Forthcoming in *The Routledge Handbook on Trust and Philosophy*, edited by Judith Simon. New York: Routledge.

Trust and Distributed Epistemic Labour

Boaz Miller
Department of Community Information Systems
Zefat Academic College
boaz.m@zefat.ac.il

Ori Freiman
The Graduate Program in Science, Technology and Society
Bar-Ilan University
ori.freiman@biu.ac.il

1. Introduction: Trust - the Glue that Binds the Products of Distributed Research into Knowledge

In a seminal paper "Epistemic dependence", Hardwig (1985) observes that he has acquired many of his true beliefs that are commonly regarded as knowledge from testimonial sources, such as experts and the media. These include beliefs about the causes of lung cancer or economic inflation. He claims that he does not personally possess evidence for the truth of these claims, or at most only weak evidence. This is because he, qua an individual, lacks and is realistically unable to gain the necessary expertise to evaluate the evidence supporting all these beliefs. Even if he had the required expertise, it would take him more than his lifetime to acquire the relevant evidence to justify all his beliefs. This evidence is dispersed among many members of his epistemic community. Rather than basing such beliefs on evidence, so Hardwig argues, he must form them on trust in the testimonies of other people who he believes possess the required evidence.

Hardwig stresses that the asymmetry he identifies does not only characterize expert-lay relationships, but also expert-expert relationships. Due to the increased specialization in modern science, even in their own field, experts do not always have the knowledge and

abilities for evaluating their peers' claims, and even when they do, they do not have enough time and resources to evaluate them in practice (see also Rolin, this volume).

Consequently, Hardwig poses a dilemma to the traditional analysis of knowledge: if we want to maintain that such propositions are known, then either only an epistemic community and none of its members individually knows them, because the community is the only body that possesses the evidence required to justify them (Hardwig 1985), or individuals know these propositions vicariously, i.e., without possessing justification for them (Hardwig 1991).

If we follow Hardwig's lead, trust is as fundamental to knowledge as epistemic justification, e.g., evidence. Trust is the glue that binds researchers' testimonies about products of their distributed epistemic labour into collective knowledge. Trust, in lieu or in addition to epistemic justification, is the element that grants the status of knowledge to individuals' true beliefs about the products of epistemic labour of other members of their epistemic community.

Hardwig's argument, or something like it, has been accepted by many social epistemologists who study distributed scientific research, and they have explored its complications and ramifications, which will be reviewed in this paper. Other social epistemologists, however, especially those who do not focus on scientific knowledge, have resisted Hardwig's argument. Their grounds for resisting Hardwig's argument are based on the claim that an individual subject typically possesses sufficient justification for her testimonially obtained beliefs to reach the status of knowledge. They acknowledge that an individual may lack *direct* evidence for these beliefs, but they insist that she normally possesses *indirect* evidence for these beliefs in the form of multiple testimonial confirmations of the same report, evidence about her informants' sincerity and competence, and evidence about experts' success rate in making true predictions (Adler 1994; Fricker 2002: 374; Goldman 2001: 106-7). Therefore, an individual need not normally base her

beliefs on blind or partly blind trust, and Hardwig's dilemma is avoided.

Miller (2015) defends Hardwig's claims against these objections. By drawing on examples from the practice of science, Miller argues that there are sufficiently prevalent cases in which the evidence that an individual scientist possesses is insufficient to grant her knowledge according to standard theories of knowledge. Nevertheless, such scientists often offer testimony based on their insufficient evidence, and it is trusted by their peers and the lay public. For example, scientists often make discovery claims although the evidence they have for their claims are defective. Such cases, so Miller argues, are indistinguishable from the point of view of the recipients of the testimony from cases in which the testifying scientist possesses sufficient evidence for her claims. If we want to explain, why a scientist does not acquire knowledge when he trusts a colleague whose evidence is defective, but does acquire knowledge when the colleague possesses the required evidence, even if in both cases the scientist personally possesses the exact same evidence about his colleague's trustworthiness, then we must acknowledge that the evidence that justifies the scientist's belief is possessed by another person, rather than located solely in his own mind. Hence, Hardwig's challenge stands.

If we accept Hardwig's argument, or something similar to it, many questions arise. These questions are both conceptual and empirical, and they pertain to the nature of the trust that binds individuals' testimonies into collective knowledge. This paper thematically surveys these questions. In the next section we explore what grounds trust. We first sketch what trust is, and ask what factors are involved in granting it. We then ask how trust in distributed research is established and maintained. Possible answers ranging from ethnographic field studies, game-theory considerations and problems, and a moral (rather than epistemic) approach to trust are discussed.

