Philosophical Psychology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cphp20

An overview on trust and trustworthiness: individual and institutional dimensions

Elisabetta Lalumera

To cite this article: Elisabetta Lalumera (2024) An overview on trust and trustworthiness: individual and institutional dimensions, Philosophical Psychology, 37:1, 1-17, DOI: 10.1080/09515089.2024.2301860

To link to this article: https://doi.org/10.1080/09515089.2024.2301860

Published online: 15 Jan 2024.

PHILOSOPHICAL PSYCHOLOGY
2024, VOL. 37, NO. 1, 1–17
https://doi.org/10.1080/09515089.2024.2301860

Routledge
Taylor & Francis Group

EDITORIAL

An overview on trust and trustworthiness: individual and institutional dimensions

Philosophical Psychology is dedicating this issue on trust and trustworthiness to Katherine Hawley (1971–2021) for two reasons. First, she was an expert in the area. Hawley was one of the most relevant voices in the philosophical debate on trust, distrust, trustworthiness, and untrustworthiness. Second, she was a role model for professional philosophers who engage with world problems and collaborate with scientists (Brown, 2023). She authored two books for nonspecialist readers (Hawley, 2012, 2019), worked for various non-academic institutions, and collaborated on interdisciplinary projects, offering her expertise on how to be trustworthy (which is the title of one of her books).

This dedication comes with a disclaimer. This journal is interdisciplinary, publishing philosophical work informed by cognitive science and psychology in general, and we are aware that the psychological dimensions of trustworthiness do not fall within the scope of Hawley's work. Our goal was not to collect discussion pieces on Hawley's work, but to dedicate a collection of new and diverse papers on the dimensions of trustworthiness to a philosopher who has made a major contribution to the subject. Nevertheless, as I will briefly illustrate in this introduction, most of the papers collected here do deal with claims and ideas from Hawley's work. And the contributions to this special issue are truly diverse, ranging from classical epistemology and conceptual analysis to clinical mental health psychology, to science-based philosophy of economics and cognitive science. Together, they address both the personal and institutional dimensions of trustworthiness.

1. Trust and trustworthiness

The concepts of trust and trustworthiness are extremely useful in explaining personal connections between people, professional relationships (for example, the one between a carer and a health professional), as well as large-scale social phenomena such as the public's perception of science and new technologies. These concepts are undoubtedly complex and, according to some, partially evaluative or "thick" in the metaethical sense – arguably, to trust and to be trustworthy are not simply descriptions but also imply positive judgments of a certain sort of attitude and attribute, respectively.

As with any complex and likely thick concept, there is a danger that trust and trustworthiness (as well as distrust and untrustworthiness) will be defined, operationalized and used differently in different research areas or even depending on the individual study (Gerken, 2021; Rousseau et al., 1998). This is one reason why it is always beneficial to examine these conceptual tools under the scrutiny of philosophical discussion in order to hone them for specific assignments (Hawley, 2019, p. 1).

After several years of relative silence, work on trust has recently gained prominence in mainstream philosophy, starting from the seminal paper by Annette Baier (1986), both with regard to inter-personal trust, and to trust in institutions (Domenicucci, 2018), science (Oreskes, 2019), and new technologies (Taddeo, 2017). This is show by the recent publication of collected volumes dedicated to these themes (Baghramian, 2020; Faulkner & Simpson, 2017; Simon, 2020), as well as a large number of articles on trust that focus on conceptual questions (how it is defined), metaphysical questions (what kind of thing it is, such as a belief, or a disposition, or an affect, and whether it is primarily a three-place relation between two subjects and a task), epistemic aspects (when trust is reasonable or justified), and normative aspects (when it is well-founded and whether it has an intrinsic value) (See (McLeod, 2023) for an overview).

An easy way to get started with the philosophical discussion is by focusing on what sets trust apart from ordinary reliance. I rely on my alarm clock to wake me up, but that is not trust; I rely on my partner's routine of showing up for appointments a quarter-hour early, but I don't trust him to do so (if it doesn't happen, no problem on my part). Reliance is based on competence and predictability, but trust is not just about competence and predictability, and there are different philosophical accounts about the missing ingredient that trust only possesses. Proposals for the missing ingredient include trustee's characteristics, such as goodwill (Baier, 1986; Cogley, 2012), moral obligations (Nickel, 2007), or self-interest to maintain the relationship with the trustor (Hardin, 2002). But also, emotions, expectations and beliefs of the trustor are natural candidates. One early suggestion in this latter camp, from Richard Holton, was that when we trust, we hold a "participant stance" toward the object of trust, which is absent when we just rely on something, as in the alarm clock example (Holton, 1994). Moreover, as many agree on, trust makes us vulnerable; betrayed trust breeds resentment (Jones, 1996). A key theme linked to the latter claim is that trust, unlike reliance, is a so-called thick concept in that it is positively evaluable and ethically charged. Some philosophers also persuasively argued that distrust is not a mere absence of trust, rather, it is a negatively connoted evaluative attitude toward someone. Indeed, we can distrust a bad prime minister, but not the morning alarm clock if it does not ring at the time we expect (Hardin, 2004; Hawley, 2014).

