



Argumentation, cognition, and the epistemic benefits of cognitive diversity

Renne Pesonen¹

Received: 3 February 2022 / Accepted: 17 June 2022

© The Author(s) 2022

Abstract

The social epistemology of science would benefit from paying more attention to the nature of argumentative exchanges. Argumentation is not only a cognitive activity but a collaborative social activity whose functioning needs to be understood from a psychological and communicative perspective. Thus far, social and organizational psychology has been used to discuss how social diversity affects group deliberation by changing the mindset of the participants. Argumentative exchanges have comparable effects, but they depend on cognitive diversity and emerge through critical interaction. An example of a cognitive psychological theory is discussed that explains how mutual reasoning affects how we think, make decisions, and solve problems, as well as how cognitive biases may facilitate an efficient division of cognitive labor. These observations are compared with the existing results in the social epistemology of science. Moreover, I explicate the conceptual differences between the distributed and social processing of information. While argumentative exchanges belong to the latter domain, most existing simulations model distributed processing, which may compromise their real-world relevance and proper conceptual interpretation. However, I aim not to criticize the existing simulation methods but to promote an approach from the cognitive psychology of reasoning that complements the current use of organizational psychology and computer simulations by investigating a different set of mechanisms relating to similar phenomena of interest in the social epistemology of science.

Keywords Social epistemology · Diversity · Argumentative theory · Information elaboration · Distributed cognition

1 Introduction

Argumentation is crucial for scientific practice not only for establishing the conclusions of research but for negotiating what goals to pursue and by what means and how

✉ Renne Pesonen
renne.pesonen@tuni.fi

¹ Faculty of Social Sciences, University of Tampere, Kalevantie 4, 33100 Tampere, Finland

to interpret empirical results, theoretical concepts, and scientific arguments themselves. In some theoretical fields, such as philosophy, arguments basically constitute the identity and subject matter of theories. Still, especially formal approaches in the social epistemology of science largely ignore argumentation even if they seek to understand the mechanisms of consensus formation, collective intelligence, and the division of cognitive labor.

One way to approach diversity in socially distributed cognitive labor is to investigate a community of independently working individuals who are specialized in different tasks and exchange their results through established information channels. This approach has been taken, e.g., in epistemic network models (Zollman, 2010). However, the emergence, stabilization, and evolution of such networks necessitates mutual understanding, trust, and some agreement on the norms concerning the shared conduct. In argumentative exchanges, different ideas and people interact to establish these affairs, and hence they are an obvious place to seek the mechanism where cognitive and social diversity make their impact on communal epistemic processes.

Social diversity can be defined at least by diversity in social values and interests or by diversity in so-called social locations (e.g., Intemann, 2010). I use the term primarily in the latter meaning where locations refer to a membership in groups that are perceived to be relevant for social identities. Often this means relatively stable demographic categories, such as gender, class, and ethnicity, but such identities can result from any attribute people use to tell some people apart from others. In this broad sense, values and interests may also mark different groups and locations. Relevant categories and attributes presumably vary across groups, contexts, and persons, but for the current discussion the relevance of such categorization is that perceived in- and out-group demarcations modulate trust and conformism and expectations of differences in opinions, experiences, and other factors pertinent to cognitive diversity (see Carter & Phillips, 2017; Fazelpour & Steel, 2022).

In addition to argumentation, another somewhat absent element in the literature is cognitive psychology. This is understandable given that cognitive psychology has traditionally focused on the knowledge processes of individuals. These two omissions are conceivably related if argumentation is seen as a fine-grained cognitive activity where the social element pertains only to the reporting of the reasoning behind our beliefs—that is, the fixation of belief is considered as a private, explicit, and rational process, and argumentation is the selective reporting of that process to others. If this is correct, argumentative exchanges may be reasonably treated simply as exchanges of information that leave the actual private reasoning process intact. However, this simplified view is mistaken, as will be elaborated throughout this paper.

We sometimes do report the actual reasoning behind our attitudes and decisions, but a good deal of evidence shows that often the reasons are opaque to ourselves and their elaboration is largely rationalization. Rationalization is sensitive to intersubjective norms, i.e., we explain our opinions and decisions in terms we think makes sense to others. This has been observed clearly in research on social psychology and choice behavior (Wilson, 2002), but also on deductive reasoning (Evans & Wason, 1976; Evans, 1996), political attitudes (Strandberg et al., 2018), and moral cognition (Haidt, 2001), to name a few. Rationalization is not a simple reporting process but a thinking process that tends to solidify attitudes, at least if they go unchallenged, and this can

happen even with opinions we actually do not hold or did not hold initially (Barden & Tormala, 2014; Strandberg et al., 2018). Rationalization is often initiated by social needs to make our putative reasons explicit, which makes them a subject of public review and enables feedback to operate on them and our decision processes. It also makes our reasoning (or lack thereof) explicit to ourselves. Argumentation is a social activity where we mutually explicate, judge, and adjust the rationalities of our beliefs and actions and negotiate legitimate points of disagreement.

To the best of my knowledge, there are no good reasons to suppose that the reasoning of scientists fundamentally differs from that of the general population. While discipline-dependent decrease in formal reasoning errors are reported, errors are still prevalent (Tversky & Kahneman, 1983; Griggs & Ransdell, 1986; Jackson & Griggs, 1988) and scientific training and expertise does not reliably facilitate logical reasoning strategies (Mercier & Heintz, 2014) or eliminate common biases with psychological or social origins (May, 2021). Indeed, perhaps the best known research program on reasoning biases originated from the perceived failure of psychologists to apply sound statistical principles (Kahneman & Klein, 2009). Case studies in the history of science also reveal that even highly acclaimed scientists are at risk of succumbing to epistemologically flawed practices after retreating from academic participation into solitary research (Mercier & Sperber, 2017).