2. What Grounds Trust? How Trust is Built, Given, and Maintained?

In the last section, we saw that researchers must extend trust to each other beyond the evidence they can personally possess for the truth of each other's claims. But researchers neither indiscriminately trust each other nor should they. Rather, they distinguish between different people and different circumstances, and use a variety of strategies for deciding whom to trust on what issues and to what extent. This section reviews research on trust in social epistemology as well as science and technology studies that addresses the question of the grounds on which researchers trust each other. This question has both empirical and philosophical dimensions, and the answers to it depend inter alia on the disciplinary background of those who study it.

Before we review the different answers to this question, we should first note that trust is not an all-or-nothing stance. That is, it is not the case that a person either categorically trusts or distrusts another person or institution. Rather, trust may be preferential, selective, and come in degrees: a person may trust another person on some matter to some extent, but distrust her on another issue (Miller 2014: 71-3). In addition, trust may be either rational or irrational. Trust is rational when the trustor has evidence about the trustworthiness of the trustee on a given mater, or at least when the trustor does not have evidence that the trustee is untrustworthy; trust is irrational when she does not. But even when rational, the evidence a person has for trusting another person does not conclusively rule out the possibility that the trustee will violate the trustor's trust (Simpson 2011: 30).

What factors determine, then, whom, to what extent, and on which issues researchers involved in distributed research trust? Let us start with ethnographic field studies of distributed scientific research, which answer this question empirically. Such studies reveal several possible answers. Based on a field research of an interdisciplinary collaborative research team, Wagenknecht (2015) identifies two strategies that scientists use to assess their collaborators' testimonies and credibility over time. While Wagenknecht calls them

"strategies alternative to trust" (2015: 164), these practices are better construed as strategies *for* giving trust.[i] One such strategy is assessing a collaborator's explanatory responsiveness. Wagenknecht depicts various practices of cross-examination in which researchers evaluate their collaborators' responses to clarification questions about their work, which are in turn used to establish their trustworthiness. The second strategy is appealing to formal indicators, such as affiliation with prestigious research institutions or publication in high-ranking journals, which are regarded as gate-keepers for credibility. In particular, a researcher affiliated with a prestigious institution may be accorded more trust for a *prima facie* suspect claim than a researcher affiliated with a less prestigious institution.

Origgi (2018: Ch. 4) notes an inherent problem with such trustworthiness indicators, both for people who use them to evaluate others, as well as for people who try to accumulate them to build their own reputation. This problem stems from two uncertainties: an uncertainty about the value of these indicators, namely, how much they are actually correlated with reliability, and a second-order uncertainty about what value other people give them. Origgi (2018: Ch. 9) further notes specific problems with their values in science, which stem from systematic distortions, such as administrators' preference for quantified indicators such as journals' impact factors, which do not necessarily best indicate a researcher's real trustworthiness (see also Origgi, this volume).

The problem of whom to trust is aggravated in interdisciplinary collaborations, when experts from different fields may work under different, possibly incompatible background assumptions and professional standards. By analyzing cooperative activities, Andersen & Wagenknecht (2013) note several ways in which researchers overcome this barrier: (1) one person (a leader) is typically responsible for the integration of research from different fields; (2) researchers gradually learn each other's' background assumptions; (3) researchers negotiate background assumptions with their collaborators from other fields.

Drawing also on an empirical field study of the laboratory, Collins (2001) argues that

trust between collaborating experimenters is grounded in their appreciation of their peers' tacit knowledge. According to Collins, when in doubt, only when researchers are persuaded through personal observation and interaction that their peers have the tacit knowledge required to successfully carry out the experiments they describe in their published papers, do they trust their experimental results.

Another approach to answer the question of the grounds for trust in distributed research does not appeal to empirical research, but to game-theoretical considerations (see also Voss & Tutic as well as Dimock, this volume). Blais (1987) models a researcher's decision to testify truly to another researcher as an iterated prisoner-dilemma game, where testifying truly is analogous to cooperating. He argues that a simple "tit for tat" tactic of rewarding reliable researchers and punishing unreliable ones is a stable strategy that may explain how scientific collaboration is possible.[ii] There are, however, problems with this approach. Frost-Arnold (2013) argues that scientific collaboration does not always fit this model. Some requests for information sharing occur only once, hence do not correspond to an iterated game; some researchers cannot know when their trust is violated; others, especially junior researchers and graduate students, are not in a position to retaliate to trust violations. But they nevertheless do extend trust. Hardwig (1991) argues that in reality, institutional sanctions for fraud or trust violation are often absent or ineffective, and therefore cannot account for researchers' trustworthiness.

As an alternative to this approach, Frost-Arnold (2013) argues that moral trust, rather than self-interested (i.e., game-theoretical) trust or epistemic trust, enables scientific collaboration. This is because scientists not only trust others when they receive others' work, but also when share their work with others. When they share their work, they trust their colleagues not to plagiarize it or steal their credit for it. When they do so, they trust their peers' moral character, rather than merely the truth of their claims. She argues that such trust is grounded in their evaluation of their peers' moral character.