Parallel to the philosophy of trust, specific work on trustworthiness in the personal and collective dimensions is also growing (Carter, 2023; Hawley, 2019; Kelp & Simion, 2023). Here again, a useful question for framing the debate is: What additional elements does being trustworthy consist of, besides being reliable for a trustor and, optionally, for a particular task or domain? Trustworthiness has been explained as a moral virtue (Potter, 2002), as a disposition to fulfill one's commitments with the trustor under good will or simpliciter (Baier, 1986; Kelp & Simion, 2023), and, more weakly, to a disposition to avoid unfulfilled commitments (Hawley, 2019). As said above, another idea is that trustworthiness amounts to having specific motives, such as moral ones, or practical interests that converge with the ones of the trustor. As with trust, trustworthiness may be ethically connoted or thick; to describe someone as trustworthy is to appreciate their value, at least epistemically (if one does not embrace the theory of trustworthiness as a virtue). As a result, failing to recognize trustworthiness when the conditions exist, or rendering a person incapable of being trustworthy in a certain scenario, are epistemic injustices with ethical impact – i.e., cases in which someone is wronged as a knower (Hawley, 2017a; Medina, 2020).

2. Trustworthiness first

But why study trustworthiness and not just trust? Indeed, for many philosophers trustworthiness is defined in terms of trust, i.e., trust is explanatorily primary, as what trustworthiness is derives from what trust is (Carter, 2023). This, however, does not detract from the fact that trustworthiness matters for philosophical accounts of trust, besides having a significative explanatory power in the social sciences and in applied ethics. If we want to understand when trust is well-grounded, we must address trustworthiness, the property to which trust is oriented. As said, dealing with trustworthiness entails accepting that trust and distrust have a normative aspect that is rooted not just in the trustor's attitude but also in the trustee's characteristics. Onora O'Neill holds the strong claim that trust is not valuable per se, but only when it is directed at the trustworthy (O'Neill, 2020). I might be justified in putting my trust in an expert who later turns out to be a charlatan if I evaluated all available information and sensibly deferred some of these checks, but that trust is arguably worthless (Origgi et al., 2021). This implies, among other things, that in order to analyze cases of loss of trust or complete mistrust, one must determine whether the recipient of trust ever was or has ceased to be trustworthy or to be regarded as such.

This philosophical normative preoccupation is also a personal, practical, and even political concern. Trust and trustworthiness often misalign, but trusting the untrustworthy and not trusting the trustworthy always "has

a cost" (Jones, 2013). In my professional life, if I don't trust a capable and sincerely interested colleague to work on a joint project, I end up working twice as hard and, per a fairly widespread intuition, I'm doing her a moral wrong. In the relationship between health professionals and those under their care – or, as they used to say, between doctor and patient – the mismatch between trust and trustworthiness can result in a patient ceasing to be compliant, having doubts, seeking endless second opinions, and starting the vicious cycle of requesting new and different services known as 'Too Much Medicine" (Fritz & Holton, 2019). On the other hand, trusting the unreliable from the patient's part may result in instances of professional manipulation (Origgi et al., 2021). Symmetrically, if the healthcare provider does not acknowledge the trustworthiness of the person being treated, they lose valuable information and commit epistemic injustice, as we know from the growing literature on this subject (Carel & Kidd, 2014). It's also noteworthy that lack of trustworthiness of a person in care might be genuine and not the result of a prejudiced attribution, and in those cases it must be identified as such in order to benefit that person (Hawley, 2015).

Moving from the individual to the collective perspective, during the COVID-19 outbreak, many people lost trust institutions, healthcare authorities and medical experts. The collective irrationalities that emerged during that time – discussed in a recent issue of this journal, see (Miyazono & Iizuka, 2023) – including extreme anti-vaccination sentiment, conspiracy theories, and anti-science epistemic bubbles can be and have been partially explained by demonstrating how governments, medical authorities, and scientists ceased to be, or to appear, trustworthy (Goldenberg, 2021). Since a decade ago, the so-called "crisis of the experts", of which the epidemic was a spectacular manifestation, cannot be explained just by pointing to the populace's misinformation or irrationality, but rather by a mismatch between public trust and trustworthiness of the experts and the institutions.

A "trustworthiness-first" approach to distrust may be identified as a trend in more recent analysis of these phenomena (Bueter, 2021; Goldenberg, 2016; Lalumera, 2018; Pertwee et al., 2022). This does not mean that if there are conspiracy theories, or antiscientific attitudes, it is the fault of scientists or governments, it just means that these are complex phenomena at least two symmetrical aspects, (dis)trust and (un) trustworthiness. In medical ethics, now that paternalism is no longer popular and the value of individuals' autonomy over health professionals is largely accepted, being trustworthy may become the normative foundation of the credibility of physicians and institutions, as well as actions that sometimes limit autonomy, such as refusing a desired but unnecessary intervention to a person in care or, on a population scale, mandating vaccination (O'Neill, 2002).

3. Conceptual issues on trust and related notions

In **"Trust's Meno problem: Can the doxastic view account for the value of trust?"** (Patrizio, 2023), Ross Patrizio confronts us with a subtle difficulty: reliance needs something more in order to be trust, but adding components to reliance does not seem to account for the value that trust has. Let us suppose that my trust in my colleague Elena to complete the project we are working on is more than just reliance; let us pick among the possible extra ingredients my belief in Elena's commitment; now let us assume that I can behave in the same way with Elena both with and without that belief about her; what does trust have over and above simulated trust, i.e., only reliance? If we accept the intuition of a value gap between any component analysis of trust and trust itself, we are dealing with the Meno problem, a classical philosophical problem from Plato's dialogue of the same name. The Meno problem is about knowledge, but Patrizio claims that it also applies to the analysis of trust and, at first glance, poses a challenge for many current theories of trust (especially those involving a belief of the trustor). But it's a challenge that can be overcome. Patrizio uses an idea from Hawley here, namely that trust is not ethically neutral. According to Hawley, trusting someone (for good reasons) is a praise and a positive value attribution, and the opposite is true as well: not trusting someone is an insult and not only a possible disadvantage for them (Hawley, 2012) Broken trust breeds resentment in the moral sense (Strawson, 1974). For examplee, even if a professor treats all of her PhD students as though she trusts them equally – for example, by putting them in charge of a project – but in reality, only trusts one, the others will still suffer ethical harm. This conclusion of Patrizio's article positions trust and being trustworthy in the sphere of values – the intuition boldly stated at the beginning of this Introduction.

