Individuals, Collectives, and Individuals in Collectives: The In-eliminable Role of Dependence

Ulrike Hahn

Correspondence:
Ulrike Hahn, Dept. of Psychological Science,
Birkbeck College, University of London
Malet Street, WC1E 7HX, London, U.K.
u.hahn@bbk.ac.uk

Abstract

Our beliefs are inextricably shaped through communication with others. Furthermore, even conversation we conduct in pairs, may itself be taking place across a wider, connected, social network. Our communication, and with that our thoughts, are consequently typically those of individuals in collectives. This has fundamental consequences with respect to how these beliefs are shaped. This paper examines the role of dependence on our beliefs and seeks to demonstrate its importance with respect to key phenomena involving collectives that have been taken to indicate irrationality. The paper argues that (with the benefit of hindsight) these phenomena no longer seem surprising when one considers the multiple dependencies that govern information acquisition and evaluation of cognitive agents in their normal, that is, social context.

Keywords: social networks, dependence, polarization, rationality.

Introduction

Much of our thinking about the nature and accuracy of our beliefs and opinions has, historically, focussed on individuals. An intellectual tradition stretching back to classical Greek philosophy has acknowledged that arguments and evidence frequently emerge in a dialectical exchange, between different parties. However, it has viewed those parties themselves as little more than as "argument dispensers". In particular, it was long held that only the arguments themselves mattered, and that consideration of the argument *source* is fallacious (see in the traditional catalogue of fallacies ad hominem argument, the appeal to authority, or the appeal to popular opinion, e.g., Woods, Irvine, and Walton, 2004). In keeping with this, philosophical and psychological interest in testimony has deepened only fairly recently, despite the fact that the majority of what we believe to know as individuals we know (partly or wholly) through the testimony of others (see e.g., Coady, 1992; Sporer, 1982).

Even where argument sources have come into view, the implicit research focus has largely been on a real or imagined single other from whom a cognitive agent receives information (e.g., Petty & Cacciopo, 1986; Hahn et al., 2009). This is reflected, for example, within argumentation theory in the near total absence of research on polylogues (argumentative exchanges involving many parties); and where polylogues have been considered they have largely been seen only through the lens of dialogues (see also Lewiński, & Aakhus, 2014). Recent shifts in research interest, however, have started to redress that balance: the resurgent interest in 'wisdom of crowds' effects; research on social epistemology within philosophy; a huge surge of interest in topics such polarisation and the spread of misinformation, and information and opinion dynamics more generally. These shifts have been fuelled further by new possibilities afforded by very large data sets of online communication (e.g., Colleoni et al., 2014).

Despite this gradual reorientation, a bias, arguably, lingers in the consideration of human communication and thought, particularly with respect to rationality. This lingering focus on, at best, single sources has arguably hindered the development of a deeper understanding of a range of phenomena. What consideration of social networks brings to the fore is that even those exchanges that ostensibly take place within a dyad, are themselves typically *embedded in wider networks of information exchange*. Communication within a collective is consequently not limited to contexts where everyone is simultaneously present. It is not the exception, it is the rule --even where the wider collective is not in focus. A failure to appreciate this has obscured the fundamental feature of communication within and across collectives: dependence.

Echoing other recent calls for figure-ground reversals in the study of cognition (Dingemansea et al., 2023), this paper seeks to draw out both the nature of that dependence and its implications. To this end, the paper seeks to demonstrate how a range of interconnected phenomena associated with belief and opinion dynamics that have attracted enduring interest as puzzling and surprising, are, arguably, anything but. Instead, our surprise reflects the extent to which one of the most fundamental determinants of how our beliefs are established has, in the past, been overlooked.

To develop the arguments of this paper, we will consider a simple model of communication across social networks by simple 'rational' Bayesian agents. Hahn, Hansen and Olsson (2020) used agent-based simulations with this model to demonstrate the role of the structure (topology) of communication networks on the accuracy of agents' beliefs. In this model, first proposed by Olsson (2011), there is a single claim at issue. Agents stochastically receive evidence from the world. They also receive testimonial evidence from agents to whom they are connected within the network. All evidence (whether from the world or from testimony) simply consists of an assertion that the claim is true or false. Each agent combines all evidence it receives via Bayes' Rule. The agent's current degree of belief in the truth or falsity of the claim determines what an agent communicates: if the agent's current degree of belief exceeds a certain threshold, the agent communicates that the claim is true (false), otherwise the agent stays silent. Full details of model and simulation can be found in Hahn et al. (2020), but no further details are required to appreciate the points made throughout the present paper.

Mirroring behavioural demonstrations of the impact of network structure on accuracy in behavioural experiments (e.g., Joensson et al., 2015; Becker et al., 2017), Hahn et al's. (2020) simulations, too, found effects of structure on accuracy. Such effects of network structure were detectable even in networks that contain both the same number of nodes and the same number of links. Fig. 1. takes data from Hahn et al.'s study and displays it in a slightly different way. Fig. 1 shows the individual level, the collective level and shows how both are affected by the topology of the network. Specifically, it shows the average individual error and the collective error, conceptualised as the accuracy of the group mean. Both types of error are affected by network structure, but in different ways. In fact, both levels are mathematically connected given a third quantity: the variance (diversity) in the individual judgments. Communication across a wider social network impacts both individual agent's accuracy and how similar their beliefs through information exchange. Both of these, in turn, jointly determine collective accuracy.[1]

Information exchange across a social network happens across the direct and indirect paths linking the agents within the network. It is the information flow across those paths that creates *dependence* between members of the network, and that dependence is integral to the ultimate outcomes of the communication, affecting both individual accuracy, and diversity across agents, and with that collective accuracy.

This makes accuracy vary systematically as a function of structure and makes the impact of communication different for the individuals and the collective as a whole (see Fig. 1). A given individual, with the same properties, the same information from the world, the same number of 'neighbours' she (directly) communicates with, and the same amount of information from them, will still end up with differences in the

[1] Specifically, these are linked via the so-called diversity prediction theorem (Page, 2010). When error, as here, is measured by the squared deviation from the truth, the collective error equals the average individual error minus the diversity, measured as the variance. For this and other formal frameworks that establish the same general functional relationships between individual accuracy, collective accuracy, and agent inter-relations see also, Hahn, 2022.

accuracy of her beliefs, depending on where in the network she resides (see Hahn et al., 2020, Fig. 7 for graphical demonstrations of this point). Individuals are thus inescapably individuals *in* collectives.

It is the key contention of this paper that a failure to properly consider this has fostered a caricature of our mental processes, because it has failed to understand the central role of informational dependence that communication networks involve. To redress this, the paper seeks to disentangle multiple types of dependence integral to belief formation across networks and to demonstrate how and why we cannot understand either individual or collective cognition without their appreciation. To make the case for the importance of these different types of dependence, the paper seeks to show how they render unsurprising three, inter-related polarization phenomena that have long exercised researchers. We introduce these three phenomena next.

