An Analysis of Bias and Distrust in Social Hinge Epistemology

Anna Pederneschi

Department of Philosophy, University of California, Irvine, California, USA.

apederne@uci.edu

85 Humanities Instructional Building, Irvine, CA 92697

Philosophical literature has focused on the concept of trust, but often considers distrust merely as an afterthought. Distrust however, because of its pervasive role in our everyday lives, can be quite damaging. Thus, understanding the rationality of distrust is crucial for understanding our testimonial practices. In this paper I analyse whether it is rational or irrational to distrust an informant on the basis of identity bias. My aim is to show that distrust is irrational when based on negative identity bias. First, I adopt Annalisa Coliva's account of social hinge epistemology where it is rationally required to assume certain basic presuppositions, such that people are generally reliable informants, to obtain propositional justification and participate in the testimonial practice. Secondly, I show how distrust based on negative identity bias can spread across other domains of interaction and jeopardize the testimonial practice as we fail to assume that people are generally reliable informants. Thirdly, I argue that considering bias as a *defeater* is beneficial to maintain the claim that bias-based distrust is irrational, as it prevents the acquisition of propositional justification. Finally, I show that distrust is rational when based on evidence and reason that the speaker is not reliable or sincere.

1

Introduction

The epistemological literature on trust offers us a variety of definitions of the concept. Taking trust to be a kind of interpersonal relation, the literature tends to focus on defining its characteristics as well as its conditions of possibility. Some accounts especially center on determining whether trust is a two place relationship - S trusts T - or a three place relationship - S trusts T to φ (Baier 1986; Holton 1994; Jones 1996; Faulkner 2007; Hieronymi 2008; Hawley 2014). Some theories dispute if trust and reliability are indeed the same relationship or they are instead distinct (Baier 1986; Jones 1996; Hardin 2002). Others take on the task of understanding the normative dimension of the trusting relationship (Hawley 2014, 2019).

In turn, not everyone within the epistemological debate on trust takes it to be just an interpersonal relation. Rather, many scholars prefer a broader conception of trust. For instance, Thomas Simpson (2012) employs the genealogical method to identify trust as a basic and ubiquitous concept which encompasses many conceptual analysis' definitions. Thi Nguyen (2022), instead argues that trust is an unquestioning attitude which is not limited to interpersonal relations, instead we can have trust for animals, objects etc. Similarly, in On Certainty (1969), Wittgenstein formulates several remarks about the central role of trust in our epistemic practices. He characterizes trust as a core stance for the formation, acquisition, and participation in any such practice. Yet, for Wittgenstein trust is not restricted to testimonial knowledge. Rather, it plays a broader role for our species, namely it acts as a *collective bonding agent* that allows us to have

and to participate in our epistemic practices. In this light, we can imagine how detrimental the opposite stance – distrust – can be.

Distortion of information on social media, where avoiding misinformation and choosing who to distrust on the basis of good reasons is particularly complicated, provide a good example of the issues distrust present. The proliferation of information online as well as unprecedented events of the past decades – such as terrorist attacks, global financial crisis, global social movements, pandemics, wars etc – have highlighted how tricky it is to choose who to trust, while it seems very easy to adopt a distrusting attitude even when we might not have enough evidence for it.

While the philosophical literature surrounding trust has generated fruitful conceptual analysis, the literature surrounding distrust remains scarce. Yet, as I have shown in my example, it is an interesting and pressing phenomenon to pin down especially for the domain of social epistemology. The debate on the concept of distrust in epistemology is limited to its definition in relation to trust. Most accounts consider distrust as an afterthought of trust and identify it simply as the absence of trust (Baier, 1986; Grovier, 1992; Jones, 1996). Others, such as Hawley (2014, 2019) and D'Cruz (2020), argue that distrust is the contrary of trust such that there could be cases in which neither judgement of trustworthiness nor judgments of untrustworthiness are warranted.

Especially in the case of testimonial exchange we want to be able to correctly evaluate which informants are trustworthy and which ones are untrustworthy to avoid forming false beliefs and being deceived. Moreover, we want to have the correct reasons and evidence which count toward making such judgments. Finally, flagging someone as an untrustworthy informant can have consequences for their role as an informant within the

testimonial practice. Thereby, we want to be sure that we are not misplacing distrust. In light of these considerations, it seems appropriate that social epistemology delve more into the phenomenon of distrust and investigate its sources as well as its consequences for the acquisition of knowledge.

A question about distrust in epistemology that has not been investigated yet asks: On which bases is it rational or irrational to distrust someone's testimony? The intuitive answer here would be that when distrust in someone's testimony is based on evidence and good reasons that the testifier is incompetent or insincere then distrust is rational. On the other hand if distrust in someone's testimony is based on non epistemic reasons such as negative identity biases – i.e. an inclination dictated by identity prejudice – then distrust is irrational. However this intuitive distinction might not be so trivial. The following example illustrates the issue of determining the rationality or irrationality of distrust in someone's testimony. It is a fact argued by many environmental scientists that global warming is a real issue that needs to be addressed. The general public agrees on trusting scientists and their efforts to understand and solve this issue, regardless of the scientists' gender. Many also find it intuitive that it is irrational to distrust women scientists on global warming because they are women, that is based on a negative identity bias. However, in some communities it is considered perfectly rational to distrust what a woman says. Think, for example, of certain patriarchal cultures, online communities or religions. Then, this example reframe the question: is it rational or is it irrational to distrust based on a negative identity bias?