With this in mind, I discuss in Sect. 3 a cognitive theory that rejects the rationalistic and individualistic understanding of human reasoning and construes it as inherently social instead. This is the *argumentative theory of reasoning* pioneered by Mercier and Sperber (2011, 2017), which holds that reasoning is closely related to communication, trust, and persuasion. While others have also emphasized the interactive and social dimensions of argumentation (e.g., the pragma-dialectical approach; see van Eemeren et al., 2014, Ch. 10) the novel idea in Mercier and Sperber's theory is that human reasoning is argumentative in function. It erodes clear distinctions between individual and social reasoning and systematizes swaths of reasoning studies in a way relevant to matters of trust, diversity, and the division of cognitive labor, which makes it a particularly useful example of a fertile contribution of cognitive psychology to social epistemology.¹

The last two sections compare and integrate the epistemic import of argumentative theory with similar results derived from organizational psychology and agent-based simulations. To explicate the relations of these three approaches, below I map their respective types of mechanisms: (1) social psychological mechanisms that affect the

¹ As an anonymous reviewer pointed out, formal methods and rigorous methodological conventions arguably make scientific reasoning different and more reliable than reasoning in non-scientific contexts. That is not disputed here. Instead, the premise here is that critical interaction among scientists (including implied as well as actual argumentative contexts) corrects reasoning more reliably than scientific training as such. Moreover, while formal methods are crucial in making reasoning more precise and complex inferences more reliable, they do not make scientists better reasoners overall. Theoretical disputes often necessitate complex inferences that integrate patterns of empirical results and theoretical arguments. This process is potentially vulnerable to confirmatory and other reasoning biases (May, 2021), even if the results that factor into such complex evaluation may not be. The events where formal methods help to conclusively establish or refute a particular claim can be construed as particularly strong cases of argumentation—unless a flaw in their choice or application is exposed. In any case, the scope of the current discussion pertains only to selective parts of scientific reasoning, namely to those where argumentative evaluation of theories, analyses, and empirical results are called for.

epistemic performance of groups through the participants' responsiveness to social information, (2) modular distributed processing that is modeled in many agent-based simulations investigating the epistemic impact of socially mediated information, and (3) socially implemented processing where information is processed and refined in the interactions of agents. Type (3) mechanisms are the main focus of this article, and argumentative exchanges are their prime example. They differ from type (1) mechanisms by a different psychological basis and specific implications for group deliberation, and they differ from the type (2) mechanism in that they are not solely communication structures affecting individual decision making but socially extended reasoning processes, as detailed in what follows.

2 The social and modular distributed processing of information

In this section, I address the distinctions between *modular distributed processing* and two meanings of social cognition: the processing of *social information* and the *social processing* of information. I elaborate this last notion in particular, which, while familiar in content, seems to be missing in the existing literature and gets conflated with socially distributed processing. The analytical value of keeping these concepts separate is in clarifying the overlapping mechanisms that contribute to collective intelligence and mediate between socially distributed and private cognition.

By (*modular*) *distributed processing*, I refer to a perspective on cognition as distributed among individuals, groups, institutions, and technology. In principle, it may remain neutral on what the processing elements are (e.g., humans, machines, groups) and disinterested in how the components internally function. It deals with the (self-)organizing properties of target systems, the aggregate outputs of their components, and the flow of information between the components. My characterization captures aspects of, e.g., Hutchins' (1995) notion of distributed *cognition* but without references to the historical and cultural scaffolding of cognitive processing. Some philosophers of science use the notion broadly with connotations not intended here. For example, Giere (2002) discussed some aspects that are usually considered as examples of extended cognition in the philosophy of mind literature and some that include social interaction as an explanatory component of cognition.

My use of the notion of distributed *processing* is intentionally more restricted. While it includes informational aspects of social interaction, it excludes those aspects that are psychologically sensitive to the social nature of the interaction. These are discussed separately below. In my usage, the key aspect is distributed processing with modular processing units, i.e., units or agents that exchange information but do not affect the internal processing of each other. Parallel distributed processing networks (Rumelhart et al., 1986) are a prime example of such systems. Epistemic networks (e.g. Zollman, 2010) also satisfy this definition, as their units of information production and processing are independently operating agents. The processing of epistemic networks is distributed in the sense that the output we are interested in is the belief state(s) of the whole community instead of individuals. To avoid possible confusion with the more encompassing notion of distributed cognition, I use the expression "*modular distributed processing*" where appropriate.

The majority of agent-based simulations in social epistemology fall into this category. In addition to epistemic networks, illustrative examples are (Weisberg & Muldoon, 2009; Hong & Page, 2001, 2004) as the latter authors in particular are explicit in that it does not matter whether the agents in their simulations are interpreted as humans or, e.g., computer programs and that a collection of agents effectively constitute a single problem solver. The agents do not engage in interactive or collaborative processing. They simply operate on the outputs left by others. Weisberg and Muldoon (2009) simulated epistemic communities that included agents whose decision-making was sensitive to what others do but not to the reasons for their behavior. I count this as a modular distributed processing model because human decisions are guided (sometimes toward choices of less intrinsic value) by available justificatory reasons (Shafir et al., 1993) that often have a basis in social interaction and values (Stanovich, 2013). Hence, for better or worse, the model replaces a core feature of socially based decision-making with a preset decision strategy. Although reason-based choice is mostly documented in economic choice behavior (e.g. Shafir et al., 1993), it would be remarkable if it turns out to be an irrelevant factor in the decision-making of working scientists, as argumentative reasoning is generally conceived as a key research-guiding practice in science.