Polarization: Shifts to Extremity, Belief Divergence, and Biased Assimilation

The term 'polarization' is many-faceted and at least nine different, interrelated meanings of the term have been distinguished in the literature (see Bramson et al., 2016). This paper is concerned with three core phenomena from the psychological literature: so-called shifts to extremity within deliberating groups, the fractioning of groups into increasingly divergent opinions, and, finally, individuals' biased assimilation of evidence. All three are widely referred to as 'polarization', but to avoid confusion we will use the more specific labels 'shift to extremity', 'belief divergence', and 'biased assimilation' in the remainder. We describe each of these in turn.

The often observed shift to extremity in deliberating groups constitutes the original 'polarization' phenomenon within the literature (see also Hahn et al., subm.). Group polarization in this sense occurs "when an initial tendency of individual group members toward a given direction is enhanced following group discussion" (Isenberg, 1986, p. 1141). The phenomenon was first introduced in the literature on "risky shifts" in decision-making (Stoner, 1968). A wealth of subsequent research (reviewed for example in, Myers & Lamm, 1976, or, Isenberg, 1986) confirmed that groups frequently came to consensus views (beliefs or attitudes) that were more extreme than the individual group members' pre-deliberation opinions. Although the size of the shift varies (e.g., large effects, see e.g., Luskin et al. 2002; Goodin & Niemeyer 2003; small effects, Merkle 1996), not just between studies but also by topic of discourse, by measure of attitude (e.g., self-report or direct observation, Miller et al., 1993), or depending on whether it is the aggregate, collective opinion or the average individual opinion that is being considered (Miller et al., 1993; Lindell et al., 2017; paralleling Fig. 1 above), the phenomenon itself is viewed as one of the "most robust patterns found in deliberating bodies" (Sunstein, 2002, pg. 177). And it has been observed with a wide range of methods, including not just lab-based studies (e.g., Myers, 1975) but also within political science using citizen debates (Lindell et al., 2017), deliberative polling, (e.g., Luskin et al., 2002) and 'citizens' juries', (Goodin & Niemeyer, 2003). Already in 1978, Lamm and Myers concluded that "seldom in the history of social psychology has a nonobvious phenomenon been so firmly grounded in data from across a variety of cultures and dependent measures" (pg. 146).

The two main explanations for shifts to extremity are not mutually exclusive (though they make different predictions that have been pursued experimentallly, see e.g., Vinokur & Burnstein, 1978): social comparison processes (e.g., Sanders & Baron, 1977) and "persuasive argumentation" (Burnstein & Vinokur, 1977). The social comparison explanation rests on the idea that humans are motivated to perceive and present themselves in a socially desirable light. As a result, they adjust their self-presentation in response to the self-presentation of others. Publicly expressed views are thus a combination of privately held opinion and beliefs about the views of others. Individuals may initially under-estimate the 'true norm' in a group ("pluralistic ignorance"). Upon exposure to others' views, group members more readily reveal their own true beliefs, shifting the group average as a result. Individuals may also adapt their opinions due to "bandwagon" effects (see Isenberg, 1986). Brown et al. (2022) provide a recent computational exploration of this general idea using agent-based modelling. The persuasive argumentation explanation, by contrast, assumes that group discussion causes shifts simply because it exposes individuals to persuasive arguments that favour the direction in which opinion then polarises. Such an account has likewise been explored in agent-based models by Maes and Flache (2013).

The shift to extremity has been seen to fuel extremism and increase social conflict (e.g., Lamm & Myers, 1978; Schkade et al., 2000; Sunstein, 2009; Broncano-Berrocal & Carter, 2021). And it has typically been accompanied by a sense that the final opinion is not wholly justified. Hence the search for explanations has typically been a search for explanations of 'deliberative failure' (Sunstein, 2006), and has invoked both cognitive biases and other epistemic 'vices' (Broncano-Berrocal & Carter, 2021). Intuitively, something about the persuasive argumentation account seems insufficient: Even where views are changing because individuals are revising their views in light of new reasons, there remains a puzzle in as much as one might expect groups to include some initial diversity of opinion. If beliefs overall become (uni-directionally) more extreme either some of these opinions must be being ignored, or one side has a stronger case. Why the latter would so frequently be the case, however, seems in need of further explanation.

The suggestion of irrationality becomes even more pronounced when considering what is, arguably, now the more widely studied form of polarization: polarization as belief divergence (e.g., Sunstein, 2002). How can a collective, given the same available body of evidence, come not only to more extreme beliefs, but fracture into diametrically opposed groups? Such polarization also raises many practical concerns, from threatening the democratic process (e.g., Fishkin, 1991; Sunstein, 2018; Dalton, 2008; Fiorina, 2016; Jost et al., 2022), through to high stakes practical concerns such as the climate (Guilbeaut et al., 2018; Schaefer et al., 2012) and vaccine debate (Horne et al., 2015; Kata, 2012; Mønsted & Lehmann, 2022). A now sprawling literature on polarization includes the potential role of the internet (Wojciesszak & Mutz, 2009; Sunstein, 2018; Pariser, 2011; Dubois & Blank, 2018; Flaxman et al., 2016) and the advent of social media in promoting partisanship (Adamic & Glance, 2005; Tucker et al., 2017, Himmelboim et al., 2013; Del Valle & Borge Bravo, 2018; Bail et al. 2018; Bakshy et al., 2015; Barbera et al. 2015; Törnberg, 2022), conspiracy theories and fake news (Vosoughi et al., 2018; Del Vicario et al., 2016). This wealth of research, however, typically failed to distinguish sufficiently clearly between rational and irrational,

or epistemic and motivational accounts (e.g., Sunstein, 2009) and has been faulted, more generally, for providing insufficient understanding of *individual level mechanisms* (see, e.g., Lindell et al., 2017).

Lord et al.'s seminal (1979) study on ``biased assimilation'' is in keeping with that individual level focus: in this polarization phenomenon, the same (mixed) evidence leads different individuals to reinforce their own, opposing initial views. In Lord et al.'s study, participants were initially assessed as for or against capital punishment. Each participant then read two (experimenter designed) journal articles: one purported to show the effectiveness of capital punishment as a deterrent of crime, the other sought to show its ineffectiveness. The expectation was consequently that these two sets of arguments would largely cancel each other out. However, participants rated the report that agreed with their prior opinion as "more convincing,". They also more readily found flaws in the counter-attitudinal report. Most importantly, participants' subsequent beliefs shifted further in the direction of their initial opinion. This finding, too, has been replicated (e.g., Lord, Lepper & Preston, 1984), and 'biased assimilation' subsequently came to be viewed as one of the key pieces of evidence for the existence of motivated reasoning (see Kunda, 1990) as a biased, irrational, form of cognition.