"Therapeutic trust" by J.Adam Carter, also deals with a conceptual issue in the philosophy of trust, with a methodology typical of epistemological discussion. Therapeutic trust, as termed by philosophers, refers to a specific species of trust undertaken with the intended aim of fostering or enhancing trustworthiness. Consider a scenario – in Carter's introduction – where you're leaving town for the weekend and require someone to take care of your house, pets, and plants. You can entrust this responsibility to a reliable friend, someone known for their track record of responsibility. Alternatively, you might decide to entrust the same tasks to your 16-year-old nephew, who lacks any such established track record. In the latter case, this trust is established with the specific purpose of fostering or enhancing trustworthiness. Although the term "therapeutic trust" does not refer to therapy in the clinical sense, the idea of therapeutic trust can be helpful in a clinical setting where a therapist may wish to grant trust to a person in order to help them develop trustworthiness, such as when working with

young people in a mental health context. Being trusted can be empowering (McGeer & Pettit, 2017). This is something we'll see in one of the papers in this collection and in the introduction essay that follows.

The conceptual problem with therapeutic trust is that it lacks many of the characteristics of core cases of trust, such as the trustor's positive outlook on the reliability of the person they are trusting, the potential for betrayal, and the potential to cause resentment in the case of betrayal – since the likelihood of betrayal is high and the trustor has assumed the risk knowingly. Also is trust adopted for moral or practical reasons, even when good epistemic reasons are lacking (Holton, 1994), and this poses a problem for accounts of trust that involve belief, because explaining how one can believe at will is notoriously difficult So, as Carter puts it, philosophers have either put forth accounts of trust in which therapeutic trust comes out as non-rational or non-genuine trust (Pace, 2021), or they have stretched their accounts of trust ad hoc in order to accommodate therapeutic trust (Jones, 2004).

Carter's solution builds upon the concept of "overriding therapeutic trust", described as a distinct form of trust, with its own set of norms – thereby escaping the difficulty of finding a definition of trust that is flexible enough to cover the therapeutic cases. The core purpose of overriding therapeutic trust is building trust through the act of trusting successfully. Its normative condition follow from this characterization, namely, there should be a successful trusting (the teenager nephew takes good care of your house), and a successful trust-building (the teenager nephew becomes a more trustworthy person), because of your trust. In the end, "good" therapeutic trust is a form of achievement, similar to accomplishing any goal through skill rather than by chance – in Carter's terms, borrowed from, it should be accurate, adroit, and apt). In this sense, it can be compared to knowledge compared to lucky true belief, or to an archer's skillful shot as opposed to a random shot hitting the target. This idea is the core of so-called performance epistemology (Carter, 2022; Turri, 2016).

Because trust tends to be described as a three-place relation – one trusts someone for something (the domain or scope of trust) – the domain of trust should be as much investigated as as the trustor's attitudes and the trustee's attributes by philosophers of trust. In fact, this is not the case, which adds to the interest of the third article of the special issue, **"Negotiating Domains of Trust"** by Elizabeth Stewart. Stewart clarifies what a domain of trust is and points out how types of mismatches between the trustor's and the trustee's intended domain of trust can lead to resentment and distrust. When I trust a good old friend to take care of my house and garden, my trust may not include the fairly delicate task of pruning roses, and I may be hurt or disappointed to return to find the roses clipped. In a similar manner, I may have expected pruning to be within the domain of my trust and

resent that this care activity has not been carried out, but my friend may not have intended the scope (and burden) of my trust to be so broad. Unless I am a rose grower by profession and my home is my nursery, the costs are clearly relatively low. Nevertheless, there might be a breach of trust with resentment or disappointment from either party. Steward also contends that when a layperson trusts an expert or professional, such as a dentist or a hairdresser, an imperfect match between domains of trust is unavoidable. I trust one for hair and the other for teeth, but I'm not sure which actions fall under the purview of dental and hair care, so the potential of misunderstanding and betrayal of trust is increased. The observation is significant for understanding distrust in science and medicine, or at the individual level in the specialist or the clinician: not having adequately negotiated or understood the domain of trust, even though innocent ignorance, may lead to distrust – this is the point of those who claim that have unrealistic expectations about science and "false hope" in medicine are keys to understand anti-scientific attitudes and loss of trust in scientific institutions (Eijkholt, 2020; Goldenberg, 2021; Musschenga, 2019). As Stewart eventually shows, however, the potential mismatch between domains of trust is not only an epistemic danger, but also a resource. By a closer attention to the negotiation of the domain of trust, both parties in a personal or institutional relation can assess their potential role in the trust breach. Were the boundaries of trust explicitly defined, and were deal breakers reasonable? The balance between too narrow or overly broad domains of trust should be evaluated. The presence of hidden assumptions should also be investigated. Did the trustor make them explicit and communicate them effectively to the trustee? The trustee's role in seeking to uncover these implicit expectations or gather more information before accepting the trust should also be considered. In general, though at times assigning epistemic blame may remain elusive, leading to uncertainty in the trust relation or its termination, in other instances, negotiations concerning domains can help restore trust, with a focus on defining expectations, boundaries, and flexibility for future interactions.