Another account of the grounds for trust is given by Rolin (2004), who claims that trust in another person's testimony is prone to the influence of social biases, especially sexist biases, which may cause women's claims and research outputs to be unjustly distrusted. She stresses that when such biases operate systematically, the body of knowledge produced and accepted by an entire research community may become accordingly biased as well (see also Medina; Potter; and Scheman, this volume). Miller (2014) expands on Rolin's argument and claims that the influence of social values on trust in testimony is not just another aspect of the familiar claim that values fill the logical gap of underdetermination between theory and evidence (Longino 2002: Ch. 5), which some philosophers find unconvincing (Laudan & Leplin 1991; Norton 2008). Miller argues that values affect trust in testimony not only by filling the gap between theory and evidence, but by adjusting the weight a person gives to other people's testimonies qua evidence. Miller argues that the same mechanism operates not only with respect to testimony, but with respect to evidence in general, including evidence produced by non-humans. This may indirectly support the claim that non-humans are genuine objects of trust (see Section 5 below).

3. Inductive Risk and Calibration of Reliability Standards

When scientists report information to each other, there is always a possibility of error. Merely reporting an error does not necessarily amount to a breach of trust, since researchers take into consideration the possibility that the reports they receive are inaccurate. But how are error-rate and reliability standards established to begin with, and what constitutes breaching them? This question arises with respect to communication between researchers in the same discipline, cross disciplinary communication, and communication of scientists with the public. This section reviews answers given to these questions.

Gerken (2015) argues that there are reliability standards that are shared by all sciences. Within collaborative research, he distinguishes a testimony that constitutes input to mono-disciplinary research from a testimony that constitutes input to an interdisciplinary

research. Gerken argues that a general scientific norm of assertion is that it is appropriate to offer testimony only if it is possible to discursively justify it in the context in which it is given. He further argues that the scientific discursive standards of justification associated with this norm are closely connected with scientific virtues such as replicability, revisability, and accountability.

Gerken's account, however, is at most partial, because scientists make reports under different implicit or explicit weighings of inductive risks against each other. There is still the question of how these weighings arise, and how scientists recognize them, particularly as these weighings may differ between individual researchers and between fields. Inductive risk is the risk that stems from making a wrong epistemic judgment, such as rejecting a true hypothesis or accepting a false hypothesis (Douglas 2000; 2009: Ch. 5). Wilholt (2009) argues that shared epistemic norms and standards by researchers in a field are *social conventions* that are defined in terms of an accepted weighing of different inductive risks, where the weights are determined by competing social values, such as environmental protection versus economic innovation. As such, Wilholt *defines* trustworthiness as adherence to these conventions, and violating trust is as infringing them, either by fraudulent or sloppy research.

Miller (2014) argues that the epistemic mechanism by which values affect epistemic standards according to Wilholt's model is assigning different evidential weights to different types of evidence, for example, assigning more evidential weight to epidemiological studies than toxicology studies when determining a substance's harmfulness. While Wilholt (2009: 98) argues that epistemic norms are merely conventional, and that any convention can establish justified trust between researchers as long as it is shared and followed by all of them, Miller (2014: 75) argues that only conventions that do not allow for too high or too low rates of errors can effectively be adopted in a well-functioning research community.

Wilholt (2013) further analyzes the weighing of inductive risks that underpins these conventions, and argues that it consists of a complex trade-off between three salient values:

the reliability of positive results, the reliability of negative results, and the investigation's rate of delivering definitive results (its power). Reliably trusting another researcher involves correctly assessing the trade-offs she makes between these values, hence affording trust in distributed research is inherently value-laden.[iii]

With respect to the role inductive risks ought to play in public trust in science, a possible view is that inductive risk considerations should affect the level of certainty scientists need to meet in order to give testimony to the public. Franco (2017) argues that inductive risks should affect public scientific assertion. Suppose, for example, that scientists deem the potential negative health consequences to babies that are not breastfed as more severe than the inconvenience and career impediment from which some mothers suffer due to prolonged breastfeeding. In such a case, scientists should lower the level of certainty required to report to the public the health benefits of breastfeeding. In particular when the risks are severe, e.g., the catastrophes that may be caused by the public not accepting the theory of global climate change, scientists should express their public claims with more certainty. Steel (2016) stresses that the values that should govern such scientific testimony are not the scientists' individual values, but publically accepted values, democratically decided upon. By contrast, John (2015) argues that inductive risks should *not* affect the testimony of scientists to audiences outside the scientific community. Rather, scientists should transparently explain the relevant inductive risks and the dependency of the scientific results on them. He sees the communication of inductive risks as part of scientists' obligations to the public and to public-policy makers.