These three different forms of 'polarization' --shift to extremity, belief divergence, and biased assimilation—share multiple connections: in particular, shift to extremity and biased assimilation are both invoked to explain belief divergence. There are also important differences that will resurface in subsequent discussion in this paper. All three conceptually involve multiple individuals. Biased assimilation, at least in the original paradigm, is only detected by observing different individuals respond in opposing ways to the same evidence. However, unlike the shift to extremity which involves group deliberation, it does not involve interaction between those individuals (see also Broncano-Berrocal & Carter, 2021). Polarization as belief divergence, finally, typically describes (merely) a distributional characteristic of individuals' beliefs within a wider group or population (see Bramson et al., 2016).

The goal of the present paper is not to resolve extant debates on any of these three phenomena. It is also not the goal of this paper to help resolve whether they are rational or irrational in any given real-world situation. Rather, the paper aims to take a step back and draw out a common underlying theme across all three: dependence. With this, it seeks to make the case that our continued surprise at these phenemona (as individuals and as a research community) reflects our continued failure to fully appreciate the full consequences of our social embedding. Appreciating these means recognizing determinants of our beliefs that are beyond our individual control.

To this end, the remainder of the paper analyses idealised, rational agents and focusses on beliefs about factual statements and accuracy. It then returns both to other mental states such as attitudes and opinions and to more realistically human models at the end.

`Dependencies' in the Availability and Impact of Evidence

The 'exchange of new information' central to persuasive argumentation theory renders shifts to extremity rational to the extent that group members are simply responding to new information. This has also been captured in

models of the phenomenon (Maes & Flache, 2013). However, as just discussed, it may seem implausible that groups *regularly* shift to extremity. Why should a group, as a whole, be likely to possess more evidence in favour of one position than another?

Availability of Evidence

To explain why an imbalance in the available amount of evidence for or against a claim is likely , we return to communication within a group of idealised Bayesian (such as our agents in the Olsson model underlying Fig. 1 above). For such a group, there is little reason to assume that the distribution of available evidence is evenly distributed across arguments/evidence for and against the claim.

First, to the extent that agents have already formed a degree of belief in the claim, that belief reflects the evidence they have encountered. A group of rational agents that leans initially toward endorsing a claim is, from that perspective, one that is in possession of more or stronger arguments in favour. This is the essence of the argumentative view of the shift to extremity. Individuals lean toward an initial position *because* they have more (or stronger) arguments in favour of that position. As a consequence, more (or stronger) arguments in favour of that position will be available for exchange in the deliberation (Vinokur & Burnstein, 1978). For example, (rational) individuals believing it more likely than not that a bear is behind the recent spate of vandalism in their neighbourhood, believe this to be so because they have stronger evidence for the bear hypothesis than the alternative of, say, a disgruntled neighbour. To the extent that subsequent group discussion then surfaces evidence that is new to one or more individuals in the group, that evidence is more likely to be further evidence in favour of the bear. Revising beliefs in light of information obtained during deliberation is thus more likely to shift beliefs in favour, as opposed to against, the bear hypothesis.

Second, the consideration of idealised Bayesian agents adds further to this by additionally answering the question of why one would expect an imbalance of evidence in the first place. Even before anyone has actually encountered any arguments, the *expected* argument/evidence distribution will not be uniform. This follows from the Bayesian conceptualisation of argument or evidence itself. On that view, an argument or piece of evidence is strong or diagnostic to the extent that it is much more likely to be found if the hypothesis is true than if it is false (expressed by the so-called likelihood ratio P(e|H)/P(e|not_H), see e,g., Hahn, 2020). Stronger evidence, so defined, will (normatively) lead to greater changes in belief.

How likely that evidence is full stop (i.e., its so-called marginal probability) is determined by total probability: P(e) = P(e|H)*P(H) + P(e|not_H)*P(not_H) (in words, the probability of obtaining it if the underlying hypothesis is true, weighted by the probability of the hypothesis, plus the probability of a false positive, weighted by the probability of the hypothesis being false). This means also that (all other things equal), individuals are more likely to encounter strong evidence in favour of a hypothesis than equally strong evidence against, if indeed the hypothesis is true (see Hahn, 2023, for further elaboration and discussion with respect to notions of confirmation bias). In other words, we would expect stronger evidence in favour of the bear, if a bear is indeed the

true cause of the vandalism in the neighbourhood. While it is entirely possible that on a particular topic or occasion one currently finds stronger evidence against a true hypothesis than for it, that will not be the case *in expectation*.

In short, one may expect a group of rational agents that have received prior evidence to lean, on average, in a particular direction. The information they have available for exchange is then likely to promote additional change in that direction.

Perceived Source Reliability

With respect to the evidence they receive, the stylised Bayesian agents of the Olsson (2011) model face a challenge that is common in the real world: they do not know the true (obj.) diagnosticity of that evidence. Their estimates of the likelihoods, and with that the impact of the arguments on their beliefs, is subjective. As Hahn et al. (2018) detail, multiple strategies exist for estimating these, such as drawing on past track records of accuracy. However, individuals might not actually know the person they are communicating with. Or the topic is sufficiently outside the scope of past exchange or the other person's expertise that it seems problematic to extrapolate from past performance.

Hence, the Olsson model implements a strategy that seeks to estimate *both* the reliability of an evidence source and the probability of the claim at issue from the evidence the source provides. Other models of source reliability in the literature implement the same intuition, albeit with slightly different technical detail (see Merdes et al., 2022, for discussion, including recent modelling work that has sought to understand more fully the implications of such a strategy). Current experimental evidence suggests that something like this strategy is also adopted by lay reasoners. Studies have found that participants receiving arguments from a source in stylised scenarios spontaneously took the plausibility of those arguments to impact both their belief in the claim at issue and the perceived reliability of the source (Collins & Hahn, 2019; Collins et al., 2018).

Regardless of how subjective estimates of the likelihoods, and hence the diagnostic value of an argument/piece of evidence, are obtained, the actual impact of evidence on beliefs will not be the same for rational agents that reasonably disagree on those likelihoods. This means also that contrary to the implicit assumption guiding Lord et al.'s ``biased assimilation'' it is entirely possible for rational agents to draw different (and even opposing) conclusions from the same piece of evidence as a result (see also Hahn & Harris, 2014; Jern et al., 2014; Hahn, Harris & Corner, 2016; Druckman & McGrath, 2016). Moreoever, the value they assign to arguments may not be independent of other beliefs and evidence agents presently hold.

Dependence and Interdependence in Communication within Collectives

Considerations of the 'dependencies' highlighted in the previous section already provide some correction to perceptions of bias or irrationality seemingly implied by shift to extremity, belief divergence, and biased assimilation. Their real impact, however, only comes to the main source of dependence considered in this paper:

the dependence brought about by communication within a collective. Within a social network (online or offline) agents communicate with one another, hence the same information can reach agents along multiple, different paths. This changes fundamentally how belief formation within a collective will unfold.