With Simion and Willard-Kyle's contribution, **Trust, trustworthiness, and obligation**, we go on to the conceptual analysis of trustworthiness, again using an approach from classical epistemology. The question they begin with is: do we have to have reasons to believe another person is trustworthy, or do we have an entitlement by default, barring suspicions to the contrary (or, as the jargon goes, barring defeaters?) In this way, the opposition that characterizes the testimony debate shows up here as well, with those who demand more conditions being reductionists (because they attempt to reduce justification to something other than testimony) and those who argue that we can believe what we learn through testimony by default being anti-reductionists (Leonard, 2023). Mona Simion has

previously defended the idea that trustworthiness is a disposition to uphold one's obligations (Kelp & Simion, 2023) – as we quickly saw at the outset, this is a less demanding position than those who impose other conditions, such as virtue or goodwill, and slightly more demanding than Hawley's, who believes that the disposition not to take on obligations that cannot be upheld is sufficient to be trustworthy. The two authors here combine the anti-reductionist question with their preferred analysis of trustworthiness: why should we be entitled by default to assume that others are disposed to uphold their obligations? They answer as follows: we live in a world governed by norms that not only regulate, but also enable our social lives; when we put our trust in someone for something, there are norms obligating them to do what we expect of them, ranging from unspoken norms in affective relationships to explicit norms of professional duties. This is a contractualist thesis, and according to Simion and Willard-Kyle, it serves as the foundation for trustworthiness by default. This is simple and straightforward solution, but it might be vulnerable to objections as well. For example, it may be claimed that norms, in the sense of laws of social practices, exist precisely because people can trust each other and trust that the others will mostly consistently follow the agreed-on pattern of of behavior. According to this objection, trustworthiness may explain why we can have social norms but not vice versa. Consider attempting to instill norms and rules in very young children who are unable of keeping track of the responsibilities they have across time and in relation to others. This, however, is an objection that should be addressed in another forum.

Whether trustworthiness can be defined in terms of commitment is also the focus of Mélinda Pozzi and Diana Mazzarella's article, "**Speaker trustworthiness: Shall confidence match evidence?"**. The research question of the paper concerns the relation between a speaker's commitments, the (a posteriori) accuracy of what they say, and the evidence available to the listener, in order to assess trustworthiness. Here we are in the field of cognitive science and the methodology is experimental. Commitments are operationalized by linguistic signals of commitment, such as verbs like "to know" or adverbs like "certainly", "definitely", and so on. Pozzi and Mazzarella conducted two online experiments, where participants viewed testimonies about a car accident from confident and unconfident witnesses and had to evaluate their trustworthiness. One experiment had both witnesses wrong but with strong evidence, while the other had both witnesses correct but with weak evidence. The accuracy and evidence strength were consistent, while only the confidence levels differed between the witnesses in both experiments. After analyzing the experimental results, the authors conclude that the alignment of confidence with evidence can outweigh the alignment of confidence with accuracy. Therefore, a speaker who demonstrates strong confidence-evidence alignment might still be deemed

trustworthy, even if their confidence-accuracy alignment is poor. In light of this, Pozzi and Mazzarella propose that Hawley's criteria for trustworthiness are unevenly weighted: making commitments that one can fulfill holds more significance than the actual fulfillment of those commitments. Even if someone fails to fulfill a commitment they were justified in making, they can still be regarded as trustworthy. In other words, to maintain their perceived trustworthiness, a speaker should primarily commit to what they have substantial evidence for. Committing only to the truth of a message when one is justified represents a better indicator of epistemic responsibility than solely sharing accurate information, which might be accidentally communicated by individuals less concerned about truth., The paper contributes to our understanding of reputation, confidence and overconfidence, and perceived trustworthiness, though it is not obvious that commitments in Hawley's framework can be reduced to verbal expressions of certainty, as the next article in this special issue clearly articulates (see below).

4. Trustworthiness, vulnerability, and epistemic injustice

As Annette Baier (1986) pointed out, dealing with trusts entails dealing with vulnerabilities and knowledge interactions that are related to power dynamics. The trust we give to others and ourselves and the trustworthiness that people may attribute to us depend a great deal on our social status and whether we are in situations of fragility or vulnerability due to illness or marginalization, whether structural or temporary. Three of the papers in this special issue are connected by this theme. Aidan McGlynn's **Making life more interesting: Trust, trustworthiness, and testimonial injustice**, clearly demonstrates how the philosophy of trust and trustworthiness is inextricably linked to the concept of epistemic injustice – harm done to the subject in its capacity as knower (Fricker, 2007). In brief, according to McGlynn, Hawley's remarks on epistemic injustice (Hawley, 2017b) allows us to frame it in a broader context and to highlight some shortcomings of Fricker's characterization. Let us look at the argument in more detail. According to Fricker, epistemic injustice occurs when a person is diminished in their role as knower and witness because they are denied credibility and they are objectified, i.e., treated as a mere source of information, as opposed to a knowing agent. According to Hawley, there is more; that is, in instances of epistemic injustice, what is denied is the speaker's trustworthiness rather than their reliability. The distinction, as we have seen, is that when we grant or withhold trustworthiness, we are also bestowing or withholding epistemic value as well as a compliment or an act of respect. Expanding on Hawley's criticism and on his own earlier work on the topic, McGlynn notes that in instances of epistemic injustice, the victim is

not always objectified; rather, it is exactly their status as a person (typically, a person of a specific social or ethnic group) and agent that is the source of the epistemic harm (McGlynn, 2021). Another case where epistemic injustice is neither a problem of failing to attribute credibility nor one of objectification is tokenism, that is, one gives too much credibility to a person just because of her group identity (e.g., inviting a person to a conference as a speaker just because he or she belongs to a certain minority); the person is harmed in not appropriately assessing his or her epistemic credentials (Davis, 2016).