4. Trust Transforms Individuals into a Collective

So far we discussed trust between individual members of a distributed research community, that is, trust between individual researchers who rely on others' testimonies about the methods and products of their research. We may also ask what the status is of the community itself. Can distributed epistemic labor turn the individuals who pursue it together into a

group that can be accorded trust as a collective? In other words, can the fact that people divide research tasks between them and rely on each other's results turn them into a collective that may be trusted over and above the trust that is given to its individual members?

Bird answers this question in the positive. He writes, "individuals can compose a social unity when they cohere because of the mutual interdependence that arises from the division of labour" (2015: 54). Bird argues that such groups can be trusted as collectives because they have what Durkheim (1893/1984) calls "organic solidarity." While in societies characterized by mechanical solidarity, social cohesion is achieved because individual members have similar thoughts and attitudes, such as a shared religion, in organic solidarity, individual members do *not* necessarily share the same beliefs and attitudes. Organic solidarity is characterized by division of epistemic labor and mutual dependence of members on each other. In organic solidarity, so Bird argues, a scientist deeply epistemically depends on the products of other scientists' epistemic labor, such as background theories, experimental equipment, statistical tools, and computer software, to the extent that her epistemic products can be seen as collectively produced by the scientific community as a whole. Bird draws on Ed Hutchins's famous examples of the USS Palau (1995a), and the airplane (1995b), in which the social organization and individuals' various distinct roles on the ship or airplane closely match the functions needed to perform a knowledge generating process in organic solidarity.

Wray (2007) also adopts Durkheim's distinction between organic and mechanical solidarity and argues that the division of labor characterizing research teams enables the group, as a whole, to know things that individual members cannot know. This is since individual members of research teams contribute their own pieces of puzzle, the outcome of their epistemic labor, to the big picture. Wray shares Bird view that having organic solidarity is necessary for a society to constitute a collective object of trust, but unlike Bird, Wray thinks

that the scientific community as a whole does *not* have organic solidarity, because individuals' epistemic dependence on each other is not that great. Only smaller research teams in which there is high epistemic dependence between members have such solidarity. Hence only smaller teams, rather than a scientific community as a whole, can be a collective object of trust, or so Wary argues.

Knorr-Cetina (1999: Ch. 3) also relates the social organization of distributed research with the question of whether groups can be trusted as collectives. She analyzes the High Energy Physics (HEP) experiments at CERN. Scientists who work at CERN have distinct roles, with their own special areas of expertise. Each individual epistemically contributes to the general outcome. In addition to correspondence between the organizational structure of the experimental team and the elements of a knowledge generating process, Knorr-Cetina argues that the epistemic culture of the HEP experiments erodes individuals' sense of self and enhances their sense of collectivity. The epistemic culture makes "the experiment," with "experiment" referring to the people, equipment and process, the object of trust rather than the individual experimenters. Knorr-Cetina contrasts the HEP experiments with laboratory experiments in molecular biology, which do *not* turn the individuals who participate in them into a knowing collective because of their small scale and hierarchical social structure.

Against the views reviewed so far, Giere (2007) argues that distributed cognitive labor does *not* turn those who pursue it into a collective that can be trusted on its own. According to Giere, Knorr-Cetina's argument, and by extension others like it, rest on a mistaken assumption:

> Knorr Cetina seems to be assuming that, if knowledge is being produced, there must be an epistemic subject, the thing that knows what comes to be known. Moreover, knowing requires a subject with a mind, where minds are typically conscious. Being unable to find a traditional individual or even collective epistemic subject within the organization of experiments in HEP, she feels herself forced to find another epistemic subject, settling eventually on the experiment itself (Giere 2007: 316).

But, so Giere argues, as a collective entity, the experiment lacks features that make up genuine

epistemic agents, such as self-consciousness, consciousness of others, and intentionality. Hence it does not constitute a genuine agent. Distributed cognition does not entail distributed knowing.

So far in this section, it was assumed that for an epistemic community or a research group to be trusted as a collective, it must possess agency, which minimally includes having collective representations of reality, collective goals, and the ability to rationally collectively act to achieve these goals (List & Pettit 2011). Both Giere and his opponents make this assumption. They only differ on whether such an agent exists, hence their different answers to the question of trust in a collective. Wilholt (2016) denies this assumption, but argues that a community that divides epistemic labor among its members can be a collective object of trust regardless of whether it has agency. He argues that research groups indeed usually do not form agents because they exist only for the publication, the experiment, or the study, spread spatially, and involve many institutions. However, to answer whether, and how, the entire community can be an object of trust, he advocates a position according to which individual scientists trust the *conventions of a community*. These conventions, which affect the methodological standards of collaborative research groups, are not reducible to the beliefs of any individual; hence the entire community is an object of trust. In other words, the trustworthiness of research groups cannot be reduced to the trustworthiness of the group's members.

