-016-1299-2

Goldman, A. (1999). *Knowledge in a social world*. Oxford: Clarendon Press.

Gordon, T. (2018). Defining argument weighing functions. *Journal of Applied Logics -- IFCoLog Journal of Logics and Their Applications*, *5*(3), 747–773.

Hahn, U., & Oaksford, M. (2006a). A Bayesian approach to informal reasoning fallacies. *Synthese*, *152*, 207–236. https://doi.org/10.1007/s11229-005-5233-2

Hahn, U., & Oaksford, M. (2006b). A normative theory of argument strength. *Informal Logic*, *26*(1), 1–24.

Hahn, U., Oaksford, M., & Corner, A. (2005). Circular arguments, begging the question and the formalization of argument strength. In A. Russell, T. Honkela, K. Lagus, & M. Pöllä (Eds.), *Proceedings of AMKLC '05* (pp. 34–40). Espoo, Finland.

Hahn, U., Oaksford, M., & Harris, A. J. L. (2013). Testimony and argument: A Bayesian approach. In F. Zenker (Ed.), *Bayesian argumentation: The practical side of probability* (pp. 15–38). Dordrecht: Springer. http://doi:10.1007/978-94-007-5357-0_2

Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach* (3rd ed.). Peru, Illinois: Carus Publishing Company.

Korb, K. (2004). Bayesian informal logic and fallacy, *24*, 41–70.

Li, H., Oren, N., & Norman, T. J. (2011). Probabilistic argumentation frameworks. In S. Modgil, N. Oren, & F. Toni (Eds.), *Theory and Applications of Formal Argumentation* (pp. 1–16). Barcelona, Spain: Springer.

List, C., & Goodin, R. (2001). Epistemic democracy: Generalizing the Condorcet Jury Theory. *The Journal of Political Philosophy*, *9*(3), 277–306.

McBurney, P., Hitchcock, D., & Parsons, S. (2007). The eightfold way of deliberation dialogue. *International Journal of Intelligent Systems*, *22*, 95–132. https://doi.org/10.1002/int.20191

Oaksford, M., & Charter, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York: Oxford University Press.

Oaksford, M., & Hahn, U. (2013). Why are we convinced by the Ad Hominem Argument? In F. Zenker (Ed.), *Bayesian Argumentation* (pp. 39–58). Dordrecht: Springer.

Prakken, H. (2010). On the nature of argument schemes. In C. Reed & C. W. Tindale (Eds.), *Dialectics, dialogue and argumentation* (pp. 167–185). London: College Publications.

Rescher, N. (1977). *Dialectics: A controversy-oriented approach to the theory of knowledge*. Albany: State University of New York Press.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes, business, economies, societies, and nations*. New York: Doubleday.

Walton, D., Toniolo, A., & Norman, T. J. (2014). Missing phases of deliberation dialogue for Real Applications. *In Proceedings of the 11th International Workshop on Argumentation in Multi-Agent Systems*.