The *processing of social information* means responding selectively to perceived social attributes and situations, statuses, group identities, and, in some cases, social macro-structures. Social psychology in particular deals with these matters. In social epistemology, the studies concerning the effects of social diversity and status on trust and group deliberation tend to fall into this category. Good examples are Steel et al. (2019) and Fazelpour and Steel (2022), who investigated how perceived demographic attributes modulate trust and group performance. The former study reviews research on how responsiveness to identity-relevant social attributes affects *information elaboration*, which is defined as any group process wherein participants communicate and integrate cognitive resources dispersed within the group, such as knowledge and reasoning heuristics. Below, I outline a specific form of such an elaboration process, which I call the *social processing of information*. The next section details its underlying cognitive mechanisms, which turn out to be affected primarily by cognitive instead of social information and diversity.

The social processing of information is somewhat difficult to distinguish from the modular form of socially distributed cognition conceptually and responsiveness to social information empirically. I address the former issue first. Socially distributed information processing is obviously social processing of information at least in some sense. However, there is nothing intrinsically social in the processing as such if it is only an aggregate of individuals working on their own, even if they utilize socially mediated information. Here the notion of the social processing of information has the more specific meaning that the social interaction implements or facilitates cognitive processing that necessitates the social interaction as an explanatory component.

People are fluent in explicating reasons for their beliefs, and this explication is often initiated by the requests of their peers. The social dynamics of these situations make us process different information (i.e., justificatory reasons and tacit background assumptions) in different ways (in an explicit, metacognitive way) in comparison to what we are predisposed to do alone. This prompted explication, by itself, is not social

processing of anything but an initiation to, and a basic constituent of, social processing of information that makes us reason in distinct ways for social engagements. Several studies show that in collaborative settings people tend to use reasoning strategies and reach conclusions that would be unlikely for any of the participants working in isolation (e.g. Moshman & Geil, 1998; Simonton, 2003; Bearman et al., 2007; Maciejovsky & Budescu, 2007; Mercier & Sperber, 2011; Trouche et al., 2014).

The currently famous extended cognition theory holds that cognition is not confined to the brain but partly distributed and realized in our interactions with the environment (e.g. Clark, 2008). Paper-and-pencil calculation is the standard example. The social processing of information can be conceived as a species of extended cognition where our cognitive processing is distributed into the social environment and supported and constrained by social interaction. Importantly, as with any extended processes, the cognition of the participants is coupled through their interactions in a way that it cannot be fully analyzed without the interactive process in which it partakes. In clear cases of modular distributed processing, a division of cognitive labor exists that allows a functional decomposition of the collective processing so that the individual participants carry out their own tasks, make decisions based on their own deliberations, and treat the information they produce and receive similarly. This contrasts with clear cases of social processing, where the social interaction is an explanatory component of the information processing of a group and its members.

Thus, suppose you advocate theory T1 and I advocate its rival T2, and we agree that experiments E1 and E2 will settle which theory is better. You conduct E1 and I conduct E2, after which we combine the results and see who was right. This is the elementary setup of influential simulation studies in social epistemology (epistemic networks, in particular) and more clearly a case of distributed rather than social processing. The deliberation that made us agree on what experiments will resolve our disagreement, or to notice that our theories are rivals and commensurable in the first place, could be (and probably often is) an example of the social processing of information. One could characterize the difference as reasoning in parallel versus reasoning together; however, the gist is more precisely in the differences in cognitive processing these two settings entail on both the individual and group level.

In addition to conceptually distinguishing social from merely socially distributed processing, the second problem was to empirically tell apart the social processing of information from the epistemic effects of responsiveness to social information. Reasoning in groups is virtually always accompanied by the concurrent processing of social information. Mere situational awareness makes us responsive to the demographic attributes of the people we are dealing with as well as their reputation, status, group affiliations, and so on. The processing of social information may happen inattentively but it affects our behavior and cognitive processing and performance in groups (Carter & Phillips, 2017). This entanglement of social cognition with mutual task processing makes it difficult to tell them apart empirically. However, the conceptual difference is straightforward: responsiveness to social information by individual members versus processing (any) information together. As an example of this entanglement, the mere presence of out-group members can temper excessive in-group trust and conformism and help participants to evaluate information more critically and elaborate their minority views to others (Steel et al., 2019).

The references above show that considerations of diversity and the social processing of information can be found in the existing literature. However, these studies tend to concentrate on the information exchange aspect instead of mutual reasoning. In particular, Fazelpour and Steel (2022) model this process as an asymmetric information uptake from in- and out-group members. They utilize epistemic network formalism designed to simulate information flow among scientists, i.e., distributed rather than social processing in the sense meant here. The effects investigated by Fazelpour and Steel can be considered as borderline cases of social processing, as the sole mechanism they tap into is the selective attenuation of information depending on the social source, but otherwise the cognitive process remains unaffected both on the individual and group level.