When a group is regarded as a collective entity that produces results together, a problem arises with respect to who is responsible and accountable for the claims made by the group. Traditionally, lead authors are the locus of responsibility for a paper. Wagenknecht (2015: 178) recognizes that under hierarchical authorship of multi-author papers, the lead author has various roles such as responsibility for the composition of the paper, providing the main argument and evidential basis, making adjustments, additions, and criticism on the various parts added to the paper, and its submission. In a way, the lead author serves as the

central knot in the trust relations of the authorship group. The lead/first author is also, most of the times, the most recognized person by which other researchers trust the results of a collaborative research paper.

However, as highly distributed research becomes more and more common, it becomes increasingly unclear that lead authors have this role, as other members make vital contributions that may be outside the lead author's expertise. The scientific literature is experiencing a continuous increase in co-authored papers in almost all scientific disciplines, especially in the natural sciences (for statistics, see references within Wagenknecht 2016: 21, and within Wray 2006: 507).

Moreover, due to the globalization and commercialization of research, especially pharmaceutical research, it become increasingly impossible to identify a lead author or a person who can assume epistemic responsibility for a trial's outcomes (Kukla 2012). How, then, can trust be established? What is the collective character of an author and who is accountable for scientific papers written by a group?

In current scientific practice, these questions are still open, and are highly relevant for both scientists and social epistemologists who study them (Kukla 2012; Hubener, Kukla & Weinsberg 2017). Clement (2014) suggests a method for addressing who deserves to be an author of a scientific article, by mentioning who are behind the ideas, the work, the writing, and the stewardship. Wray (2006) argues that authors of collaborative papers be conceptualized as plural subjects and not as groups of individuals. The realistic prospects of these proposals remain to be seen in particular given the different disciplinary norms in deciding about authorship orders.

The problem of holding authors accountable in highly collaborative research is aggravated by the phenomena of ghost writing – when lead authors are a rubber stamp for other people's work (Sismondo 2004, 2007, 2009; Sismondo & Doucet 2010); by sponsorships influencing published results (see meta-analysis by Sismondo 2008); and

problems associated with publication practices such as gift-authorship and conflicts of interest (Smith & Williams-Jones 2012).

5. Technological Instruments as Objects of Trust

When distributed research is conducted, cognitive tasks are delegated to non-humans, such as instruments and computers, which manufacture, represent, and disseminate data. Wagenknecht (2014: 477-478) notes that researchers need not only extend trust to other researchers' testimonies, but also to epistemic artifacts. For example, they must trust a dataset to be accurate and complete, an instrument to be properly calibrated, etc. But are such non-humans entities objects of genuine trust, or are they merely relied upon? Trust is distinguishable from mere reliance in that reliance is merely an expectation of a certain regularity to happen, whereas trust is a normative attitude. When trust, as opposed to mere reliance, is breached, the trustor feels let down and betrayed (Baier 1986).

The premise of this paper is that trust binds the products of distributed epistemic labor into collective knowledge. We saw that trust can also transform an ensemble of individuals into a collective body worthy of trust on its own. The significance of the question of whether instruments are objects of genuine trust is therefore twofold. First, can unmediated instrumental outputs constitute part of this collective knowledge, or do they first need to be testified by a human researcher? Second, do instruments constitute part of a collective body worthy of trust, or do only humans take part in it?

Within analytic epistemology and philosophy of technology, Freiman (2014) identifies two extreme sides in the debate about whether it is categorically possible to genuinely trust technological artifacts, the orthodox camp and non-orthodox camp. The orthodox camp represents the common and traditional views of mainstream epistemology, and holds that genuine trust is based on a human quality, such as intentionality, consciousness, free will, or even good will (for a survey of reasons, see Nickel et al. 2010: 432), and therefore cannot be directed at artifacts. Unlike mere reliance, trust is "inherently

subject to the risk that the other will abuse the power of discretion" (Hardin 1993: 507; for more about the distinction between trust and mere reliance, see references within Hawley 2014: 1 as well as Goldberg, this volume). According to the orthodox view, I rely on my computer to work, but I do not trust it. When it malfunctions, I might be disappointed, but not betrayed.

The current accepted view of trust can hold that "trusting is not an attitude that we can adopt toward machinery" (Jones 1996: 14). As such, inquiries about trust and technologies commonly bypass the technologies as objects of trust and point to the humans who are behind the technologies – such as engineers, designers, etc., as the objects of trust. For example, when a person trusts a bridge not to collapse while crossing it, she actually trusts the people who designed and built the bridge and those who are responsible for its maintenance (Origgi 2008; Pitt 2010). Freiman & Miller (2019) offer a middle ground position, according to which instruments may be subject of "quasi-trust", which is distinguishable both from mere reliance, and from full-fledged normative trust.

