How Ambiguous Testimony Can Derail Group Deliberation

Leon Assaad (L.Assaad@campus.lmu.de)
Munich Center for Mathematical Philosophy, LMU Munich
Ludwigstraße 31, 80539 Munich, Germany

Abstract

Ambiguity is pervasive in verbal expressions of uncertainty: utterances such as "likely" are interpreted as different probabilities by different individuals, leading to systematic misunderstandings between speakers and hearers. What are the group-level effects of such misunderstandings? Can they lead deliberating collectives to inaccurate beliefs, or pronounced disagreements? To address these questions, we introduce a new Bayesian agent-based model of ambiguous testimony, designed to compare the effects of increasingly ambiguous communication against a baseline of ideal exchange. We find that increasing ambiguity reduces collective accuracy and increases belief dispersion. For mild ambiguity, denser communication networks mitigate these risks. Under strong ambiguity, however, increased communication can backfire: higher network density can exacerbate collective inaccuracy and polarization.

Keywords: Agent-Based Model; Disagreement; Bayesian Inference; Deliberation; Social Networks

Introduction

A jury is asked whether the defendant is guilty. A panel of experts discusses whether there will be an earthquake next week. An electorate debates whether a candidate should be prime minister. Group deliberation is ubiquitous: we engage in exchanges of arguments, evidence, or testimony, in order to come to a collective conclusion. Typically, the aims of deliberation are twofold: first, deliberation ought to reliably produce true beliefs (Goldman, 1999). Second, it should reduce disagreement and ideally foster consensus (Kitcher, 1993). Combining both goals, deliberation is successful when the group converges on a correct consensus.

But deliberation can fail: groups may reach incorrect consensus, or their beliefs drift apart and they polarize. Prominent theories trace failure back to behaviour which is (arguably) incompatible with epistemic rationality—such as motivated reasoning, cognitive biases, and the formation of secluded epistemic bubbles (Anderson, 2021; Kahan, 2013; Taber et al., 2009). By contrast, a growing body of research engages in rational reconstructions of deliberative failure: it explores whether wrong consensus or polarization can arise even when group members are rational—or at least not blatantly irrational (cf. Dorst, 2023; Kopecky, 2023; Singer et al., 2019).

In this vein, our paper investigates a plausible mechanism leading to deliberative failure: ambiguity inherent in ordinary discourse. Verbal expressions are notoriously underdetermined, and give rise to ambiguity: they allow for multiple admissible interpretations. A prominent case are verbal probability expressions (VPEs) such as "likely". Speakers often prefer them to precise numerical probabilities when communicating uncertainty—even in consequential domains such as medical diagnoses or climate science (Dhami & Mandel, 2022). These expressions are, however, inherently coarse-grained: the same expression is compatible with a range of underlying credences. Depending on the situation, both a 70% and a 95% degree of belief might be reported as "likely." This creates a level of ambiguity for hearers, who may, as a result, interpret the same utterance differently.

Indeed, decades of psychological research show that we have such idiosyncratic interpretations of common uncertainty expressions, prompting the worry of an "illusion of communication": we may misunderstand each other because we falsely assume that we uniformly interpret expressions of uncertainty (Budescu et al., 2014; Collins & Hahn, 2018).

What are the group-level effects of such misunderstandings? Suppose a jury deliberates on a suspect's guilt, but each juror has their own, slightly different interpretation of a positive/negative utterance (e.g., "I believe the suspect is guilty"). Can these individual-level misunderstandings accrue, and lead the group to inaccurate beliefs, or pronounced disagreement? And if so, under which conditions?

To address these questions, we introduce a new agent-based model (ABM) of deliberation via ambiguous testimony. Inspired by classic models from social epistemology (e.g., Hahn et al., 2024; Olsson, 2011), it represents agents who collect evidence, and exchange their beliefs about a target hypothesis. Agents are Bayesian, interpret their private evidence correctly and uniformly (cf. Assaad and Hahn, 2024), and—in an optimal baseline—communicate their fine-grained degrees of belief truthfully and without any loss of nuance. This provides an ideal baseline of deliberation in the absence of ambiguity.

To introduce ambiguity, we restrict communication to an increasingly coarse-grained vocabulary for expressing credences, so that multiple beliefs map to the same expressions. Crucially, we allow for interpretive idiosyncrasy: while agents share the same category boundaries, they may translate the same expression into different numerical probabilities within that category. Hence, coarser vocabularies make for more ambiguity, and more interpretive idiosyncrasy in our model.

We find that ambiguity and interpretive idiosyncrasy can be sufficient to generate substantial deliberative failure. As ambiguity increases, collective accuracy drops and belief dis-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

persion rises: more agents end up with incorrect beliefs, and the group splits into increasingly disagreeing camps. For mild ambiguity, denser communication networks mitigate these risks. Under strong ambiguity, however, more communication can backfire: higher network density may amplify misinterpretation and thereby exacerbate both collective inaccuracy and disagreement.