The distribution and uptake of information is certainly an important mechanism in collective cognitive labor, but it is only one mechanism among others, and "information" here refers solely to binary empirical outcomes obtained by applying a theory (i.e., success or failure). Zollman's (2010) main finding was that limiting the social distribution of information can be beneficial because it prevents the community from locking prematurely into a wrong theory due to early promising evidence. In Fazelpour and Steel's model, a similar effect results from a selective distrust of out-group members. Frey and Šešelja (2020) implemented it by adding inertia to theory change in the form of time-lags and thresholds for the likelihood ratio a competing theory needs to cross before individual scientists switch theories.

Unfortunately, Kummerfeld and Zollman (2015) demonstrated that these results depend on the arguably questionable assumption that individual scientists explore theoretical alternatives only through social influence. Nevertheless, there are several mechanisms that lead to the same group-level epistemic benefit, and what they have in common is that they resist theory switch at the individual level (i.e., individual scientists resist the said influence). As Zollman (2010) pointed out, his main result can be secured also by issuing strong priors that bias scientists to stick initially with their favorite theory despite the mounting evidence to the contrary. This kind of obstinacy may be irrational unless it results from good justificatory reasons that make the transient discounting of evidence reasonable. Evidence needs to be accepted on top of being received, and scientific debates often revolve around what conclusions the presented data actually supports and why. Such epistemic vigilance is essential, especially in the face of publication bias, i.e. the widespread practice of omitting the publication of null results, which makes it virtually impossible to estimate the actual success rate of theories.

Nevertheless, currently there have been only a few attempts to incorporate argumentation or a critical feedback mechanism into network simulations (e.g., Borg et al., 2018; Frey & Šešelja, 2020). In contrast to Zollman (2010), Frey and Šešelja found no trade-off between speed and reliability in a community's convergence to the correct theory. This was largely due to adding "critical interaction" among agents in addition to circulating experiment outcomes. However, they assumed that feedback is always truth-conductive. This may be true *on average*, but the same presumably holds for experimental results, too, and as we see in the next section, people tend to be cognitively biased in their critical interactions. Hence, criticism can be misleading just like evidence, and just as it is important to study the social dynamics of infor-

mation exchange, it is important to understand the social dynamics of argumentative exchanges to grasp how individual and group rationality are related and what factors drive consensus formation, the division of cognitive labor, and so on. Next, I turn to the argumentative theory of reasoning, which addresses these issues and exemplifies the concept of the social processing of information.

3 On the argumentative theory of reasoning

Why do humans reason? Obviously because reasoning enables us to make better decisions and improve our knowledge. On closer examination, however, the answer seems less evident. A wealth of evidence shows that most people most of the time violate practically all normative models of rationality, from logic to decision theory to probability calculus and more (e.g. Evans et al., 1993; Gilovich et al., 2002). These problems are not due to random errors or cognitive limitations as highly predictable errors loom large even with trivial reasoning tasks and formal education and expertise do not eliminate them reliably. Also, our reasoning strategies tend to be deficient. We evaluate and seek information in biased, self-serving ways that prevent rather than help us to change our minds. We are fluent in producing justificatory reasons for our decisions but bad at judging if they are actually sound and relevant.

Some reasoning experiments (e.g. Evans, 1996) reveal that when subjects try to solve an unfamiliar problem, they do think of their solutions but that does not usually change their initial (and often faulty) responses. Instead, the subjects clearly spend their time in producing *post hoc* justifications for their intuitive initial decisions. Of course, sometimes we successfully reason ourselves out of unexpected trouble, but this may be due to creative rather than logically regimented cognition. Be that as it may, the fact still remains that explicit reasoning often follows decisions and not the other way around, and when deliberation guides our choices, it surprisingly often points in the wrong direction: those who deliberate more are in many instances less satisfied with their choices in the long term (Wilson, 2002). Thus, human reasoning seems to serve poorly—or even in opposition to—its supposed function to correct our potentially deficient beliefs, attitudes, and decisions.

However, these problems are documented mostly in laboratory settings where individuals reason alone, and many of them are greatly attenuated when we reason together. We suddenly become good at evaluating the logical consistency of arguments and finding their weak points, utilizing falsificatory strategies, and so on. When small groups solve standard reasoning tasks together, they tend to converge to the correct solution at a significantly higher probability than would be expected if the performance resulted from some member knowing the correct solution and then convincing the others. The observed group behavior instead supports the alternative explanation that the correct solution often emerges in group deliberation (Moshman & Geil, 1998; Trouche et al., 2014).

Mercier and Sperber (2011, 2017) interpret this peculiar pattern of results not as implying that human reasoning is inherently deficient but that its function is traditionally misconstrued. According to their *argumentative theory of reasoning*, reasoning has evolved for the purposes of social interaction and improving communication instead

of knowledge. Most of the time we rely on trust and accept socially mediated information at face value, but blind trust makes communication unreliable and the recipients helpless victims of misinformation and deception. Other people are immensely valuable for information and collaboration, and both of these social affordances necessitate reliable communication. We can evaluate the credibility of socially shared information by its probability against relevant background knowledge or by the reputation of the information source, but such knowledge is not always available, and then the standard way to validate claims is to ask for justifications. These justifications may be further addressed and the process iterated until reasonable trust in the claim is either established or rejected.

According to the argumentative theory, human discursive reasoning has evolved to serve this function and it has two main components: persuasion and information verification. Most things we probably learn unreflectively by doing, by hearsay, and so on. But as we invite others to join our practical engagements or try to influence communal decision making, we must be willing and capable of producing justificatory reasons as other people are naturally inclined to ask for them to verify if our input is reliable. In doing so, we do not need to address our explanations in a self-critical manner. We just do our best to address the inquiries and leave the criticism to our peers. However, mere rhetoric will not do, at least in the long run. For the information validation to serve its purpose, recipients must be capable of at least following the obvious implications and checking the coherence of the presented information and its consistency against background knowledge. That is, persuasion attempts must be coherent enough to survive critical examination to actually persuade or else the verification function would fail to serve its purpose.