McGlynn draws attention to another significant contribution to epistemic injustice that can be found in Hawley's work. Hawley illustrated how trustworthiness involves striking a balance between wanting to commit, having the ability to acquire or apply the skills necessary to carry them out, and being able to say no. Everyone has a different strategy for this; however, for some people and in some circumstances, the approach is more complex, and it is hard to be trustworthy when one is under disadvantageous circumstances, whether they be ongoing or one-time (Hawley, 2019). This suggests that epistemic injustice can be the result of a more complex picture in which the person can only be untrustworthy, and the injustice resides in not being able to obtain trustworthy person status.

The difficulty of being trustworthy in a typical situation of vulnerability, illness, is illustrated by Michael Larkin and Zoe Boden in **The dynamics of interpersonal trust: implications for care at times of psychological crisis**. This article bridges the gap between the philosophy and psychology of trustworthiness in this special issue, and we move from the normative conditions of epistemology to a description of the experience of losing and renegotiating trust and trustworthiness, in cases of people with acute and severe mental health conditions, such as attempted suicide. From their cases, the authors suggest that someone's untrustworthiness can arise from relational pain and episodes of perceived failure and can infect all relationships; that becoming untrustworthy is a relational process, which takes time (and can therefore also be changed or stopped); that loss of trust in oneself leads to difficulties in trusting others. Finally, they propose that in some cases not attributing trustworthiness is a protective strategy for oneself and others, regardless of the epistemic merits of this attribution. This is the case of a person who does not talk to her mother about her suffering condition because she does not believe her to be trustworthy of not worrying too much and of being able to cope with the situation; in this way she tries to protect both the mother and herself, from the difficulty of dealing with an emotional reaction and a change in the mother-daughter relationship. Here, the problem is not to establish whether the patient's mother would or would not be capable to adapt to the news, but rather to understand

the strategy that the patient is devising for herself. This non-normative attention to trustworthiness is a complement to the philosophical, normative analysis of the other papers in this special issue.

Seth Goldwasser & Alison Springle in **Trauma, Trust, & Competent Testimony** also focus on a situation of vulnerability and its relationship to trustworthiness, the situation of a trauma survivor who is called upon to testify or recount their experience. According to the two authors, there is a typical argument that uses psychological evidence to diminish the competence of trauma victims, specifically that of remembering traumatic events. The conclusion of this argument is a judgment of incompetence toward the victim, which on the one hand does not blame the victim for their own incompetence, the origin of which is indeed explained, but on the other hand is belittled in their ability to be a trustworthy witness, since competence is a condition for trustworthiness. The argument that explains the memory problems of trauma victims would thus underlie a systematic epistemic injustice toward them. Goldwasser and Springle, on the contrary, argue that the victims' typical way of remembering traumatic events is a different way from that of those who have not experienced trauma, but to downgrade this different way to incompetence is a form of epistemic intolerance (a notion explored by (Catala, 2020).

5. Collective and institutional dimensions

Those who hold that it is rational to believe the testimony of others without the need for further evidence (aka the anti-reductionist) can offer this argument: that a speaker who informs us about something takes responsibility for what they say and how they say it, and through this commitment, if we trust this person, we can believe their testimony (Moran, 2005). According to Hawley, this epistemic model of testimony does not work if we want to account for how we are justified in trusting the testimony of group, and for what the trustworthiness of groups is (Hawley, 2017a). One of the problems for Hawley is that for groups there is no equivalent of taking responsibility for what and how something is said, as groups do not choose words voluntarily in the way individuals do. A second problem for her is that ascribing beliefs to a group is no easy task. Moreover, Hawley's concept of trust is morally thick – trust is a value, and broken trust bears moral resentment. Based on this, she questioned the idea that we can be just as morally offended by a group as we can be by another person. Collectively, her three doubts suggest that there is no difference between trust and reliance when the object of supposed trust is a group.

Matthew Bennett, in **Trusting Groups**, directly confronts Hawley's doubts and argues that difference between trust and reliance for groups can be significantly made. He explains that groups too have commitments,

in the sense that they can be driven by goals or values. His argument makes use of remarks on the fault and legal liability of organizations, for example of an oil company that is responsible for an environmental disaster. Furthermore, he rejects the idea that groups cannot take explicit responsibility for what they say and how; in fact there are official public statements from companies, as well as political parties (and, I would add, even scientific communities, e.g., medical, make use of many official position papers, see (Cambrosio et al., 2009)). In democratic societies there is a practice of attribution of values to companies, governments and other institutional entities, and there is also endorsement of values by groups; the betrayal of these values or the discovery of values that are not shared can lead to distrust and moral resentment in those who initially conceded trust. When a governmental agency ask citizens whether they trust the electoral system, Bennett reminds us, it is not just asking them whether they believe that the system works efficiently, but whether or not it is corrupt (American National Election Studies, 2020). With this, Bennett's article usefully connects the epistemology of trustworthiness, which has so far been predominantly person-centered, with studies in the social sciences and philosophy of science, in which trust and trustworthiness of groups and institutions are taken for granted and have long been an explanandum.