Hence, our findings suggest that ambiguity and interpretive idiosyncrasy—natural features of verbal communication—can be sufficient to derail deliberation, even in absence of blatant irrationality.

Ambiguous Expressions of Uncertainty

When prompted, a juror says "I am 90% certain the suspect is guilty." This understandable utterance matches a standard assumption in psychology and parts of philosophy: we entertain degrees of belief—or (un)certainty—in propositions, and they can be modelled in terms of subjective probabilities (Hartmann, 2021). When expressing uncertainty, however, people and institutions often opt for verbal expressions over numerical ones: the IPCC uses probability terms (e.g., "very unlikely"), accompanied by probabilistic interpretations, to communicate uncertainty to the public (Budescu et al., 2009). In institutional deliberation settings, expert participants are often prompted to express their judgments using coarse ordinal vocabularies, such as four-point Likert agreement–disagreement scales (Lazarus et al., 2022). Generally, research suggests that senders often have a preference to communicate uncertainty verbally, while receivers prefer numerical expressions (Dhami & Mandel, 2022; Wallsten et al., 1993).

This produces ambiguity for the receiver: coarse reports are compatible with a range of underlying degrees of belief. Suppose that speakers must express their confidence using a four-point agreement scale. Then both an agent with credence 0.80 and another with credence 0.95 might reasonably select the same category, e.g., "strongly agree". Consequently, for the hearer, the utterance underdetermines the speaker's credence: conditional on knowing the shared vocabulary, it is compatible with an interval of possible underlying beliefs. It is therefore unsurprising if different hearers translate the same message into different numerical probabilities.

Indeed, a large body of psychological research shows that VPEs are understood idiosyncratically: when faced with the same statement, hearers can—and often do—translate it into different numerical probabilities (Wintle et al., 2019). In classic studies, interpretations of "very likely" have ranged from roughly 45% to 99%, with similarly wide spreads for expressions such as "likely," "good chance," and "uncertain" (Lichtenstein & Newman, 1967). Such findings, among others, give rise to an "illusion of communication": speakers and hearers may mistakenly assume mutual understanding because they expect probability terms to be interpreted similarly across individuals (cf. Budescu et al., 2009, for a critical discussion see Collins and Hahn, 2018).[1]

Not only are VPEs idiosyncratically understood. Interpretations need not be symmetric. One might expect "antonym" pairs to function as probabilistic complements: if "likely" is interpreted as 75%, then "unlikely" should be 25%, so that the two sum to 100%. Empirically, however, this complementarity often fails. In studies where participants provide numerical translations of VPEs, the central tendencies (means or medians) associated with a term and its apparent opposite do not reliably add up to 100%. Lichtenstein and Newman (1967) find that "likely" and "unlikely" are asymmetric: their means sum to 72% + 18% = 90%. So in that pair, "likely" is the weaker term, while "unlikely" is the stronger one: "unlikely" is closer to 0% than "likely" is to 100%. Analogously, Willems et al. (2020) find that the Dutch translations of "certain" and "uncertain" are interpreted such that their central tendencies sum well above 100%, whereas "possible" and "impossible" sum to about 53% in the mean. This kind of asymmetry can, in part, be explained linguistically and pragmatically (e.g., "possible" vs. "impossible", see Mosteller and Youtz, 1990).

Relatedly, VPEs can also display directionality effects: some are positive (e.g., "probable"), others are negative (e.g., "not certain"). The choice of positive or negative expression draws attention to different outcomes, signals the speaker's intentions, and can thereby affect interpretation (Teigen & Brun, 1999, 2003). Such directionality effects can compound asymmetries: if some communicators systematically prefer positive over negative expressions, listeners may systematically over- or under-estimate their credences, systematically worsening misunderstandings.

When communication proceeds via verbal expressions of uncertainty, we risk interpreting them idiosyncratically, prompting natural worries about the effectiveness of communication. What are the effects of such ambiguity on the group-level outcomes of deliberation?

Deliberation and Testimony

Suppose a group of agents is deliberating whether a binary hypothesis H is true (H) or false ($\neg H$). Each agent starts with a degree of belief $P(H) \in [0, 1]$, informed by private evidence. Deliberation proceeds by communication on a social network, representing the group's social structure: agents exchange information with their neighbours and update their beliefs. In the ideal case, the group converges on a true belief ($P(H) \approx 1$ if H is true), yet deliberation can also fail: agents may converge on an incorrect belief, or beliefs may diverge.

This highly idealized sketch (or some close variant) forms the backbone of many models of deliberation from social epistemology. While models differ substantially in how they conceptualize communication and inquiry,[2] the tradition most important for our purposes is deliberation via testimony. In

[1] Studies also find substantial heterogeneity in people's vocabularies, i.e., which expressions they employ to communicate uncertainty (for reviews, see Budescu et al., 2014; Dhami and Mandel, 2022).