The Recursive Nature of Social Communication

Consider now two of our Bayesian agents, Agent 1 and Agent 2. Over multiple rounds, they provide each other with information and revise their beliefs in light of the evidence received. As just discussed, the impact of the evidence received is determined by the (subjective) likelihoods the agents assign its source at each point in time.

Where one agent's evidence effects the other's beliefs and those beliefs are eventually communicated back to the first agent, this creates a feedback loop: Agent 1's *perception* of the quality of the evidence received from Agent 2 not only determines its impact on the beliefs of Agent 1, but, also, because of the return flow via communication, impacts the *objective accuracy of the beliefs of Agent 2*. My perceptions of your evidence end up influencing not just my accuracy, and hence, actual reliability, but also *yours*: because my accuracy is a function of *both* the objective quality of your information and my subjective estimate thereof (because the latter determines its actual impact on my beliefs), my perceptions of your reliability influence your actual reliability.

All of this is the case even where agents do not additionally avail themselves of strategies for dynamically revising estimates of reliability (as just outlined), but simply assign a fixed, subjective value. Any dynamic revision on top, however, may exacerbate the dependence.

Of course, in real world contexts, we exchange more information than is captured in the pure testimony of the Olsson model (see also Collins et al., 2018): we (also) exchange supporting arguments in favour of a claim. But this will not break that feedback loop unless the supporting arguments we search for or select, and the diagnostic values we assign, are completely independent of any change in belief we might be undergoing.

In short, one reason why viewing the parties in an argument as mere argument dispensers that require no further consideration is flawed, is because their reliability affects the impact their evidence should have. And that same reliability will likely change objectively over the course of the discourse. Unfortunately, however, even for the simple case of the Olsson model it is difficult (or even impossible) to track these mutual influences appropriately over time, as we discuss next.

Dependence Across Social Networks

In fact, the agents of the Olsson model are not optimally Bayesian. They are so-called naïve Bayesian agents who assume that the evidence they receive from different sources (or the same source at different times) is *independent*. This simplification is clearly false, but necessary in practice.

Even a social network where information only ever flows in one direction between directly communicating agents retains the problem that the same evidence can reach an agent via different paths. Normatively, the impact of receiving three reports of a bear in the neighbourhood should differ depending on whether these reflect independent sightings or reflect three individuals passing on the same underlying report by a fourth party.

Bayesian models can capture some cases of such dependence in appropriate ways (and analysis reveals that independent evidence is not always stronger, see Pilditch et al., 2020; Bovens & Hartmann, 2003). Lay reasoners also seem sensitive to some (but not all) of those distinctions (e.g., Whalen, Griffiths & Buchsbaum, 2018; Madsen, Pilditch & Hahn, 2020; Pilditch et al., 2020). However, one cannot, in practice, solve this for an entire social network. Individuals simply do not know the wider structure of their communication networks beyond their immediate neighbours (or at best, their neighbour's neighbours), particularly in the age of online social media where networks comprise millions of nodes. Even less do they know the specific contents of communicative exchanges they are not directly party to. Finally, even in a small network such as those simulated above with the Olsson model, it would be computationally intractable to factor in those dependencies (see also Merdes et al., 2020 for extended discussion).

An immediate consequence of this falsely assumed independence is that agents may come to *overweight* the evidence. In effect, there may be 'double counting' for evidence that reaches an agent via multiple routes (e.g., the same 'underlying bear sighting' received via three different sources). That double-counting can become apparent both individually and collectively.

Fig 2 shows the belief dynamics of sample runs of the Olsson model: shown are the average degree of belief of agents in the model (starting with a prior of .5 reflecting ignorance) over multiple time steps. The red line reflects the dynamics for agents in the network. The blue lines indicate the belief dynamics of perfectly matched agents who receive exactly the same evidence from the world, but do not participate in communication. Typical runs see faster convergence for the communicating agents (Fig 2: left hand and middle plot). Occasionally, the double counting created by dependence will become so severe that the average degree of belief in the network clearly exceeds the available evidence. To demonstrate this, Fig. 2 also graphs the belief dynamics of a single 'ideal agent' who has directly received all evidence from the world that went into the network as a whole and has weighted it by the true, objective likelihood. Typically, the communicating agents in the network (red line) are closer to this ideal agent (green line) than those who, individually receive the same evidence but do not participate in communication (blue line). This demonstrates the value of communication. However, under reasonable starting conditions, they will be more 'conservative', that is, less extreme in their mean belief than the ideal agent until their beliefs converge at the truth, because, individually, they still have less information. On occasion, though, the network beliefs may become more extreme than the total data going into the network actually warrants (right panel). That overshoot is a result of dependence in communication.

The chance that such overshoot will happen depends on the degree of dependence that arises as a function of the structure of the network (its topology) and the amount of communication that takes place relative to uptake of external evidence from the world. Both in the model, and in the real world, the impact of the dependence structure embodied by the network is mediated by the actual impact of any dependent evidence on individual agents' beliefs. This impact, in turn. will be mediated by the subjective likelihoods agents assign, and by how much other, independent, evidence they have, given that all evidence is ultimately aggregated into a single posterior degree of belief. Both the agents' threshold of assertion agents and the probability of communication will then determine the extent to which it is passed on.

This obscures the magnitude of the dependence at the individual level. It is thus no surprise that humans may under-estimate the scale of such dependence, given that its consequences are, at best, indirectly observable. To help inform our understanding of the scale and nature of that influence, it thus seems useful to conclude with the simulation of an even simpler, more idealized case. Imagine a world where communication consists solely of a kind of 'pass the parcel': information comes in unit parcels, and communication consists of handing whatever token parcels one possesses at a given point in time to the person we communicate with.

Fig. 3 shows the outcome of multiple such rounds of pass the parcel across a small-world network (see Watts & Strogacz, 1998) of 10 agents. At initiation of the exchange (Timestep 0), each agent has one unit parcel of information. In each round of communication (subsequent time steps), the agent passes whatever parcels she presently possesses in her store to her interlocutor. Any parcels received from others at that time step are added to the agent's memory store. To help track information flow, the simulation shows each agents' *initial* information parcel in red, and all other information parcels in blue. The bar-plots in Fig. 3 show for each agent (subplots 1 to 10) all the parcels in its current store at that time step, separated out by which initial agent information parcel they represent (indexed within each subplot by the numbers 1 to 10). At Timestep 1, after the first round of communication, each agent has its initial unit parcel (in red) and 1 token of the initial unit parcels it has received from of each of the other agents it communicates with. The point of interest is to see how the content of each agent's store develops over subsequent rounds.