This creates a cognitive asymmetry toward our own beliefs and beliefs held by others. The signature effects of this asymmetry are *confirmation bias* in producing reasons and *belief bias* in the evaluation of arguments. Confirmation bias is the tendency to seek confirmatory evidence and dismiss falsificatory reasoning strategies toward one's own opinions and hypotheses. In various reasoning and decision-making studies subjects fail to notice how irrelevant and incoherent their rationalizations are (Evans & Wason, 1976; Wilson, 2002), showing that, if unchallenged, we are often not very reflective on the reasons we report. Belief bias means that we are relatively blind to logical errors in arguments that purport to support what we already believe, but we easily spot the errors in formally identical arguments when we do not believe the conclusions.

This may seem to paint a bleak image of human cognition, as it appears we are predisposed to hold our opinions uncritically and attack any counteracting arguments. However, when reasoners congregate in argumentative exchanges, the participants do their best to produce supporting arguments for their opinions and, crucially, a critical feedback mechanism emerges as they actively seek the faults in arguments presented by others (Mercier & Sperber, 2011). Considered from the information processing perspective, this means that an interactive search ensues to discover all the strengths and weaknesses of the opinions and arguments brought into the situation. Given that all the viewpoints receive a fair hearing, the argumentative dynamics of the group hence approximates to the classical ideal of a rational reasoner. If the intersubjectively established feedback mechanism is severed, we lapse into our default biased ways

of reasoning, potentially even as a group. However, assuming (as Mercier & Sperber does) that we are capable of producing, identifying, and accepting sound reasons, truth tends to win in group deliberations—at least eventually and under ideal conditions. The evidence is strongest with problems that have a demonstrably correct answer, and the ideal conditions include sufficient level of diversity to prevent groupthink.

Mercier & Sperber also propose that the cognitive biases may contribute to an efficient division of cognitive labor as a by-product—an idea proposed earlier by Solomon (1992). In complex epistemic enterprises, it is taxing for an individual to investigate all the possible theoretical alternatives in a thorough manner. Arguably, it is more efficient if the effort is distributed among individuals who allocate their effort to studying their preferred theory deeply and making the best case out of it. These cases are then pooled and evaluated in mutual argumentative exchanges where their faults are exposed. The authors do not elaborate this idea much further, but others have subsequently endorsed it and even proposed that this presumed group benefit is the main adaptive function of reasoning biases (see Peters, 2020). Be that as it may, while the idea is suggestive, it is not obvious that the biases of individual reasoners actually contribute to a decent division of cognitive labor. However, simulation studies may prove useful here, and the logic behind the idea is highly analogous to the result discussed in the previous section, i.e., that individuals' resistance to theory switch may benefit the epistemic community.

Recently, at least two agent-based network simulations of argumentative exchanges have been introduced (Gabbriellini & Torroni, 2014; Borg et al., 2018). Both utilize the same abstract argumentative framework that represents theories as networks whose structure the agents gradually discover. Part of the structure consists of attacks against rival theories and counterattacks that defend them from rivals. In (Borg et al., 2018), agents periodically communicate what they have found and they may either switch to a better defended theory or stay with the current one to discover more arguments, attacks, and defenses. The setup thus differs from Zollman's epistemic networks in that the agents share argumentative structures instead of experimental outcomes. In contrast to Zollman, Borg et al. found that increased information sharing mainly led to a faster and more reliable convergence to the best theory. However, in small, densely connected populations, false positives hamper the progress if the agents communicate only those pieces of information that support their theory. This parameter may be relevant if confirmation bias predisposes scientists to dismiss counterarguments they discover as irrelevant. The other model (Gabbriellini & Torroni, 2014) is directly influenced by Mercier & Sperber, and it accommodates dynamically changing trust as a factor of argument acceptance. Unfortunately, neither of these studies investigate if the resistance to switch affects the division of cognitive labor as proposed above.

In summary, the widely documented peculiarities of human reasoning and their explanation by the argumentative theory makes the prime empirical case for the social processing of information, and it does it in a way that has several links to focal issues in social epistemology. Importantly, the explanation pinpoints actual mechanisms that cross individual and group-level cognition. The key point is that information processing is qualitatively different between people in contrast to individual reasoners. Information validation implements a critical feedback mechanism by facilitating falsificatory reasoning strategies, and confirmatory biases resist premature opinion switch

and elevate the changes that all the viewpoints are properly addressed. In combination, these mechanisms make group deliberation more objective and rational in the classical sense of tracking the truth or at least the best available argument. However, this "objectivity" is actually intersubjectivity, which is sensitive to both social and cognitive diversity as detailed below.

4 The mechanisms of social information processing and the epistemic value of diversity

In the dual pathway model by Carter and Phillips (2017), social diversity is proposed to affect group performance via two independent but counteracting mechanisms: (1) The recognition of social diversity cues expectations of cognitive differences, and even when these expectations are not accurate, they may facilitate the more open-minded and thorough processing of task-relevant information. (2) The same process can also activate group biases, distrust, and a competitive mindset, resulting in a negative impact. However, as argued by Fazelpour and Steel (2022), this latter pathway can yield to a third mechanism, where (3) mitigating excessive trust benefits the group by making the participants more critical toward the information they receive and more willing to articulate dissent. All of these mechanisms are sensitive to social diversity and, hence, are psychologically rooted in the processing of social information.