[2] As argument exchange (Borg et al., 2017), as averaging of credences (Hegselmann & Krause, 2002), or as the sharing of experimental results (Zollman, 2010).

everyday contexts, testimony takes the form of a report expressing one's confidence in a proposition—for instance, answering "Will it rain tomorrow?" with "I think so." In this sense, testimonial reports are signals about a speaker's credence in a hypothesis H, and, consequently, about H itself.

Formal epistemologists have developed Bayesian[3] ABMs of deliberation via testimony, prominently the *Laputa* model (Angere, 2010; Olsson, 2011). In *Laputa*, agents testify to communicate their stance on H (typically testifying only if they surpass a certainty threshold). Receivers update on said testimony using Bayesian source-reliability models, which specify their "trust" in a source (Bovens & Hartmann, 2003; Collins et al., 2018; Hahn et al., 2018): the more one believes the source to be reliable, the greater the likelihood ratio of its report, meaning a greater belief update in the direction of their testimony. These models have been used to study how network structure and communication frequency affect deliberative success, as well as failures such as polarization (Angere & Olsson, 2017; Pallavicini et al., 2021).

Bayesian models of testimony are typically restricted to binary messages (Hahn et al., 2024; Merdes et al., 2020). Depending on their credence, sources can only send a positive report (claiming "H is true"), a negative report (claiming "$\neg H$ is true"), or abstain. The interpretation of these reports is governed by the receiver's reliability estimate. This will not do for our purposes: we need a model that captures more or less ambiguous testimony, leading to more or less room for idiosyncratic interpretation. In the next section, we introduce a new Bayesian model of testimony that captures many degrees of message granularity, and, therefore ambiguity.

A Model of Ambiguous Deliberation

Our model is built on an ideal baseline.[4] Agents form their initial credences in the discussed hypothesis ($P(H) \in [0, 1]$) by personal inquiry. They inquire optimally, in the sense that they know their own reliability (i.e., the chances that they will get to the correct answer). In the baseline, they then communicate their degree of belief precisely: "My belief is $P(H)$." This testimony enables the listener not only to understand precisely which degree of belief the sender has, but also their reliability. Testimony has perfectly transmitted all relevant information.

Using this base, we introduce ambiguity: agents are restricted to a vocabulary of n expressions. For instance, $n = 4$ can be interpreted as agents using a "yes", "maybe yes", "maybe no", "no" vocabulary. We assume that senders choose signals via this simple rule: the credence interval $[0, 1]$ is split into n equally sized partitions, one for each expression. If the sender's credence falls within an expression's interval, then they send that expression—e.g., they will send "yes" (in the $n = 4$ case) if their degree of belief is greater than 75%.

We will (optimistically) assume that all agents are aware of this vocabulary, and the choice rule. This still creates the following ambiguity: they do not know where the sender lies in the interval. Our crucial assumption is idiosyncratic interpretation: each agent has their own numerical interpretation of each signal. Each interpretation falls within the designated interval, and is chosen at random: it is independently sampled for each agent from a uniform distribution. Hence, in the case of $n = 4$, one agent might interpret a "yes" as a signal of $P(H) = 0.8$, another as $P(H) = 0.95$.

This has two immediate consequences: first, as n decreases, interpretations will, in expectation, differ more strongly (for they are sampled from a greater interval). Second, since an agent's interpretation of each signal is independent, their interpretations will likely be asymmetrical: a "yes" may be interpreted as a 0.85, and a "no" as a 0.07.

The simulation proceeds as follows: agents are placed on a social network. Their initial belief is determined by personal inquiry. Then, they share one testimonial expression with each of their neighbours. These neighbours receive this testimony, and update their respective degrees of belief. For each simulation run, we hold fixed the same ground truth, network, and private evidence draws, and only change the communication channel; thus differences across n isolate the effect of ambiguous testimony. Decreasing from $n = 10$ to $n = 2$, we therefore run 10 cases in parallel; we will refer to each parallel case as a "group", and compare the groups' performances directly.

Before we explain our model in technical detail, let us note some idealizations, which should be relaxed in further work: first, the assumption that vocabularies are mapped by speakers to equally sized probability intervals does not reflect real VPE lexica (e.g., the IPCC's, see Budescu et al., 2009). Second, we assume that the interpretations of expressions remain stable for agents within a simulation; an agent will, for instance, invariably interpret a "yes" as 0.85, regardless of who utters it. While there is some empirical evidence suggesting consistency of interpretations (Budescu & Wallsten, 1985), changes in context can alter them. However, since our simulations will run for only very short time horizons (more below), we believe this to be a defensible idealization for now. Third, we assume that agents resolve ambiguity by choosing a precise credence, rather than entertaining imprecise probabilities (an avenue for future research). Lastly, and perhaps most impactful, is our assumption that agents share a fixed vocabulary, and know of the credence interval producing different expressions. This, we take to be an optimistic assumption, and conjecture that overlapping intervals, and differing vocabularies, will make communication harder, and deliberation less effective.