The distinction between reliance and trust in institutions is the key point of Shaun Gallagher and Enrico Petracca's article, **Trust as the Glue of Cognitive Institutions**. The notion of cognitive institution comes from the application of the extended cognition paradigm to the philosophy of economics, in particular to the thesis that institutions are shared mental models (Denzau & North, 1994; Petracca & Gallagher, 2020). The purpose of the article is to demonstrate how cognitive institutions work better when there are trusting relationships both internally and with the public.

Katherine Furman's contribution, **Beliefs, Values and Emotions: An Interactive Approach to Distrust in Science**, is a sketch for a theory of distrust in science. The author usefully reminds us that a theory has the function of explaining by finding causally relevant elements, while a concept delimits or defines a phenomenon. More specifically, the article's goal is to map the relationships between beliefs, values, and emotions in order to further build a theory of distrust that practitioners working in these contexts may utilize. Drawing from recent work in philosophy of science, in the section on values and beliefs, Furman summarizes the ways in which scientific research is value-laden and the ways in which the information we receive about science is influenced by values – where "scientific information" includes policy recommendations and medical advice as well as any information we might learn about scientific discoveries. She also observes that values can adapt to science and shift. Regarding the relationship between feelings and beliefs, Furman shows how feelings can affect how

evidence is evaluated (as in the example of parental vaccine hesitancy), and how, generally speaking, the normal vulnerability that comes with illness alters our trust in medical research and medical professionals. Lastly, regarding the value-emotion dyad, Furman contends and provides examples of how values violations can result in unfavorable emotional reactions and how emotions can signal the crossing of a value boundary. Furman's sketch for a belief-value-emotion theory of distrust may serve as a helpful framework for more research on each combination of the three elements.

We return to epistemology with Anna Pederneschi's article, **An Analysis of Bias and Distrust in Social Hinge Epistemology**, but this time by addressing a new issue, that of distrust by a group. Here the problem is that of identifying the reasons available to the distrusting group, from their own epistemic standpoint. Pederneschi's work falls under the umbrella of hinge epistemology, which holds that the basis for the justification of our beliefs ultimately rests in a set of hinge propositions that are modifiable, albeit only partially, and serve as the scaffold of our natural and social existence – developing an idea from the later writings of Ludwig Wittgenstein (Coliva, 2012). Hinge epistemology has been characterized as a third way to two classical positions with respect to the so-called Moore's paradox, one maintaining that I am not justified in believing that there is a hand here when I raise my hand in front of my eyes, until I have a reason to respond to a skeptic who speaks to me of illusion or parallel reality; the other claiming that our perceptual beliefs are directly justified when we have them – i.e., by default I am entitled to believe that there really is a hand here, until someone bears the burden of proof to the contrary. If we admit a set of hinge propositions where justification comes to an end (like mathematical axioms), we have a defense against the skeptic. At the same time, we do not have to buy the idea of direct justification of perceptual beliefs, because only some perceptual beliefs are hinges, and many hinges are not perceptual beliefs, but rather of the type "the Earth existed long before me", "there are other human beings besides me" and so on.

Hinge epistemology has been applied to another problem of classical epistemology, the justification of testimony (Coliva, 2019). Does it also apply to social trust and distrust? In the article Pederneschi employs a notion of hinge trust developed within this paradigm, which is a default trust toward people that is simultaneously certain from the subject's point of view and maximally vulnerable to be defeated. Bias can erode the hinge trust; these are by definition bad reasons, but they might be entrenched in the epistemic practices of a community, so that members use them as if they were hinges. For instance, Ben and Brenda's friends may have a deep systemic bias against the credibility of women and in favor of that of men, so they may mistrust Brenda the architect when she explains the new buildings in Milan's Isola district,

but trust Ben the amateur who has read two Instagram posts about trendy architects and comments on Brenda's explanations. Is the group of friends justified to distrust Brenda? After all, the friends of Ben and Brenda are sensible in relying on one of the most central convictions of their epistemic lives, and therefore they might turn out rational in distrusting Brenda. However, this seems to be an unattractive verdict. Coliva's strategy, which Pederneschi adopts in the paper, is that there is a distinction between hinges *de jure* (genuinely knowledge-conducive hinges) and hinge *de facto* (those contingently in effect in individual communities, which can be biased). On this view, the group's bias works like a hinge, but it is not a de jure hinge (no bias is safely truth-conducive), and therefore the group's distrust is not rational. Arguably, it is not clear how this distinction may be traced in principle without violating the tenets of hinge epistemology, which is in any case internalist – that is, all the justificatory material, so to say, should be accessible to the subject whose beliefs we want to evaluate -; so the interesting problem of the rationality of biased distrust arguably remains open. In a different epistemological framework, C.Thy Nguyen recently characterized trust as an open and unquestioning attitude (Nguyen, 2022) and defined a hostile epistemic environment as one that continuously poses a threat of betrayal to unquestioning trust (Nguyen, 2023). We might conceive of this environment of epistemic hostility as including systematic biases and consider Coliva's and Pederneschi's idea of open trust as akin to Nguyen's – mutatis mutandis. In such a mixed framework, distrust due to bias would be rationalized by the extended hostile epistemic environment. Social accounts of rationality, where the environment plays a central role, are gaining momentum, possibly indicating a further direction in which the philosophy of trust and trustworthiness can develop (Contessa, 2023; Levy, 2021).