The plot for Timestep 2 shows the systematic differences across agents in the amount of information they possess from each individual agent, including those with which they do not communicate directly, as a result of varying position in the network. It is these differences that ultimately give rise to the effects of structure on the accuracy of individual and collective beliefs described in the introduction. The plot shows also just how much of the agent's store after 2 steps consists of information the agent originally had which is now *coming back to that agent* (red bars). The plot for Timestep 10, finally, gives an indication of how all of those differential influences will persist over time.

In an ideal world, each agent would end up with *only* the 10 distinct information parcels that were present at the initiation of the communication, as these constitute the distinct pieces of information present across the network as a whole. For that to happen, however, the network structure would have to be fully known to

individual agents, *and*, it would need to be possible to uniquely assign the information tokens to the appropriate information type ("initial Agent 1 info" etc..). Neither is possible in practice. The former will be impossible because the network structure is typically not known, and where it is, rapidly becomes computational too costly to trace. The second will be impossible because the individual tokens are themselves partly or wholly folded into aggregate, evaluative judgments of information by each agent and thus not communicated individually. Consequently, at best, some of the reduplication can be undone.

Revisiting Opinion Extremity, Belief Divergence, and Biased Assimilation

From all of the preceding material, it should become clear that shifts to extremity, belief divergence within groups, and biased assimilation, are not only *possible* for rational agents (or, better, agents that are as rational as realistically possible), they are *unsurprising*, in as much as they derive from fundamental features of belief formation for those agents when placed in collectives.

Shifts to extremity are unsurprising because arguments and evidence of a given strength or quality are not uniformly distributed, and the in-eliminable dependence that comes with communication across a collective serves to amplify the cumulative impact of arguments and evidence beyond their true diagnostic value.

Belief divergence will arise because that same dependence structure will selectively promote different pieces of information in different parts of the network, as demonstrated in the persistent differences in the parcel distributions across agents in Fig. 3: imagine simply that, say, four of the initial information parcels spoke for, and six against, a particular claim. The differences in the distributions for each agent (across the 10 different subplots) mean that different overall conclusions would be reached.

This, finally, illustrates how, at an aggregate level, there is no realistic baseline of "unbiased" assimilation: even if agents in the simple simulation of Fig. 3 weighted each token information parcel in their store completely equally, the ultimate impact of the 10 initial, unit (and thus equal) parcels would not be the same.

The fact that even a single piece of information received (just once) from the same source by two different agents need not, normatively, be treated equal amplifies this. Any functional dependence between perceived source reliability and plausibility of the evidence vis a vis one's current degree of belief in a claim (as underlies multiple formal models of source reliability, and seems to be present in actual human behaviour, see Section 2 above) will lead to systematic 'bias' in assimilation. This will add a ratchet that amplifies all three phenomena: shifts to extremity, polarization, and biased assimilation.

All of this becomes clear from considering simple stylised models. The very simplicity of these models underscores how fundamental and ineliminable the driving constraints are. Nevertheless, it is also useful and important to think about more realistic situations. In the remainder, we consider two aspects of such realism.

From Beliefs to Opinions

The first of these is the extension from beliefs to opinions. The Olsson (2011) model, and, with it, considerations about rational Bayesian agents, is about *beliefs*. At issue is a claim about the world that is either true or false (a 'proposition'), and probabilities represent agents' current degree of belief in the truth of that claim. Much of our real-world discussion (and hence data on shift to extremity, polarization, and biased assimilation), however, concerns *opinions*.

Opinions (or attitudes, see Eagly & Chaiken,1993) are *valuations* (such as "chocolate tastes nice", "green is a pretty colour"), not factual statements.[2] So we may ask about the extent to which the considerations of previous sections apply to opinions and opinion dynamics also. Arguably, the only points made above that do *not* apply directly are the points made about the distribution of arguments of a given strength as a function of the truth or falsity of the claim at issue, because opinions are not true or false.

Beyond that, it is reasonable to assume that different persuasive messages concerning opinions, too, differ in perceived strength, that we may include features of the message source to moderate their impact, and that the initial distribution of opinions in a group is unlikely to be perfectly matched. This suggests we can just also the naïve Bayesian model as a simple *process model of opinion formation*, and all other aspects of the analysis stay the same.

In fact, the theoretical distinction between beliefs and opinions does not seem salient to lay reasoner in the first place, so assuming common psychological processes for both does not seem unreasonable.
The issue, then, becomes whether the model is too simplistic in other ways.

From Bayesian Agents to Real People

Real people and real belief formation in collectives will not be like the model. We know that lay reasoners, at the very best, approximate Bayesian inference in other contexts (Hahn & Harris, 2014). We increasingly know also from experimental investigation that lay reasoner's sensitivity to dependence seems limited (e.g., Yousif et al., 2019; Pilditch et al. 2020). This makes it seem unlikely that a descriptively adequate model of actual human behaviour would do better.

Actual humans cannot do better than the most rational model possible. It would thus have to be the case that such a model existed. Real world testimony may involve both what Collins et al. (2018) call "mere testimonial assertion" (whereby an agent asserts, *as evidence*, that the claim at issue is true) and the "transmission role of testimony" (whereby an agent transmits supporting arguments for the claim at issue). The Olsson model incorporates only testimonial assertion, not the communication of other, supporting evidence. As already suggested above, this changes little in principle. For one, testimonial assertion is arguably *part of* the speech act of communicating evidence. Given pragmatic principles to the effect that the speaker believes the evidence and that

[2] The difference becomes apparent by considering that two people cannot disagree about a putative fact (say, 'the Earth is round') and both be (wholly) right, but they can disagree about whether chocolate tastes nice or green is prettier than blue.

it is relevant to the listener (Levinson et al., 1983), it is pragmatically odd to provide (only) a strong argument for claim one doesn't believe. Testimonial assertion will thus remain part of what is happening in communication across a network. It may sometimes (or even often) be possible to identify supporting arguments by their content as arguments which one has encountered before, and thus resist any double counting (in the earlier example "John said he saw a bear" received from different sources). But this will also often *not* be the case ("someone saw a bear").

More realism in this regard may thus attenuate effects of dependence, but it will not wholly eliminate them. Adding in supporting arguments thus simply becomes part of the general considerations about the relative influence of dependent to independent evidence in any concrete setting. The precise balance already various as a function of what is at issue, what evidence for it exists, and the topology of the network across which it is communicated. In that sense, adding in supporting arguments ultimately adds nothing fundamentally new or different.

Finally, real people could be much worse than any rational agent model, because of motivational or otherwise distorting biases (see e.g, Hahn & Harris, 2014; Lewandowsky et al., 2013; Kahan, 2013; Miller et al., 2016). There is little reason to doubt such biases exists. There are, however, empirical questions about how much of a difference these actually make in practice given the in-eliminable influence of dependence. Given its influence, the real bias looks to have been our persistent belief that cognition in collectives could have been otherwise, and *not* exhibited all the features that have so long been taken to be indicative of bias.