Ideal Baseline

The ideal baseline is inspired by the Bayesian NormAN framework (short for Normative Argument Exchange across Networks, Assaad et al., 2023, 2025). The state of the target

[3]Bayesian epistemology posits that rational agents' beliefs in propositions behave like probabilities and should be updated via some version of Bayes' rule (Sprenger & Hartmann, 2019).

[4]Our model code and simulation data are publicly available on The Open Science Framework (OSF) at https://osf.io/2pmau/overview?view_only=d8abe8bdea524e649802c149c107b2a3.

hypothesis H is binary: either H or $\neg H$. We assume a base rate of 50%: in about half of the simulations H is true. It is the agents' task to find this out.

To inquire, an agent a performs a private binary "test" $\mathrm{E}_a \in \{E_a, \neg E_a\}$. These tests are conditionally independent given H (cf. Olsson, 2013), and how informative one's test is depends on its likelihood ratio $LHR_{E_a} := \frac{P(E_a|H)}{P(E_a|\neg H)}$. If $LHR_{E_a} = 1$, the test is non-diagnostic; the further away from 1, the more informative it is. Each agent is initialized with an LHR, which determines the true positive and false positive rates of their test.[5] Informative tests, i.e., LHRs far away from 1, will likely lead to a truth-conducive result: if $P(E_a|H) = 0.9$, and H holds, there is a 90% chance one's test will be positive. We assume that agents are "calibrated" (Assaad & Hahn, 2024): they know their own LHR_{E_a} and also the base rate of the hypothesis. Hence, they update from the common prior $P(H) = 0.5$ by Bayes' rule, in an optimal fashion. In this sense, an agent's LHR is their reliability: the stronger the LHR, the more likely their test is correct; and since they know of their own diagnosticity, the greater the belief update $P(H \mid E_a)$. Hence, each test is, in expectation, truth-conducive.

How reliable the group is is a free parameter of our model. Each LHR is drawn from a modified normal distribution with a mean of 1.[6] The greater the standard deviation SD of the distribution, the more extreme the LHR values, and the more reliable the group. In the baseline, agents communicate their credence precisely via a continuous report $\mathrm{REP}_a^{(\mathrm{ideal})} \in [0, 1]$, where $\mathrm{REP}_a^{(\mathrm{ideal})} = x$ means "my degree of belief in H is x." Since agents share the same prior and are calibrated, receiving $\mathrm{REP}_a^{(\mathrm{ideal})} = x$ lets the hearer infer the reliability of the sender's private evidence: there is a unique likelihood ratio LHR which, when combined with the known prior, yields posterior x.[7] Thus, under ideal communication, a sender's report transmits exactly the diagnostic impact of their private evidence. Nothing is lost in communication. Lastly, hearers assume that each testimonial report is independent (see Olsson, 2011). Formally, an agent a's belief is computed as: $P_a^{(\mathrm{ideal})}(H|\mathrm{E}_a, \mathrm{REP}_1^{(\mathrm{ideal})}, \ldots, \mathrm{REP}_k^{(\mathrm{ideal})})$, if they have heard from k neighbours.

More or Less Ambiguous Testimony

To make testimony ambiguous, we restrict agents to a shared vocabulary of n messages. Formally, a report is now: $\mathrm{REP}_a^{(n)} \in \{s_1^{(n)}, s_2^{(n)}, \ldots, s_n^{(n)}\}$, where the latter is the set of expressions. Let each $s_i^{(n)}$ correspond to a partition of the unit interval: $[0, 1] = I_1^{(n)} \cup I_2^{(n)} \cup \cdots \cup I_n^{(n)}$, where each $I_i^{(n)}$ is an interval of width $1/n$. The sender's choice function is simply: if $P(H) \in I_i^{(n)}$, send $s_i^{(n)}$. Hearers idiosyncratically interpret these signals: for each signal $\{s_1^{(n)}, \ldots, s_n^{(n)}\}$, hearer a has their own interpretation $\{j_{a,1}^{(n)}, \ldots, j_{a,n}^{(n)}\}$. Each signal is being interpreted as a credence within the interval $j_{a,i}^{(n)} \in I_i^{(n)}$, sampled uniformly at the start of each simulation. This interpretation fixes the LHR of the report, and therefore the update $P(H \mid \mathrm{REP} = s_i^{(n)})$ of the receiver.[8]

We run 10 levels of ambiguity in parallel: based on the same set of evidence, agents communicate via the ideal baseline, and via 9 channels, from $n = 2$ to $n = 10$. Hence, each agent a entertains 10 parallel beliefs: $P_a^{ideal}(H)$, $P_a^{n=10}(H)$, $\ldots$, $P_a^{n=2}(H)$. For agent a, these beliefs are computed as: $P_a^{(\mathrm{n})}(H|\mathrm{E}_a, \mathrm{REP}_1^{(n)}, \ldots, \mathrm{REP}_k^{(n)})$.