References

American National Election Studies. (2020). *The ANES guide to public opinion and electoral behavior*. https://electionstudies.org/data-tools/anes-guide/

Baghramian, M. (Ed.). (2020). *From trust to trustworthiness*. Routledge. https://doi.org/10.4324/9780429060724

Baier, A. (1986). Trust and Antitrust. *Ethics*, *96*(2), 231–260. https://doi.org/10.1086/292745

Brown, J. (2023). Introduction. *The Philosophical Quarterly*, *73*(3), pqad054. https://doi.org/10.1093/pq/pqad054

Bueter, A. (2021). Public epistemic trustworthiness and the integration of patients in psychiatric classification. *Synthese*, *198*(19), 4711–4729. https://doi.org/10.1007/s11229-018-01913-z

Cambrosio, A., Keating, P., Schlich, T., & Weisz, G. (2009). Biomedical conventions and regulatory objectivity: A few introductory remarks. *Social Studies of Science*, *39*(5), 651–664. https://doi.org/10.1177/0306312709334640

Carel, H., & Kidd, I. J. (2014). Epistemic injustice in healthcare: A philosophial analysis. *Medicine, Health Care, and Philosophy*, *17*(4), 529–540. https://doi.org/10.1007/s11019-014-9560-2

Carter, J. A. (2022). Trust as performance. *Philosophical Issues*, *32*(1), 120–147. https://doi.org/10.1111/phis.12214

Carter, J. A. (2023). Trust and trustworthiness. *Philosophy and Phenomenological Research*, *107*(2), 377–394. https://doi.org/10.1111/phpr.12918

Catala, A. (2020). Metaepistemic injustice and intellectual disability: A pluralist account of epistemic agency. *Ethical Theory and Moral Practice*, *23*(5), 755–776. https://doi.org/10.1007/s10677-020-10120-0

Cogley, Z. (2012). Trust and the trickster problem. *Analytic Philosophy*, *53*(1), 30–47. https://doi.org/10.1111/j.2153-960X.2012.00546.x

Coliva, A. (Ed.). (2012). Moore's proof, liberals, and conservatives – is there a (Wittgensteinian) third way? In *Mind, meaning, and knowledge* (1st ed., pp. 323–351). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199278053.003.0013

Coliva, A. (2019). Testimonial hinges. *Philosophical Issues*, *29*(1), 53–68. https://doi.org/10.1111/phis.12140

Contessa, G. (2023). It takes a village to trust science: Towards a (thoroughly) social approach to public trust in science. *Erkenntnis*, *88*(7), 2941–2966. https://doi.org/10.1007/s10670-021-00485-8

Davis, E. (2016). Typecasts, tokens, and spokespersons: A case for credibility excess as testimonial injustice. *Hypatia*, *31*(3), 485–501. https://doi.org/10.1111/hypa.12251

Denzau, A. T., & North, D. C. (1994). Shared mental models: Ideologies and institutions. *Kyklos*, *47*(1), 3–31. https://doi.org/10.1111/j.1467-6435.1994.tb02246.x

Domenicucci, J. (2018). Trusting institutions. *Rivista Di Estetica*, *68*(68), Article 68. https://doi.org/10.4000/estetica.3485

Eijkholt, M. (2020). Medicine's collision with false hope: The false hope harms (FHH) argument. *Bioethics*, *34*(7), 703–711. https://doi.org/10.1111/bioe.12731

Faulkner, P., & Simpson, T. (2017). *The philosophy of trust*. Oxford University Press.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Clarendon Press.

Fritz, Z., & Holton, R. (2019). Too much medicine: Not enough trust? *Journal of Medical Ethics*, *45*(1), 31–35. https://doi.org/10.1136/medethics-2018-104866

Gerken, M. (2021). Trust issues. *Metascience*, *30*(3), 391–393. https://doi.org/10.1007/s11016-021-00659-8

Goldenberg, M. J. (2016). Public misunderstanding of science? Reframing the problem of vaccine hesitancy. *Perspectives on Science*, *24*(5), 552–581. https://doi.org/10.1162/POSC_a_00223

Goldenberg, M. J. (2021). *Vaccine hesitancy: Public trust, expertise, and the war on science*. University of Pittsburgh Press.

Hardin, R. (2002). *Trust and trustworthiness*. Russell Sage Foundation.

Hardin, R. (2004). *Distrust*. Russell Sage Foundation.

Hawley, K. (2012). *Trust: A very short introduction*. OUP Oxford.

Hawley, K. (2014). Trust, distrust and commitment. *Noûs*, *48*(1), 1–20. https://doi.org/10.1111/nous.12000

Hawley, K. (2015). Trust and distrust between patient and doctor. *Journal of Evaluation in Clinical Practice*, *21*(5), 798–801. https://doi.org/10.1111/jep.12374

Hawley, K. (2017a). Trust, distrust, and epistemic injustice. In I. J. Kidd, J. Medina, & G. Pohlhaus (Eds.), *The Routledge handbook of epistemic injustice* (1st ed., pp. 69–78). Routledge.

Hawley, K. (2019). *How to be trustworthy*. Oxford University Press.