Nickel (2013) argues that trust regarding technologies means not only trusting the people who are behind the technologies, but also social institutions, as entities. According to his entitlement account of trust in technological artifacts and socio-technological systems, an trustor has evidences that indicates to her the trustworthiness of the humans behind the technologies and their interests in serving the interests of the users. In such a way, for example, a technology failure to perform will lead to an effective sanction by institutional structures; and others are willing to stake their reputations on the technologies' performances.

The question of whether instruments that are involved in distributed research are genuine objects of trust is relevant also to the issue of trust in groups. Assuming that collectives can be genuine objects of trust, do they consist of only humans or also non-humans? As we saw, Knorr-Cetina (1999) argues that in the case of the HEP experiments, the

experiment as a whole, both the humans who perform it and the instruments they use, is an collective entity. Against this, Giere (2007) argues that this claim is metaphysically extravagant, and raises principled ontological problems about setting the boundaries of the alleged collective agent.

Assuming, however, that humans and non-humans can constitute an extended cognitive system, under which conditions can this happen, and what role does trust play in them? In a famous paper, Clark and Chalmers (1998: 17) describe four conditions, known as the "trust and glue" conditions, for a person and external artifacts to be part of one extended cognitive systems. One of these criteria is that the person automatically trusts the outputs of the artifact, e.g., measurement results. It is unclear whether Clark and Chalmers use "trust" to denote genuine trust, or understand it only as a form of reliance, but however understood, relations of trust must exist between a human and an artifact for them to be part of an extended cognitive system.

When Clark explains how he wishes to apply the trust criterion, he acknowledges that if a person becomes suspicious of his artifact, it "would at that point cease to unproblematically count as a proper part of his individual cognitive economy" (2011: 104). This is in striking contrast to our treatment of dubious brain-bound recollection, whose status as mental does not waver. Externally stored information is part of a subject's cognitive system and may thus realize a subject's mental states only insofar as it is implicitly trusted, and ceases to be so when it is not, while internal information sources, such as memory or perception, remain part of the subject's cognition regardless. Clark argues that this distinction between external and internal information sources makes no difference in practice, because occasions of doubt are rare and responsibilities for checking are rare. By contrast, Record and Miller (2018) argue that responsible instrument users should sometimes doubt the epistemic output of external artifacts on which they rely, rather than trust them implicitly. They illustrate their argument with the use of a GPS app on a

smartphone, and argue that a responsible driver should not blindly follow the GPS instructions, lest she endangers herself. It therefore seems that the existence of trust relations between a user and an instrument is too strong a condition for them to constitute together an extended cognitive system. The issue of the conditions under which humans and technological artifacts constitute an extended cognitive system, and what role trust plays in them, remains underexplored in the philosophical literature.

6. Conclusion

To sum up, the introductory section of this paper presented Hardwig's argument from trust in the testimonies of others, which entails that trust is the glue that binds researchers' distributed labour products into communal knowledge. In order for trust, whatever it is, to work well, it must have certain properties. Additionally, different kinds of tasks require different kinds of glue. The social epistemology of trust studies these properties as well as the ways in which they bind individuals, knowledge, and communities, together.

The second section explored what grounds trust. We reviewed research on trust from the fields of science and technology studies and social epistemology. Answers included appeals to assessing collaborators' explanatory responsiveness and appealing to formal indicators such as affiliation and credibility. Other answers ground trust in the appreciation of peers' tacit knowledge. We then considered a self-interest account of trust, which is embedded in game-theoretical considerations. The section ended with exploring the role of the moral character of peers, social biases, and social values, in the formation of trust relations.

The third section turned to questions concerning establishing communal reliability standards, which are required for the formation of justified trust, and no less important – breaching them. These issues were dealt in respect to researchers communicating within the same discipline, across to other disciplines, and outside science – to the public and decision makers. Each communication route involves different epistemic considerations and its own

underpinning inductive risks.

In the fourth section, we raised the question of whether distributed epistemic labour can turn individuals into a group that can be accorded trust as a collective entity; that is, whether a group can be trusted over and above the trust that is given to its individual members. The question surveyed several relevant debates within group ontology about correctly characterising the division of labour and the collective objects of trust.

The last section raised the issue of technological artifacts in distributed research and collective knowledge. It began by raising the fundamental question of whether technological artifacts can categorically be considered as trustworthy. The common view within epistemology regards apparent trust in technologies as ultimately appealing to the humans who are behind the technologies, rather than the technological artifacts as objects of genuine trust. We then turned to present some existing accounts of trust that do take technological artifacts into consideration. These accounts consider technologies as well as social institutions as entities. We showed that the answer to the question of whether technological artifacts are genuine objects of trust is relevant to the issue of trust in groups. Last, we turned to the role trust plays in extended cognitive systems, where its existence it is considered a criterion for constituting such a system.We ended by noting that the roles trust plays in constituting a cognitive system of humans and technological artifacts are still largely underexplored.

References

Adler, J. E. (1994) "Testimony, trust, knowing," *The Journal of Philosophy* 91(5): 264-275.