Mercier and Sperber (2011) make comparable observations associated with cognitive diversity. The critical feedback mechanism does not emerge in groups that are too like-minded because there is little dissent to be expressed. Thus, its functioning necessitates some degree of cognitive diversity. However, too much diversity may cause problems in mutual understanding; the communicative process breaks down and no consensus or persuasion results. In addition to a sufficient level of cognitive diversity, this mechanism causally depends on the above condition (3), i.e., a sufficient level of trust so that participants are willing to cooperate but not accept claims at face value.

Thus, both cognitive and social diversity have similar effects on group deliberation: No diversity, no disagreement, and no critical feedback; but too much diversity erodes trust and mutual understandings and prevents the convergence of opinion. However, the effects of social and cognitive diversity stem from different mechanisms—responsiveness to social information and the social character of human reasoning, respectively. Well-placed trust is an important factor in both, and it can be injected by a variety of means, either cognitively by displaying sound reasoning and competence on the subject matter or socially through status and reputation, authority, affiliation, and so on.

The key mechanism that produces epistemic benefits from cognitive diversity is mutual corrective feedback, which counteracts belief bias and shifts the focus on selected kinds of task-relevant information, such as justificatory reasons, tacit background assumptions, and alternative interpretations of data. The resulting argumentative exchanges are structured by the participants' interaction, which guides the production, search, and selection of information. Hence, the outcome is the social processing of information, where the processing unit is the interacting group. This is

to be differentiated from merely socially distributed processing, where information is passed between agents, who utilize it on their own, and agents can be conceived as autonomous information processing modules who are sensitive to socially distributed signals but whose processing is not guided or supported by social interaction.

These considerations cast doubts on, e.g., Hong and Page's (2004) "diversity trumps ability" theorem and its relevance to science. In their simulation, each agent applies its abilities individually in a dedicated time slot and the next agent continues from there. As the authors note, a collection of agents is effectively a single problem solver armed with the abilities of several agents. Their simulation has been criticized on conceptual, formal, and methodological grounds (Thompson, 2014; Reijula & Kuorikoski, 2021), but from an empirical perspective we can also add that the group interaction of cognitively diverse agents has the capacity to result in mutually corrective processing that transcends the cognitive repertoire of the participants individually. How highly skilled individuals actually perform in groups in comparison to more cognitively diverse but less competent ones depends obviously on the group but also on the task (Hill, 1982) and arguably also on how the feedback mechanism psychologically activates and operates.

A good candidate for the mechanism comes from the *default-interventionist* dual-process theory of reasoning (Evans & Stanovich, 2013), whereby human cognition consist of control and inference systems that activate on error detection. Most of the time, we carry out our routines unreflectively and make decisions intuitively. When a conflict emerges, the control mechanisms intervene in our behavior to halt the ongoing activity and shift our attention to analyze the unexpected trouble. Apparently, mere requests to reflect and explicate the reasons of our actions do not reliably activate analytical thinking but simply create a new task to produce justificatory reasons. It is precisely the conflict with expectations that activates reasoning for conflict resolution, and counterarguments to our confabulations may function as a special case of conflict that forces us to think more thoroughly what we are saying and doing as we are trying to convince others.

If this is the case, like-minded groups of cognitively similar agents may effectively function much like a single problem solver because the lack of conflict fails to activate analytical reasoning as the exchanges within the group unfold along the mutually expected tracks. This does not prevent groups from engaging in discursive exchanges, but the mutual elaboration of reasons may simply solidify existing attitudes and the lack of critical thinking may go unobserved as the corrective mechanism fails to activate. Cognitively diverse groups, however, may enjoy cognitive benefits that stem from the interplay of basic control mechanisms of human cognition and the function of argumentative reasoning.

Hence, we have an empirical argument from cognitive psychology according to which diversity may indeed trump ability. The real-world relevance of this version of the argument remains to be seen, but its logic is quite different from Hong & Page's, which relies on the breadth of available cognitive heuristics. Instead, the present argument relies on how in the presence of cognitive diversity argumentative exchanges make participants counteract each other's belief biases which plague the evaluation of arguments from accepted data or premises to their putative implications. Nevertheless, groups of skilled experts may enjoy other advantages, such as the more fluent use

of remote analogies for creative problem solving in comparison to groups with less domain competence, but the evidence on this is mixed (Bearman et al., 2007).

Lastly, these considerations suggest a topic for further investigation: If groups benefit from the presence of cognitive diversity, does this also hold for groups of groups? That is, if we have groups consisting of cognitively similar individuals, what happens when cognitively different groups engage in mutual reasoning? Do they still reap the benefits of critical interaction, or do in-group biases prevent the feedback mechanism from correcting the participants? If individual reasoning biases facilitate the effective distribution of cognitive labor, do conformist group biases have the same effect when groups congregate? Can social similarity forge trust between individuals across groups separated by cognitive diversity? These questions are relevant especially for understanding the social epistemology of multidisciplinary research where common ground is presumably slim and cognitive diversity wide.