Social Network

A group of $N = 100$ agents is connected on a small-world network from Watts and Strogatz, 1998, mirroring earlier NorMAN setups (Assaad & Hahn, 2024; Assaad et al., 2025). By adjusting the mean neighbourhood size through varying k (and setting the rewiring-probability to 0.1), we transition from a sparse, to a nearly completely connected network.

Simulation Runs and Assessment

Each simulation initializes whether H is true, places agents on the network, initializes the agents' reliabilities, and stochastically generates their results. Each agent has an initial belief $P(H \mid E_a)$. Then, they testify exactly once to each of their neighbours. Our choice of single testimony serves to avoid known problems of dependence and double counting in classical testimony models (Pilditch et al., 2025). In our case, since each test is independent given H, each piece of testimony is as well, making the agents' independence assumption true.

We vary the following parameters: First, the number of neighbours, where every agent has $2k$ neighbours ($k \in \{1, 2, 3, 5, 10, 25, 49\}$). Second, we change the groups' reliability, by changing SD from 1 to 5 in steps of 1. Simulating all factorial combinations, we ran every setting 1000 times, making for a total of 35000 runs. To assess deliberative success, we track whether each of our 10 groups is accurate, and whether they have come to consensus. For accuracy, we use the widely used Brier score (e.g., Huang, 2023): an agent's Brier score is $(P(H) - V(H))^2$, where $V(H) = 1$ if the hypothesis is true and 0 if false. We use the mean Brier score as a measure of group accuracy; it ranges from 0, perfectly correct consensus (if H, then all are at $P(H) = 1$), to 1, perfectly wrong consensus, and is at 0.25 if all agents

[5] Following the "Symmetry" assumption from Olsson, 2011 and Angere, 2010, we assume that $P(E \mid H) = P(\neg E \mid \neg H)$, i.e., a test's true positive and true negative rates are equal.

[6] We use a normal distribution with a mean of 1, which is truncated at 1—it is "half" a normal distribution. With a 50% chance, a drawn value x is flipped to $1/x$. This makes for a roughly symmetric evidence distribution: sometimes positive evidence is pro-H ($LHR > 1$), sometimes positive evidence is anti-H ($LHR < 1$), with equal probability, and the strength is mirrored (LHR vs $1/LHR$).

[7] Using Bayes' theorem in the form $P(H \mid \mathrm{E}) = \frac{P(H)}{P(H)+(1-P(H))\,LHR_{\mathrm{E}}^{-1}}$, the receiver can solve for the implied likelihood ratio: $LHR_{(\mathrm{REP}=x)} = \frac{x(1-P(H))}{P(H)(1-x)}$.

[8] We add a failsafe: no signal is interpreted as exactly $P(H) = 1$ or $P(H) = 0$. Interpretations are clamped to 0.99 and 0.01 respectively, preventing infinite likelihood ratios.

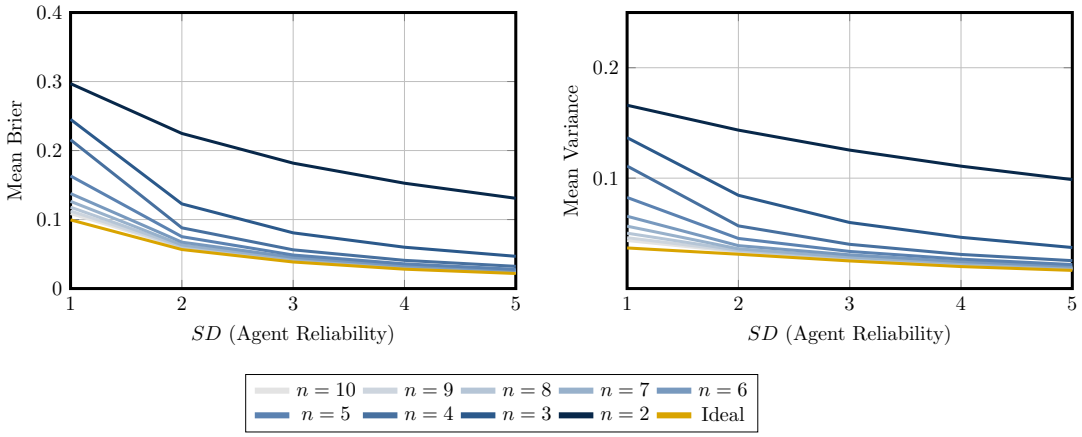

Figure 1: Performance vs. reliability. Mean group accuracy (Brier score, left) and belief dispersion (variance, right) as a function of evidence reliability (SD), averaged over all network connectivities k. Lines show the mean across simulation runs. Colours denote communication channels, from the ideal baseline (yellow) to the coarsest $n = 2$ vocabulary (dark).