Hawley, K. J. (2017b). Trustworthy groups and organisations. In P. Faulkner & T. Simpson (Eds.), *The philosophy of trust* (pp. 230–249). Oxford University Press.

Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy*, *72*(1), 63–76. https://doi.org/10.1080/00048409412345881

Jones, K. (1996). Trust as an affective attitude. *Ethics*, *107*(1), 4–25. https://doi.org/10.1086/233694

Jones, K. (2004). Trust and terror. In P. DesAutels & M. U. Walker (Eds.), *Moral psychology: Feminist ethics and social theory* (pp. 3–18). Rowman & Littlefield.

Jones, K. (2013). Distrusting the trustworthy. In *Reading Onora O'Neill* (pp. 186–198). Routledge.

Kelp, C., & Simion, M. (2023). What is trustworthiness? *Noûs*, *57*(3), 667–683. https://doi.org/10.1111/nous.12448

Lalumera, E. (2018). Trust in health care and vaccine hesitancy. *Rivista Di Estetica*, *68*(68), 105–122, Article 68. https://doi.org/10.4000/estetica.3553

Leonard, N. (2023). Epistemological problems of testimony. In E. N., Zalta, U., & Nodelman (Eds.), *Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/spr2023/entries/testimony-episprob/

Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford University Press.

McGeer, V., & Pettit, P. (2017). The empowering theory of trust. In P. Faulkner & T. Simpson (Eds.), *The philosophy of trust* (pp. 14–34). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198732549.003.0002

McGlynn, A. (2021). Epistemic objectification as the primary harm of testimonial injustice. *Episteme*, *18*(2), 160–176. https://doi.org/10.1017/epi.2019.9

McLeod, C. (2023, Fall). Trust. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entriesrust/

Medina, J. (2020). Trust and epistemic injustice. In *The Routledge handbook of trust and philosophy* (pp. 52–63). Taylor and Francis Inc. http://www.scopus.com/inward/record.url?scp=85106162837&partnerID=8YFLogxK

Miyazono, K., & Iizuka, R. (2023). Special issue on COVID-19 collective irrationalities: An overview. *Philosophical Psychology*, *36*(5), 895–905. https://doi.org/10.1080/09515089.2023.2221929

Moran, R. (2005). Getting told and being believed. *Philosophers' Imprint*, *5*(5), 1–29. https://doi.org/10.1093/acprof:oso/9780199276011.003.0013

Musschenga, B. (2019). Is there a problem with false hope? *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, *44*(4), 423–441. https://doi.org/10.1093/jmp/jhz010

Nguyen, C. T. (2022). Trust as an unquestioning attitude. In T. S. Gendler, J. Hawthorne, & J. Chung (Eds.), *Oxford studies in epistemology volume 7* (1st ed., pp. 214–244). Oxford University Press. https://doi.org/10.1093/oso/9780192868978.003.0007

Nguyen, C. T. (2023). Hostile Epistemology. *Social Philosophy Today*, *39*, 9–32. https://doi.org/10.5840/socphiltoday2023391

Nickel, P. J. (2007). Trust and obligation-ascription. *Ethical Theory and Moral Practice*, *10*(3), 309–319. https://doi.org/10.1007/s10677-007-9069-3

O'Neill, O. (2002). *Autonomy and trust in bioethics*. Cambridge University Press.

O'Neill, O. (2020). Questioning trust. In J., Simon (Ed.), *The Routledge handbook of trust and philosophy* (pp. 17–27). Routledge.
Oreskes, N. (2019). Why trust science? In *Why trust science?* Princeton University Press. https://doi.org/10.1515/9780691189932
Origgi, G., Branch-Smith, T., & Morisseau, T. (2021). Why trust Raoult? How social indicators inform the reputations of experts. *Social Epistemology: A Journal of Knowledge, Culture and Policy*. https://doi.org/10.1080/02691728.2022.2042421
Pace, M. (2021). Trusting in order to inspire trustworthiness. *Synthese*, *198*(12), 11897–11923. https://doi.org/10.1007/s11229-020-02840-8
Patrizio, R. F. (2023). Trust's Meno problem: Can the doxastic view account for the value of trust? *Philosophical Psychology*, 1–20. https://doi.org/10.1080/09515089.2023.2206837
Pertwee, E., Simas, C., & Larson, H. J. (2022). An epidemic of uncertainty: Rumors, conspiracy theories and vaccine hesitancy. *Nature Medicine*, *28*(3), 456–459. https://doi.org/10.1038/s41591-022-01728-z
Petracca, E., & Gallagher, S. (2020). Economic cognitive institutions. *Journal of Institutional Economics*, *16*(6), 747–765. https://doi.org/10.1017/S1744137420000144
Potter, N. N. (2002). *How can I be trusted?: A virtue theory of trustworthiness*. Rowman & Littlefield.
Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617
Simon, J. (2020). *The Routledge handbook of trust and philosophy*. Routledge.
Strawson, P. F. (1974). *Freedom and resentment and other essays*. Routledge.
Taddeo, M. (2017). Trusting digital technologies correctly. *Minds and Machines*, *27*(4), 565–568. https://doi.org/10.1007/s11023-017-9450-5
Turri, J. (2016). Knowledge as achievement, more or less. In M. Á. F. Vargas (Ed.), *Performance epistemology: Foundations and applications* (pp. 124–134). Oxford University Press.

Elisabetta Lalumera
Department for Life Quality Studies, University of Bologna
elisabetta.lalumera@unibo.it
http://orcid.org/0000-0002-0345-0838