The distinction between social and modular distributed reasoning may help in selecting formal frameworks for simulations of cross-team collaboration. In multidisciplinary research, the fundamental setup is different from the standard epistemic networks in that different agents or teams do not work competitively to figure out the best theory among several rivals. Instead, in multidisciplinary research the division of cognitive labor means that different teams bring their competencies into different phases and aspects of research to pursue a common complex goal. Hence, in this case the collaboration structure arguably resembles more of Hong & Page's simulations than epistemic network models. The convergence of opinion is not the goal but the optimal solution, and all the participants need not even understand the outcome in detail.

5 Conclusion

I have argued that the social epistemology of science could benefit from research on argumentation and human reasoning more than currently seems to be the case. Argumentation is a key scientific practice that affects how data is utilized and how individuals think and make decisions, and its functioning needs to be understood from an empirical psychological and communicative perspective. However, critical feedback is only one of the mechanisms that emerge in group interaction. For example, the use of analogies seems to be facilitated in group problem solving (Bearman et al., 2007). Analogies may be cued for communicative needs, but this may also yield unintended benefits, as analogical reasoning is widely associated with creative thinking, such as forming novel hypotheses. Therefore, insights from cognitive psychology more generally may help in identifying the relevant mechanisms and epistemic benefits of collective intelligence. It may also help in interpreting and empirically calibrating the existing simulation models, and utilizing the resources of cognitive psychology does not mean that we need to model the complex cognitive processes of individual minds.

Simulation models of argumentative exchanges are already available as well as a readily useful cognitive psychological literature. Hence, we need not start from scratch to integrate these elements into the ongoing research. Nothing in this paper

hangs on whether the argumentative theory of Mercier & Sperber is correct (at least in detail). Their theory is simply a suggestive way to systematize swaths of research on human reasoning in a way that is highly relevant to the current topics in the social epistemology of science. My choice to discuss simulation studies in this paper is mainly due to them constituting one of the main currents in the ongoing research, and they make it particularly explicit what mechanisms are supposed to operate in collective cognitive labor and how these mechanisms are supposed to function.

I elaborated the notion of the social processing of information, which I practically equated with argumentation but which is potentially a more general aspect of socially extended cognition pertaining to any reasoning process that is guided and implemented by collaborative interaction. It differs from the related notion of information elaboration by being slightly more constrained, and its underlying mechanisms differ from the group effects that result from affecting the mindset of the participants through responsiveness to social information. The other contrasting concept, especially important in connection with simulation studies, is modular distributed processing. The distributed processing framework is certainly a legitimate way to conceptualize and model the mechanisms of collective intelligence. I do not claim that every simulation study needs to concern with the dynamics of mutual deliberation, but the communicative and collaborative structure of the target systems should be analyzed perhaps more carefully in order to make informed decisions as to when a mere distributed framework suffices and when it limits the real-world interpretation of the results either empirically or conceptually.

As an example of the latter, some epistemic network simulations show that resistance to switch theory despite evidence may benefit the epistemic community. This obstinacy can be construed as irrationality that paradoxically yields epistemic benefits. However, Bayesian rationalists are not sensitive merely to evidence but also to priors, which may be affected by justificatory reasons and theoretical arguments, and data can be conceived just as a special case of argument. The argumentative theory of reasoning suggests that individuals tend to stick with their opinions until they run out of counterarguments or reasons to argue. This may be a bias, but not a result of extreme belief but intellectual resistance. Nevertheless, this observation is not antagonistic to—but rather yet another way of discovering—the common message from several simulation studies that the rationality of a group can be independent of the rationality of its members (Mayo-Wilson et al., 2011; Grim et al., 2019).

Helen Longino's (1990) influential account of objectivity in science overlaps markedly with themes discussed in this paper. According to Longino, objectivity comes in degrees and it is a characteristic of community's practice rather than of individual scientist's thought processes. Objectivity is secured through critical discussion that exposes and corrects background assumptions. Critical interaction does not guarantee that we arrive at truth but it weeds out personal preferences in how data is interpreted and conclusions accepted. Widely shared background assumptions may still remain hidden and insulated from critical evaluation. I have reviewed evidence that further support these ideas and in particular that group deliberation makes our reasoning more objective. I have characterized the resulting notion of objectivity as intersubjectivity, which I believe is consonant with Longino's account of objectivity

not as an observer-independent truth but as the critically achieved consensus of the scientific community.

As an extension to Longino's account, I have discussed evidence that the corrective effect of argumentative interaction does not only pertain to tacit background assumptions but also to basic psychological reasoning mechanisms which produce and evaluate justificatory reasons in predictably biased ways. Longino lists several preconditions for objective inquiry, namely that there must be recognized avenues of criticism and shared standards that critics can invoke. Moreover, the community as a whole must be responsive to criticism and intellectual authority must be shared equally among qualified practitioners. The first pair concerns the conditions that ensure that orderly critical interaction will routinely take place. The latter pair ensures that all the viewpoints get properly addressed and the critical discussion has the intended corrective effect on the participants. I have discussed results that indicate that group dynamics depends not only on explicit norms and institutional arrangements but also on social psychological factors that may be difficult to spontaneously recognize and control.

Lastly, the link between social epistemology and cognitive psychology should not be conceived as a one-way street. Computational models have always been a core method in cognitive sciences, and how social epistemologists apply formal methods from economics, ecology, and computational sociology to investigate collective intelligence may prove useful for the social fronts of cognitive psychology. Even if the simulations are not high-fidelity models of their target systems, they may provide useful optimality models and hypothetical mechanisms for rational belief and socially based decision-making.