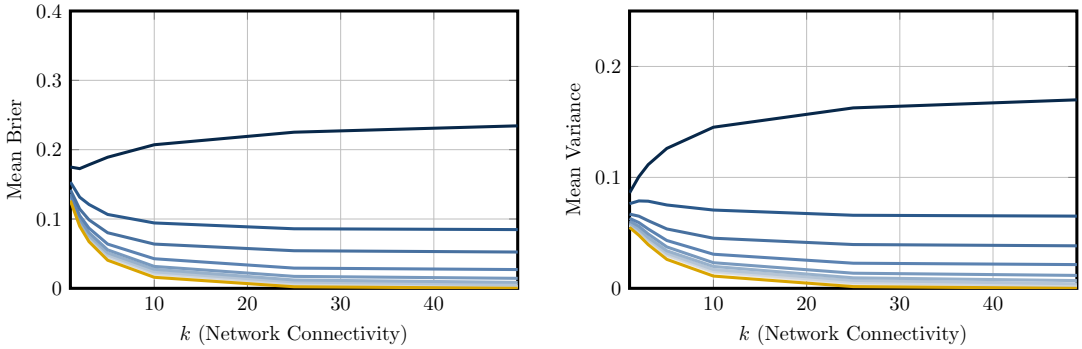

Figure 2: Performance vs. connectivity. Mean group accuracy (Brier; left) and belief dispersion (variance; right) as a function of network connectivity k, averaged over all reliabilities. Colours as in Fig. 1.

believe $P(H) = 0.5$ (the prior). As a measure of consensus/dispersion, we use simple statistical variance of beliefs (following Bramson et al., 2017). Variance of beliefs ranges between 0, where all agents are in perfect consensus, to 0.25, which means maximal bi-polarization: half of agents believe $P(H) = 1$, the others believe $P(H) = 0$.

Results

Overall, greater ambiguity makes deliberation less successful: smaller vocabularies (lower n) reduce group accuracy (higher Brier) and increase belief dispersion (higher variance). While we see no (systematic) convergence to incorrect consensus, we find that this effect is present across the parameter ranges we explore: in expectation, smaller n vocabularies both increase inaccuracy and variance, with the dichotomous "yes"/"no" case ($n = 2$) faring particularly badly.

Unsurprisingly, all groups do better when agents are more reliable: Fig. 1 shows performance as a function of reliability (averaging over all networks). Very ambiguous groups do considerably worse in terms of variance and accuracy, but as reliability increases, all groups approach consensus and perfect accuracy, thus moving closer to the ideal baseline.

The impact of connectivity is more complex: Fig. 2 shows performance as a function of connectivity k, averaged over all reliabilities. For the ideal baseline, more connectivity is better: both variance and inaccuracy decrease, and at high connectivity the group reaches near-perfect consensus. Low levels of ambiguity follow suit. However, for more ambiguous groups ($n = 3, 4, 5$), the beneficial effect of greater connectivity is pronounced at first, but quickly plateaus. For $n = 2$, the most ambiguous group, variance and Brier increase with more connections: more communication worsens deliberation.

This negative effect of rising k interacts with reliability. Fig. 3 shows simulations where we explored a range of very low reliabilities, $SD \in [0.25, 0.5, 0.75, 1]$ (we ran each setting 1000 times over all network densities, for a total of 28000 runs). The ideal group still converges to true consensus as it becomes better connected: pooling the collective evidence, if interpreted correctly, is sufficient to come to the truth. However, under unreliability, all groups become more dispersed as they get more connected. Groups $n = 2, 3, 4, 5$ also become markedly less accurate. Combining an unreliable evidence base with ambiguous testimony thus produces a negative network effect: more communication worsens deliberation.

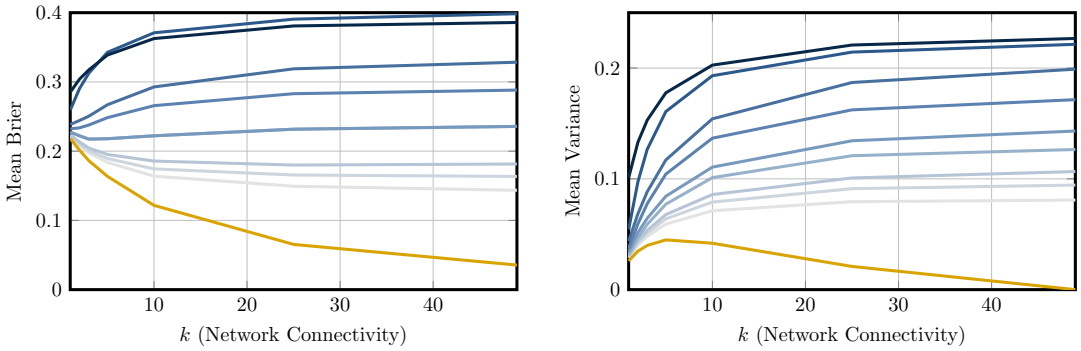

Figure 3: Performance vs. connectivity under low reliability ($SD \in [0.25, 0.5, 0.75, 1]$), averaged over SD values. Mean group accuracy (Brier; left) and belief dispersion (variance; right) as a function of network connectivity k. Colours as in Fig. 1.