Andersen, H., and Wagenknecht, S. (2013) "Epistemic dependence in interdisciplinary groups," *Synthese* 190(11): 1881-1898.

Baier A. (1986) "Trust and antitrust," *Ethics* 96(2): 231-260.

Beatty, J. (2006) "Masking disagreement among experts," *Episteme* 3(1): 52-67.

Bird, A. (2015) "When is there a group that knows? Distributed cognition, scientific knowledge, and the social epistemic subject," in J. Lackey (ed.), *Essays in Collective Epistemology*, Oxford: Oxford University Press.

Blais, M. (1987) "Epistemic Tit for Tat," *The Journal of Philosophy* 84: 363-375.

Clark, A. and Chalmers D. (1998) "The extended mind," *Analysis* 58(1): 7-19.

Clark, A. (2011) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford: Oxford University Press.

Clement, T. P. (2014) "Authorship Matrix: A Rational Approach to Quantify Individual Contributions and Responsibilities in Multi-Author Scientific Articles," *Science and Engineering Ethics* 20(2): 345-361.

Collins, H. M. (2001) "Tacit knowledge, trust and the Q of sapphire," *Social Studies of Science* 31(1): 71-85.

Dimock, this volume

Douglas, H. (2000) "Inductive risk and values in science," *Philosophy of science* 67(4): 559-579.

Douglas, H. (2009) *Science, policy, and the value-free ideal*, Pittsburgh, PA: University of Pittsburgh Press.

Durkheim, E. (1893/1984) *The division of labor in society*. Translated by W. D. Halls. New York: The Free Press.

Franco, P. L. (2017) "Assertion, nonepistemic values, and scientific practice," *Philosophy of Science* 84(1): 160-180.

Freedman, K. L. (2015) "Testimony and epistemic risk: the dependence account," *Social Epistemology* 29(3): 251-269.

Freiman, O. (2014) "Towards the epistemology of the internet of things: techno-epistemology and ethical considerations through the prism of trust," *International Review of Information Ethics* 22(2): 6-22.

Freiman O., and Miller B. (2019) "Can artificial entities assert?," In S. Goldberg (Ed.), *The Oxford Handbook of Assertion.* Oxford: Oxford University Press.

Fricker, E. (2002) "Trusting others in the sciences: a priori or empirical warrant?," *Studies in History and Philosophy of Science* 33(2): 373-383.

Frost-Arnold, K. (2013) "Moral trust and scientific collaboration," *Studies in History and Philosophy of Science* 44: 301-310.

Gerken, M. (2015) "The epistemic norms of intra-scientific testimony," *Philosophy of the Social Sciences* 45(6): 568-595.

Giere, R. N. (2007) "Distributed cognition without distributed knowing," *Social Epistemology* 21(3): 313-320.

Goldberg, this volume

Goldman, A. I. (2001) "Experts: which ones should you trust?," *Philosophy and Phenomenological Research* 63(1): 85-110.

Hardin, R. (1993) "The street-level epistemology of trust," *Politics & Society* 21(4): 505-529.

Hardwig, J. (1985) "Epistemic dependence," *The Journal of Philosophy* 82(7): 335-349.

Hardwig, J. (1991) "The role of trust in knowledge," *The Journal of Philosophy* 88(12): 693-708.

Hawley, K. (2014) "Trust, distrust and commitment," *Noûs* 48: 1-20.

Huebner, B., Kukla, R., and Winsberg, E. (2017) "Making an author in radically collaborative research," in M. Strevens, S. Angere, E. J. Olsson, K. Zollman, & D. Bonnay, (eds.)*, Scientific Collaboration and Collective Knowledge,* Oxford: Oxford University Press.

Hutchins, E. (1995a) *Cognition in the Wild*, Cambridge MA: MIT Press.

Hutchins, E. (1995b) "How a cockpit remembers its speeds," *Cognitive Science* 19: 265-288.

John, S. (2015) "Inductive risk and the contexts of communication," *Synthese* 192(1): 79-96.

Jones, K. (1996) "Trust as an affective attitude," *Ethics* 107(1): 4-25.

Knorr Cetina, K. (1999) *Epistemic cultures: How the sciences make knowledge*, Harvard: Harvard University Press.

Kukla, R. (2012) "'Author TBD': Radical collaboration in contemporary biomedical research," *Philosophy of Science* 79(5): 845-858.

Laudan, L., and Leplin, J. (1991) "Empirical equivalence and underdetermination," *The Journal of Philosophy* 88(9): 449-472.

List C. and Pettit P. (2011) *Group agency: the possibility, design, and status of corporate agents*, Oxford: Oxford University Press.

Longino, H. (2002) *The fate of knowledge*, Princeton: Princeton University Press.