Conclusions

The seeming irrationality implied by shifts to extremity and belief divergence have prompted a long history of demonstrations that rational agents may on occasion exhibit irrationality in groups: from research on information cascades (Anderson & Holt, 1997; Acemoglu et al., 2011 ; Bikchandani et al, 1992), through to a plethora of work on polarisation in rational models (besides the model as discussed here, Olsson, 2013, see also e.g., Madsen et al., 2018; O'Connor & Weatherall, 2018, 2019).

That interest (present authors included) reflects our surprise at these behaviours. The enduring value of rational reconstruction in this context, arguably, does not lie in the fact that rational models may show such behaviour. The real value is that trying to build a model of rational agents doing the best they can, reveals the root of these behaviours to lie with something that cannot be eliminated wholly from communication in and across a collective: dependence. Communication across a social network, that is communication by individuals within a collective, is pervasive. With such communication comes dependence. The real news is that it has taken us so long to see this and appreciate its implications.

Acknowledgements

The research reflected in this paper was supported by the Humboldt Foundation's Anneliese Meyer Research Award and an AHRC/DFG grant awarded to the author. Special thanks go to Momme von Sydow and Christoph Merdes for their coding, analysis, and many, many conversations that went into understanding how communication across social networks actually happens.

References

Acemoglu, D., Como, G., Fagnani, F., & Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, *38*(1), 1–27. http://doi.org/10.1287/xxxx.0000.0000

Acemoglu, D., Dahleh, M. A., Lobel, I., & Ozdaglar, A. (2011). Bayesian learning in social networks. *The Review of Economic Studies*, *78*(4), 1201-1236.

Adamic, L. A., & Glance, N. (2005, August). The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery* (pp. 36-43). ACM.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211-36.

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *The American Economic Review*, 847-862.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F. & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216-9221.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130-1132.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531-1542.

Baron, R. S., Hoppe, S. I., Kao, C. F., Brunsman, B., Linneweh, B., & Rogers, D. (1996). Social corroboration and opinion extremity. *Journal of Experimental Social Psychology*, *32*(6), 537-560.

Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the national academy of sciences*, *114*(26), E5070-E5076.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992-1026.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford University Press on Demand.

Bramson, A., Grim, P., Singer, D. J., Fisher, S., Berger, W., Sack, G., & Flocken, C. (2016). Disambiguation of social polarization concepts and measures. *The Journal of Mathematical Sociology*, *40*(2), 80-111.

Broncano-Berrocal, F., & Carter, J. A. (2021). *The philosophy of Group polarization: Epistemology, Metaphysics, psychology*. Routledge.

Burnstein, E.. & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology. 13.* 315-332.

Coady, C. A. J. (1992). *Testimony: A philosophical study*. Clarendon Press.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), 317–332.

Collins, P. J., & Hahn, U. (2019). We might be wrong, but we think that hedging doesn't protect your reputation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Online first.

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, section Cognition.

Dalton, R. J. (2008). "The Quantity and the Quality of Party Systems: Party System Polarisation, Its Measurement, and Its Consequences." *Comparative Political Studies* 41:899-920.

Del Valle, M. E., & Borge Bravo, R. (2018). Echo Chambers in Parliamentary Twitter Networks: The Catalan Case. *International Journal of Communication*, *12*, 21.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554-559.

Dingemansea, M., Liesenfelda, A., Rasenberga, b, M., Albertc, S., Amekad, F. K., Birhanee, f, A., ... & Wiltschkogg, M. (2023). Beyond Single-Mindedness: A Figure-Ground Reversal for the Cognitive Sciences. *Cognitive Science*, *47*(1), e13230.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, *21*(5), 729-745.

Douven, I. (2019). *Computational Models in Social Epistemology*. Chapter 45. In: M. Fricker, P. Graham, D. Henderson, N. J. Pedersen, & J. Wyatt (Eds). *The Routledge Handbook of Social Epistemology*.

Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, *9*(2), 111-119.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.

Festinger, L. (1950). Informal social communication. *Psychological Review*, *57*(5), 271–282. https://doi.org/10.1037/h0056932

Fiorina, M. P. (2016). Has the American public polarized?. *Hoover Institution*.

Fishkin, J. S. (1991). *Democracy and deliberation: New directions for democratic reform* (Vol. 217). New Haven, CT: Yale University Press.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298-320.

Friedkin, N. E. (1999). Choice shift and group polarization. *American Sociological Review*, 856-875.

Garcia, D., Abisheva, A., Schweighofer, S., Serdült, U., & Schweitzer, F. (2015). Ideological and temporal components of network polarization in online political participatory media. *Policy & Internet*, *7*(1), 46-79.

Golub, B., & Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), 112-49.

Goodin, R. E., & Niemeyer, S. J. (2003). When does deliberation begin? Internal reflection versus public discussion in deliberative democracy. *Political Studies*, *51*(4), 627-649.

Gruzd, A., & Roy, J. (2014). Investigating political polarization on Twitter: A Canadian perspective. *Policy & Internet*, *6*(1), 28-45.

Guess, A., & Coppock, A. (2018). Does counter-attitudinal information cause backlash? Results from three large survey experiments. *British Journal of Political Science*, 1-19.

Guilbeault, D., Becker, J., & Centola, D. (2018). Social learning and partisan bias in the interpretation of climate trends. *Proceedings of the National Academy of Sciences*, *115*(39), 9714-9719.

Hahn, U. (2020) Argument quality in real world argumentation. Trends in Cognitive Sciences, 24, 363-374.

Hahn, U. (2022). Collectives and Epistemic Rationality. *Topics in Cognitive Science*, 14, 602–620.

Hahn, U. (2023). Revisiting confirmation bias: Strong arguments in favour are easier to find. Manuscript in prep.

Hahn, U., & Harris, A. J. (2014). What does it mean to be biased: Motivated reasoning and rationality. *The Psychology of Learning and Motivation*, *61*, 41–102.

Hahn, U., Harris, A. J., & Corner, A. (2016). Public reception of climate science: Coherence, reliability, and independence. *Topics in cognitive science*, *8*(1), 180-195.

Hahn, U., Merdes, C. & von Sydow, M. (2018). How Good is Your Evidence and How Would You Know? *Topics in Cognitive Science, 10,* 660-678.

Hahn, U., Hansen, J.U. & Olsson, E.J. (2020) Truth tracking performance of social networks: how connectivity and clustering can make groups less competent. Synthese, 197 (4), 1511-1541.

Hahn, U., Harris, A. J., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, *29*(4), 337-367.

Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication*, *18*(2), 154-174.

Horne, Z., Powell, D., Hummel, J. E., & Holyoak, K. J. (2015). Countering antivaccination attitudes. *Proceedings of the National Academy of Sciences*, *112*(33), 10321-10324.

Hsu, H. L., & Park, H. W. (2012). Mapping online social networks of Korean politicians. *Government Information Quarterly*, *29*(2), 169–181.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, *50*(6), 1141.

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, *59*(3), 690-707.