Explaining the Effects: Bias via Asymmetry

Why does more connectivity prove bad under ambiguity? A key factor is asymmetry. Recall that our model allows for asymmetry: interpretations of complementary expressions need not be probabilistic complements. This can systematically bias an agent's stream of evidence. As an illustration, consider the contrast between the ideal baseline and the simplest case of ambiguity: $n = 2$, a "yes"/"no" or "likely"/"unlikely" vocabulary. Suppose that H is true and that agent a receives four ideal reports, expressing credences of 90%, 80%, 30%, and 40% respectively. In the ideal, this yields a posterior of $P_a^{\text{ideal}}(H) \approx 0.911$. Now suppose $n = 2$. The four reports become "yes", "yes", "no", "no". If the receiver interprets "yes" as 72% and "no" as 18% (mirroring the mean interpretations of "likely"/"unlikely" in Lichtenstein and Newman, 1967), then their posterior is $P_a^{(n=2)}(H) \approx 0.242$. Hence, we see pronounced divergence between the ideal baseline and the coarse $n = 2$ belief. In this case, asymmetry has biased the agent against the hypothesis.

Bias via asymmetry can make a receiver's testimony stream inaccurate, when it would be truth-conducive under correct interpretation. The example above shows that, even though the receiver has heard from four neighbours, their degree of belief has moved away from $P(H) = 1$. This helps explain why more neighbours, i.e., greater connectivity, can be detrimental for accuracy. Bias also helps explain why greater connectivity disperses groups of ambiguous communicators: if one agent is biased towards the hypothesis and another against, then their interpreted testimony streams can drive their beliefs apart.

Why does this negative effect of connectivity appear most clearly when the group is unreliable (Fig. 3)? When agents are unreliable, they are likely to receive weak and contradictory signals: their beliefs will be closer to the $P(H) = 0.5$ base rate, with many being on the incorrect side of the base rate. Hence, agents will receive testimony pointing towards and against H—though correct signals will, in expectation, be stronger (see Hahn, 2025). In the ideal baseline, receiving agents correctly interpret this nuance. However, under ambiguity and asymmetry, even weak incorrect signals can be exacerbated ("over-weighted"), drawing the receiver towards wrong beliefs. By contrast, when agents are reliable, most signals point towards the correct hypothesis, and there are only few misleading signals that can be exacerbated by asymmetry.

To show that it is really asymmetry, not merely idiosyncrasy, which drives this effect, we reran the same simulations imposing symmetry. While every agent had their own idiosyncratic interpretation, they were symmetric: interpretations of complementary expressions summed to exactly 100%. In this case, more connectivity is better for all groups. The simulation data can be found on OSF (link in footnote 4).

Conclusion

Verbal expressions of uncertainty are often ambiguous and interpreted idiosyncratically, creating individual-level misunderstandings. Our simulations suggest that such misunderstandings can significantly worsen group deliberation: greater ambiguity decreases accuracy and increases disagreements, even when agents are otherwise optimal, unbiased reasoners. When agents are reliable, increased communication mitigates the risks of ambiguity. The detrimental effect of more communication under ambiguity appears most clearly when agents are unreliable: in that case, asymmetries in interpretation can bias the evidence stream, so that greater connectivity amplifies misinterpretation.

Our simulations identify a mechanism reminiscent of other "less-is-more" results from the formal literature on testimony (Hahn et al., 2024; Olsson, 2011). In contrast to those models, our agents never become mistrustful of others, nor do they segregate into subgroups. Instead, deliberation fails because agents resolve realistic ambiguity in a suboptimal way.

The tentative upshot of our model is this: to improve deliberation, it may be important to ensure not only that uncertainty expressions are uniformly understood via reducing ambiguity (e.g., by adding numerical translations; Budescu et al., 2009). It may also be important to ensure interpretive symmetry, in the sense that complementary expressions function as probabilistic complements. Asymmetry is a way of smuggling in biases towards a conclusion, risking undermining the benefits of deliberation for the individual and the collective.

Acknowledgments

I gratefully acknowledge funding from the Konrad-Adenauer-Stiftung through a PhD scholarship and support from the Chair of Philosophy of Science at the Munich Center for Mathematical Philosophy, LMU Munich. I am very thankful to Kevin Zollman for valuable discussions that helped shape this project. I thank Ulrike Hahn, Stephan Hartmann, and Alexander Reutlinger for their invaluable advice and support.

References

Anderson, E. (2021). Epistemic bubbles and authoritarian politics. *Political epistemology*, 11–30.

Angere, S. (2010). Knowledge in a social network. *Synthese*, 167–203.

Angere, S., & Olsson, E. J. (2017). Publish late, publish rarely!: Network density and group performance in scientific communication. In *Scientific collaboration and collective knowledge* (pp. 34–62). Oxford University Press.

Assaad, L., Fuchs, R., Jalalimanesh, A., Phillips, K., Schöppl, K., & Hahn, U. (2023). A bayesian agent-based framework for argument exchange across networks. *arXiv preprint arXiv:2311.09254*.