Medina; Potter; and Scheman, this volume

Miller, B. (2014) "Catching the WAVE: The Weight-Adjusting Account of Values and Evidence," *Studies in History and Philosophy of Science* 47: 69-80.

Miller, B. (2015) "Why (some) knowledge is the property of a community and possibly none of its members," *The Philosophical Quarterly* 65(260): 417-441.

Nickel, P. J. (2013) "Trust in technological systems," in M.J Vries, A.W.M. Meijers & S.O. Hansson (eds.), *Norms in technology*, Dordrecht: Springer.

Nickel, P. J., Franssen, M., & Kroes, P. (2010) "Can we make sense of the notion of trustworthy technology?," *Knowledge, Technology & Policy* 23(3-4): 429-444.

Norton, J. (2008) "Must evidence underdetermine theory?," in M. Carrier, D.A. Howard, & J. Kourany (Eds.), *The challenge of the social and the pressure of practice: Science and values revisited*, Pittsburgh: University of Pittsburgh Press.

Origgi, G. (2008) *Qu'est-ce que la confiance?* Vrin.

Origgi, G. (2018) *Reputation: What It Is and What It Matters,* Princeton: Princeton University Press.

Origgi, this volume

Pitt, J. C. (2010) "It's not about technology," *Knowledge, Technology and Policy* 23 (3-4): 445-454.

Record, I. and Miller B. (2018) "Taking iPhone seriously: epistemic technologies and the extended mind," in D. Pritchard, A. Clark, J. Kallestrup, O.S. Palermos & J.A. Carter (eds.), *Extended Epistemology*, Oxford: Oxford University Press.

Rolin, K. (2004) "Why gender is a relevant factor in the social epistemology of scientific inquiry," *Philosophy of Science* 71(5): 880-891.

Rolin, this volume

Simpson, T. W. (2011) "e-Trust and Reputation," *Ethics and Information Technology* 13(1): 29-38.

Sismondo, S. (2004) "Pharmaceutical maneuvers," *Social Studies of Science* 34(2): 149-59.

Sismondo, S. (2007) "Ghost management: how much of the medical literature is shaped behind the scenes by the pharmaceutical industry?," *PLoS medicine* 4(9): e286.

Sismondo, S. (2008) "Pharmaceutical company funding and its consequences: a qualitative systematic review," *Contemporary Clinical Trials* 29(2): 109-113.

Sismondo, S. (2009) "Ghosts in the machine: publication planning in the medical sciences," *Social Studies of Science* 39(2): 171-198.

Sismondo, S., and Doucet, M. (2010) "Publication ethics and the ghost management of medical publication," *Bioethics* 24(6): 273-283.

Smith, E., and Williams-Jones, B. (2012) "Authorship and responsibility in health sciences research: a review of procedures for fairly allocating authorship in multi-author studies," *Science and Engineering Ethics* 18(2): 199-212.

Steel, D. (2016) "Climate change and second-order uncertainty: defending a generalized, normative, and structural argument from inductive risk," *Perspectives on Science* 24(6): 696-721.

Wagenknecht, S. (2014) "Opaque and translucent epistemic dependence in collaborative scientific practice," *Episteme* 11(4): 475-492.

Wagenknecht, S. (2015) "Facing the incompleteness of epistemic trust: managing dependence in scientific practice," *Social Epistemology* 29(2): 160-184.

Wagenknecht, S. (2016) *A Social epistemology of research groups: collaboration in scientific practice*, London: Palgrave Macmillan.

Wilholt, T. (2009) "Bias and values in scientific research," *Studies in History and Philosophy of Science* 40(1): 92-101.

Wilholt, T. (2013) "Epistemic trust in science," *The British Journal for the Philosophy of Science* 64(2): 233-253.

Wilholt, T. (2016) "Collaborative research, scientific communities, and the social diffusion of trustworthiness," in M. S. Brady & M. Fricker (eds.), *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*, Oxford: Oxford University Press.

Wray, K. B. (2006) "Scientific authorship in the age of collaborative research," *Studies in History and Philosophy of Science* 37(3): 505-514.

Wray, K. B. (2007) "Who has scientific knowledge?," *Social Epistemology* 21(3): 337-347.

[i] For Wagenknecht, trust is an indiscriminating stance toward a particular person's credibility, which does not differentiate between different testimonies this person may give, whereas we construe trust as an attitude that differentiates both between people and testimonies, possibility from the same person.

[ii] Game theorists tend to see trust as mere reliance in a game. But there is still the possibility of being betrayed by an irrational actor, which a game-theoretical model does not account for.

[iii] Freedman (2015) argues that inductive risks set norms of testimony in general, not just in science. She argues that interests are involved in the assessment of the normative status of beliefs because evidence, on its own, does not set the bar of how much certainty is needed in any given case. The amount of evidence needed depends on the hearer and the hearer's inductive risk of believing what has been told.