Jasny, L., Waggle, J., & Fisher, D. R. (2015). An empirical examination of echo chambers in US climate policy networks. *Nature Climate Change*, *5*(8), 782.

Jern, A., Chang, K. M. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, *121*(2), 206.

Jolley, D., & Douglas, K. M. (2014). The effects of anti-vaccine conspiracy theories on vaccination intentions. *PloS one*, *9*(2), e89177.

Jost, J. T., Baldassarri, D. S., & Druckman, J. N. (2022). Cognitive–motivational mechanisms of political polarization in social-communicative contexts. Nature Reviews Psychology, 1(10), 560-576.

Jönsson, M. L., Hahn, U., & Olsson, E. J. (2015). The kind of group you want to belong to: Effects of group structure on group accuracy. *Cognition*, *142*, 191-204.

Kahan, D.M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making* 8 (4): 407–24.

Karlsen, R., Steen-Johnsen, K., Wollebæk, D., & Enjolras, B. (2017). Echo chamber and trench warfare dynamics in online debates. *European Journal of Communication*, *32*(3), 257-273.

Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm–An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, *30*(25), 3778-3789.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480.

Lamm, H., & Myers, D. G. (1978). Group-induced polarization of attitudes and behavior. In *Advances in experimental social psychology* (Vol. 11, pp. 145-195). Academic Press.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. (2018). The science of fake news. *Science*, *359*(6380), 1094-1096.

Lee, J. K., Choi, J., Kim, C., & Kim, Y. (2014). Social media, network heterogeneity, and opinion polarization. *Journal of Communication*, *64*(4), 702-722.

Lelkes, Y. (2016). Mass Polarization: Manifestations and Measurements, *Public Opinion Quarterly*, 80, S1, 392–410, https://doi.org/10.1093/poq/nfw005

Levinson, S. C., Levinson, S. C., & Levinson, S. (1983). *Pragmatics*. Cambridge university press.

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—therefore,(climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science*, *24*(5), 622-633.

Lewiński, M., & Aakhus, M. (2014). Argumentative polylogues in a dialectical framework: A methodological inquiry. *Argumentation*, *28*, 161-185.

Li, L., Scaglione, A., Swami, A., & Zhao, Q. (2013). Consensus, polarization and clustering of opinions in social networks. *IEEE Journal on Selected Areas in Communications, 31*(6), 1072-1083.

Lindell, M., Bächtiger, A., Grönlund, K., Herne, K., Setälä, M., & Wyss, D. (2017). What drives the polarisation and moderation of opinions? Evidence from a Finnish citizen deliberation experiment on immigration. *European Journal of Political Research*, *56*(1), 23-45.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarisation: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, *37*(11), 2098.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: a corrective strategy for social judgment. *Journal of personality and social psychology*, *47*(6), 1231.

Luskin, R.C., Fishkin, J.S. & Jowell, R. (2002). Considered opinions: Deliberative polling in Britain. *British Journal of Political Science* 32: 455–487..

Madsen, J. K., Bailey, R. M., & Pilditch, T. D. (2018). Large networks of rational agents form persistent echo chambers. *Scientific Reports*, *8*(1), 12391.

Madsen, J. K., Hahn, U., & Pilditch, T. D. (2020). The impact of partial source dependence on belief and reliability revision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(9), 1795.

Mäs, M., & Flache, A. (2013). Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. *PloS one*, *8*(11), e74516.

Merdes, C., von Sydow, M. & Hahn, U. (2021). Formal models of source reliability. Synthese, *198*(23), 5773-5801.

Merkle, D. (1996). The National Issues Convention Deliberative Poll. *Public Opinion Quarterly* 60: 588– 619.

Miller, A. G., McHoskey, J. W., Bane, C. M., & Dowd, T. G. (1993). The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology*, *64*(4), 561.

Miller, J. M., Saunders, K. L., & Farhart, C. E. (2016). Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust. *American Journal of Political Science*, *60*(4), 824-844.

Mønsted, B., & Lehmann, S. (2022). Characterizing polarization in online vaccine discourse—A large-scale study. PloS one, 17(2), e0263746.

Myers, D. G. & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin. 83.* 602-627.

Myers, D. G. (1975). Discussion-induced attitude polarization. *Human Relations*, *28*(8), 699-714.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175.

O'Connor, C., & Weatherall, J. O. (2018). Scientific polarization. *European Journal for Philosophy of Science*, *8*(3), 855-875.

O'Connor, C., & Weatherall, J. O. (2019). *The misinformation age: how false beliefs spread*. Yale University Press.

Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme*, *8*(02), 127–143.

Olsson, E. J. (2013). A Bayesian simulation model of group deliberation and polarisation. In *Bayesian Argumentation* (pp. 113-133). Springer Netherlands.

Page, S. (2010). Diversity and complexity. In *Diversity and complexity*. Princeton University Press.

Petty, R. E., Cacioppo, J. T., Petty, R. E., & Cacioppo, J. T. (1986). *The elaboration likelihood model of persuasion* (pp. 1-24). Springer New York.

Quattrociocchi, W., Caldarelli, G., & Scala, A. (2014). Opinion dynamics on interacting networks: media competition and social influence. *Scientific Reports, 4*, 4938.

Sanders. G. S., & Baron, R. S. (1977). Is social comparison irrelevant for producing choice shifts? *Journal of Experimental Social Psychology. 13.* 303-314.

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, *116*(16), 7662-7669.

Scheufele, D. A., Hardy, B. W., Brossard, D., Waismel-Manor, I. S., & Nisbet, E. (2006). Democracy based on difference: Examining the links between structural heterogeneity, heterogeneity of discussion networks, and democratic citizenship. *Journal of Communication*, *56*(4), 728-753.

Schkade, D., Sunstein, C.R., & Hastie,R. (2010) When Deliberation Produces Extremism, *Critical Review*, 22:2-3, 227-252.

Sporer, S. L. (1982). A brief history of the psychology of testimony. *Current Psychological Reviews*, *2*(3), 323-339.

Stoner, J. A. F. (1968). Risky and cautious shifts in group decisions: The influence of widely held values. *Journal of Experimental Social Psychology. 4.* 442-459.

Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, *10*(2), 175-195.

Sunstein, C. R. (2006). Deliberating groups versus prediction markets (or Hayek's challenge to Habermas). *Episteme*, *3*(3), 192-213.

Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. Oxford University Press.

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Törnberg, P. (2022). How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences*, *119*(42), e2207159119.

Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). From liberation to turmoil: social media and democracy. *Journal of Democracy*, *28*(4), 46-59.

Vinokur, A., & Burnstein, E. (1978). Depolarization of attitudes in groups. *Journal of Personality and Social Psychology*, *36*(8), 872.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442.

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. Cognitive science, 42(1), 168-187.

Wojcieszak, M. (2015). Polarization, political. The International Encyclopedia of Political Communication, 1-7.

Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement?. *Journal of Communication*, *59*(1), 40-56.

Woods, J., Irvine, A., & Walton, D. N. (2004). *Argument*: *Critical thinking, logic and the fallacies* (Rev. ed.). Toronto, Ontario, Canada: Prentice Hall.

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology & Society*, *30*(5), 316-327

Yousif, S. R., Aboody, R., & Keil, F. C. (2019). The illusion of consensus: A failure to distinguish between true and false consensus. *Psychological Science*, *30*(8), 1195-1204.

Figures

Fig. 1. The figure displays data taken from Hahn et al., 2020 that shows performance across different network structures, for networks matched in size (number of nodes). All but the fully connected network additionally have the same number of connections between nodes as well. Networks “regular4” and “regular4distant” each represent a lattice and differ only in a single connection that has been rewired to create a distant link across the network.
In the figure, the data from Hahn et al. have been transformed to represent error, measured as the squared distance to the true parameter value for the claim at issue (i.e., here *P(claim)=1*). Individual error (Ind Error) is the average error across individuals. Collective error (CE) represents the error of the group average. This is equal or lower than the average individual error reflecting wisdom of crowd effects.

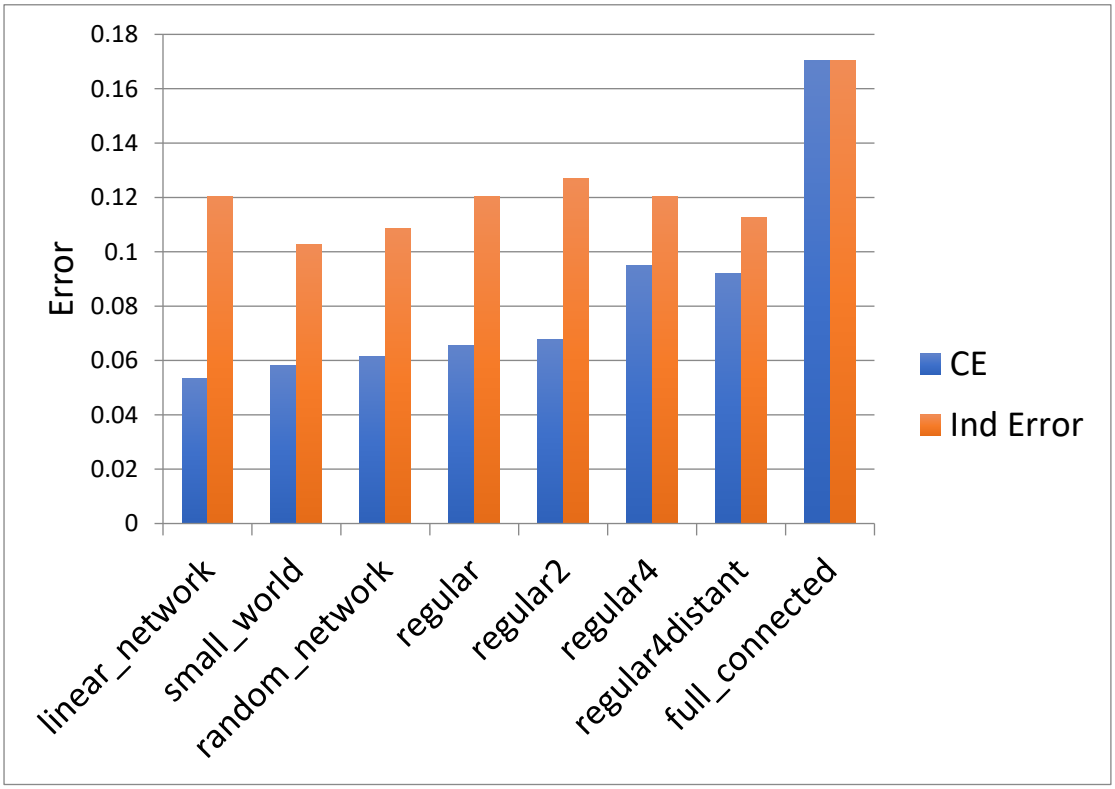

Fig. 2. Sample runs of the Olsson (2011) model using a small world network (Watts & Strogatz, 1998). Displayed is the average degree of belief (probability) across time. In this simulation, all agents start with a prior of p = .5 reflecting ignorance. The red line shows agents in the network, the blue line shows a group of matched 'shadow agents' (see also Hahn et al., 2019): each an exact copy of an agent in the network that receives the same evidence from the world, but does not communicate. The green line shows a single 'ideal agent' receiving all evidence from the world going into the network. The value of communication is seen in the fact that the networked agents (red line) are closer to the ideal agent, than shadow agents. On occasion, however, beliefs for the network become more extreme than warranted by the total evidence to network/ideal agent (right hand plot) due to dependence.

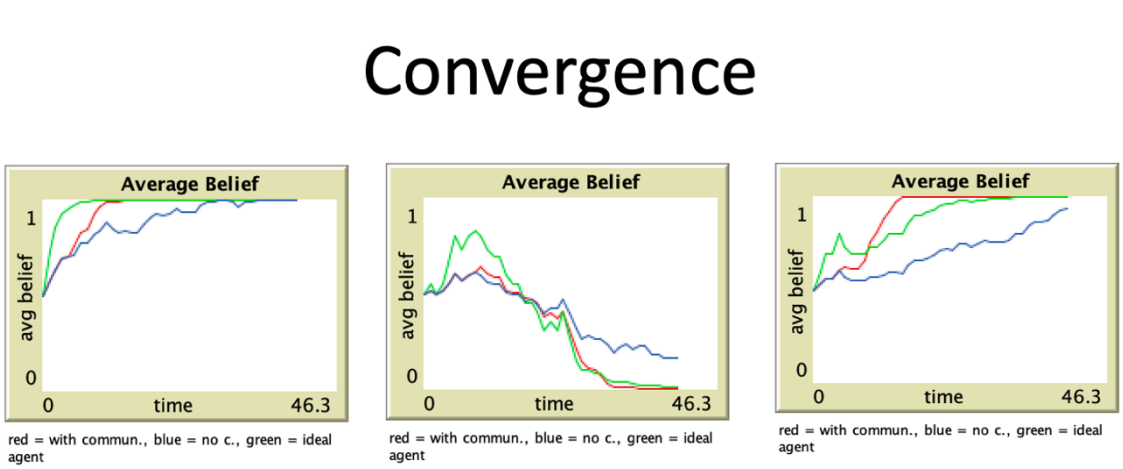

Fig. 3. The figure shows the flow information in a small world network (top left) of 10 agents. The 10 subplots at each Timestep are histograms representing the information store of each of the 10 agents in the network, the 10 bars in each histogram represent the number of copies of the initial unit parcels from each of the 10 agents they have received. The copies of each agent's own initial information are shown in red.