Acknowledgements I would like to thank Kristina Rolin, Inkeri Koskinen, Jaakko Kuorikoski, Samuli Reijula, and Pasi Valtonen for helpful discussion about the paper. This work has been funded by the Academy of Finland research project Social and Cognitive Diversity in Science: An Epistemic Assessment.

Funding The Academy of Finland.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barden, J., & Tormala, Z. L. (2014). Elaboration and attitude strength: The new meta-cognitive perspective. *Social and Personality Psychology Compass*, 8(1), 17–29.
- Bearman, C., Ball, L. J., & Ormerod, T. C. (2007). The structure and function of spontaneous analogising in domain-based problem solving. *Thinking & Reasoning*, 13(3), 273–294.
- Borg, A., Frey, D., Šešelja, D., et al. (2018). Epistemic effects of scientific interaction: Approaching the question with an argumentative agent-based model. *Historical Social Research*, 43(1), 285–307.
- Carter, A. B., & Phillips, K. W. (2017). The double-edged sword of diversity: Toward a dual pathway model. *Social and Personality Psychology Compass*. <https://doi.org/10.1111/spc3.12313>.
- Clark, A. (2008). *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford: Oxford University Press.
- Evans, J. S. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, 87(2)
- Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum Associates.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241.
- Evans, J. S. B. T., & Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, 67(4), 479–486.
- Fazelpour, S., & Steel, D. (2022). Diversity, trust, and conformity: A simulation study. *Philosophy of Science*, 89(2), 209–231. <https://doi.org/10.1017/psa.2021.25>.
- Frey, D., & Šešelja, D. (2020). Robustness and idealizations in agent-based models of scientific interaction. *The British Journal for the Philosophy of Science*, 71(4), 1411–1437.
- Gabriellini, S., Torroni, P. (2014). A new framework for ABMS based on argumentative reasoning. In: Kamiński B, Grzegorz K (Eds.), *Advances in social simulation: proceedings of the 9th conference of the european social simulation association* (pp 25–36). Springer, Berlin
- Giere, R. (2002). Scientific cognition as distributed cognition. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 285–299). Cambridge: Cambridge University Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.
- Griggs, R. A., & Ransdell, S. E. (1986). Scientists and the selection task. *Social Studies of Science*, 16(2), 319–330.
- Grim, P., Singer, D. J., Bramson, A., et al. (2019). Diversity, ability, and expertise in epistemic communities. *Philosophy of Science*, 86(1), 98–123.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Hill, G. W. (1982). Group versus individual decision making: Are $n + 1$ heads better than one? *Psychological Bulletin*, 91(3), 517–539.
- Hong, L., & Page, S. E. (2001). Problem solving by heterogenous agents. *Journal of Economic Theory*, 97(1), 123–163.
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46), 16385–16389.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: The MIT Press.
- Intemann, K. (2010). 25 years of feminist empiricism and standpoint theory: Where are we now? *Hypatia*, 25(4), 778–796.
- Jackson, S. L., & Griggs, R. A. (1988). Education and the selection task. *Bulletin of the Psychonomic Society*, 26(4), 327–330.
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6), 515–526.
- Kummerfeld, E., & Zollman, K. J. S. (2015). Conservatism and the scientific state of nature. *The British Journal for the Philosophy of Science*, 67(4), 1057–1076.
- Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

- Maciejovsky, B., & Budescu, D. V. (2007). Collective induction without cooperation? learning and knowledge transfer in cooperative groups and competitive auctions. *Journal of Personality and Social Psychology*, 92(5), 854–870.
- May, J. (2021). Bias in science: Natural and social. *Synthese*, 199(1–2), 3345–3366.
- Mayo-Wilson, C., Zollman, K. J. S., & Danks, D. (2011). The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, 78(4), 653–677.
- Mercier, H., & Heintz, C. (2014). Scientists' argumentative reasoning. *Topoi*, 33(2), 513–524.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–111.
- Mercier, H., & Sperber, D. (2017). *The Enigma of reason*. Cambridge, MA: Harvard University Press.
- Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking & Reasoning*, 4(3), 231–248.
- Peters, U. (2020). What is the function of confirmation bias? *Erkenntnis*. <https://doi.org/10.1007/s10670-020-00252-1>.
- Reijula, S., & Kuorikoski, J. (2021). The diversity-ability trade-off in scientific problem solving. *Philosophy of Science*, 88(5), 894–905.
- Rumelhart, D. E., McClelland, J. L., & Group, T. P. R. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: The MIT Press.
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1–2), 11–36.
- Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psychological Bulletin*, 129(4), 475–494.
- Solomon, M. (1992). Rationality and human reasoning. *Philosophy of Science*, 59(4), 439–455.
- Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning*, 19(1), 1–26.
- Steel, D., Fazelpour, S., Crewe, B., et al. (2019). Information elaboration and epistemic effects of diversity. *Synthese*, 198, 1287–1307.
- Strandberg, T., Sivén, D., Hall, L., et al. (2018). False belief and confabulation can lead to lasting changes in political attitudes. *Journal of Experimental Psychology: General*, 147(9), 1382–1399.
- Thompson, A. (2014). Does diversity trump ability? An example of the misuse of mathematics in the social sciences. *Notices of the American Mathematical Society*, 61(9), 1024–1030.
- Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, 143(5), 1958–1971.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293–315.
- van Eemeren, F. H., Garssen, B., Krabbe, E. C. W., et al. (2014). *The pragma-dialectical theory of argumentation*. Dordrecht: Springer.
- Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2), 225–252.
- Wilson, T. D. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.
- Zollman, K. J. S. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, 72(1), 17–35.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.