Assaad, L., Fuchs, R., Phillips, K., Schöppl, K., & Hahn, U. (2025). Capturing argument in agent-based models. *Topoi*.

Assaad, L., & Hahn, U. (2024). Rational polarization: Sharing only one's best evidence can lead to group polarization. *Proceedings of the annual meeting of the cognitive science society*, *46*.

Borg, A., Frey, D., Šešelja, D., & Straßer, C. (2017). Examining network effects in an argumentative agent-based model of scientific inquiry. *International workshop on logic, rationality and interaction*, 391–406.

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology*. Oxford University Press.

Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., & Holman, B. (2017). Understanding polarization: Meanings, measures, and model evaluation. *Philosophy of science*, *84*(1), 115–159.

Budescu, D. V., Broomell, S., & Por, H.-H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological science*, *20*(3), 299–308.

Budescu, D. V., Por, H.-H., Broomell, S. B., & Smithson, M. (2014). The interpretation of ipcc probabilistic statements around the world. *Nature Climate Change*, *4*(6), 508–512.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational behavior and human decision processes*, *36*(3), 391–405.

Collins, P. J., & Hahn, U. (2018). Communicating and reasoning with verbal probability expressions. In *Psychology of learning and motivation* (pp. 67–105, Vol. 69). Elsevier.

Collins, P. J., Hahn, U., Von Gerber, Y., & Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content. *Frontiers in psychology*, *9*, 18.

Dhami, M. K., & Mandel, D. R. (2022). Communicating uncertainty using words and numbers. *Trends in Cognitive Sciences*, *26*(6), 514–526.

Dorst, K. (2023). Rational polarization. *Philosophical Review*, *132*(3), 355–458.

Goldman, A. I. (1999). *Knowledge in a social world*. Oxford University Press.

Hahn, U. (2025). Confirmation bias and the plausible distribution of evidence. *Mind & Society*, 1–18.

Hahn, U., Merdes, C., & von Sydow, M. (2018). How good is your evidence and how would you know? *Topics in Cognitive Science*, *10*(4), 660–678.

Hahn, U., Merdes, C., & von Sydow, M. (2024). Knowledge through social networks: Accuracy, error, and polarisation. *Plos one*, *19*(1), e0294815.

Hartmann, S. (2021). Bayes nets and rationality. *The Handbook of Rationality*, 253.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, *5*(3).

Huang, A. C. (2023). Track records: A cautionary tale. *The British Journal for the Philosophy of Science*.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision making*, *8*(4), 407–424.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford University Press, USA.

Kopecky, F. (2023). Argumentation-induced rational issue polarisation. *Philosophical Studies*, 1–25.

Lazarus, J. V., Romero, D., Kopka, C. J., Karim, S. A., Abu-Raddad, L. J., Almeida, G., Baptista-Leite, R., Barocas, J. A., Barreto, M. L., Bar-Yam, Y., et al. (2022). A multinational delphi consensus to end the covid-19 public health threat. *Nature*, *611*(7935), 332–345.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic science*, *9*(10), 563–564.

Merdes, C., Von Sydow, M., & Hahn, U. (2020). Formal models of source reliability. *Synthese*, 1–29.

Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical science*, 2–12.

Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme*, *8*(2), 127–143.

Olsson, E. J. (2013). A bayesian simulation model of group deliberation and polarization. *Bayesian argumentation: The practical side of probability*, 113–133.

Pallavicini, J., Hallsson, B., & Kappel, K. (2021). Polarization in groups of bayesian agents. *Synthese*, *198*, 1–55.

Pilditch, T. D., Hahn, U., & Lagnado, D. (2025). The problem of dependency. *Synthese*, *205*(4), 143.

Singer, D. J., Bramson, A., Grim, P., Holman, B., Jung, J., Kovaka, K., Ranginani, A., & Berger, W. J. (2019). Rational social and political polarization. *Philosophical Studies*, *176*, 2243–2267.

Sprenger, J., & Hartmann, S. (2019). *Bayesian philosophy of science*. Oxford University Press.

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, *31*, 137–155.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational behavior and human decision processes*, *80*(2), 155–190.

Teigen, K. H., & Brun, W. (2003). Verbal probabilities: A question of frame? *Journal of Behavioral Decision Making*, *16*(1), 53–72.

Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the psychonomic society*, *31*(2), 135–138.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, *393*(6684), 440.

Willems, S., Albers, C., & Smeets, I. (2020). Variability in the interpretation of probability phrases used in dutch news articles—a risk for miscommunication. *Journal of Science Communication*, *19*(2), A03.

Wintle, B. C., Fraser, H., Wills, B. C., Nicholson, A. E., & Fidler, F. (2019). Verbal probabilities: Very likely to be somewhat more confusing than numbers. *PLoS One*, *14*(4), e0213522.

Zollman, K. J. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, *72*(1), 17–35.