In this paper I argue that it is irrational to distrust someone based on a negative identity bias. And, distrust is only legitimate when based on evidence and epistemic reasons. For the purposes of this paper I will call illegitimate distrust the irrational kind

of distrust based on negative identity bias. I will call legitimate distrust the rational kind based on evidence and epistemic reasons[1].

My argument rests on the methodological background of hinge epistemology. In section 1 I adopt the normative framework of rationality offered by Annalisa Coliva and her brand of social hinge epistemology as well as her notion of trust. Briefly, Coliva argues that our rationality requires the assumption of certain basic presuppositions – hinges – such as "T is generally a reliable informant" to rationally participate in the testimonial practice and thus acquire justification for our beliefs. Under this theory trust is intended in the Wittgensteinian sense sketched at the beginning of this introduction—a basic stance of dependence on ourselves and each other. In the second section I show that bias based distrust is irrational because it violates the requirement of rationality for rational participation in the testimonial practice . Moreover, this kind of distrust is based on negative identity bias, it shares some features with it which make it spread across domains of interaction and social groups, jeopardising the entire epistemic practice. I will then turn in section 3 to the definition of negative identity bias within social hinge epistemology. I adopt the definition of bias as collateral belief which *defeats* the acquisition of testimonial justification. I will argue that this definition of bias is more beneficial as it confirms that illegitimate distrust is irrational. In section 4, I will focus on legitimate distrust in testimonial practices. I argue that legitimate distrust is only allowed within a particular domain of interaction where no testimonial knowledge or

---

[1] The terminology I choose here to identify irrational distrust and rational distrust is meant to avoid confusion with the terminology in social psychology. Indeed, in a previous version of this paper I refer to the two kinds of distrust respectively as unmotivated distrust and motivated distrust. Thankfully, one anonymous reviewer made me notice that such terminology would have been confusing for readers familiar with the literature on motivated reasoning in social psychology, where this reasoning is the one based on non-epistemic reasons such as biases. Then motivated distrust would have been the irrational kind based on bias, while unmotivated distrust would have been the rational kind based on epistemic evidence. I decided to steer clear of any misunderstanding and I opted for a different option of illegitimate and legitimate distrust which reflect the normative rationality background as I use to ground my argument.

justified belief is transmitted in the first place because the testifier is not knowledgeable and/or sincere. Therefore, legitimate distrust does not spread across domains of interaction. It is rational because the hearer is not required to assume any presupposition. Crucially legitimate distrust does not jeopardize the entire testimonial practice or the exchange of information.

I conclude this paper with a response to the initial issue, namely on which bases it is rational or irrational to distrust someone's testimony. Namely, the intuition that it is rational to distrust someone's testimony on the basis of evidence and good reasons while it is irrational to distrust someone's testimony on the basis of negative identity bias is proven valid.

1. Social Hinge Epistemology

In Extended Rationality (2015) Coliva proposes a distinctive reading of OC which gives raise to her brand of hinge epistemology, according to which justification and knowledge take place within a system of assumptions. Thereby, Coliva enters the debate on perceptual justification by offering a moderate alternative to the dominant positions held by liberals, on one hand, and conservatives, on the other. Briefly, the liberal view defended by James Pryor (2000) argues that perceptual justification is obtained every time one has the relevant evidence, absent defeaters. While the conservative view requires additional information concerning, for instance, the existence of an external world, not being a BIV, etc., to be appropriately justified in order to have perceptual justification. Crispin Wright (2014) claims that this latter solution leads to circularity and thus to skepticism about perceptual justification. Wright address this problem by offering a variation of conservatism that avoids charges of circularity by granting

perceptual justification on the basis of non-evidential reasons[2]. Coliva's moderatism stands between these two poles by claiming that it is constitutive of rationality to assume such collateral information – which she identifies with hinges, properly so regarded – to obtain perceptual justification (absent defeaters). By these means, moderatism agrees both with conservatives in requiring additional information, besides a certain course of experience (absent defeaters), and with liberals in avoiding the stringent condition of justifying the relevant hinges. Similarly, Coliva's account of testimonial justification (2015; 2019; n.d.), consists in a moderate alternative to the debate between reductionists and antireductionists. Furthermore, these two views can be themselves divided in global or local reductionists and global or local anti-reductionists. The analogy that Coliva notices with the analysis of the debate on perceptual justification is that local or global reductionists and global anti-reductionists hold a position that echoes the conservative one. Indeed, they consider testimonial justification as depending on the justifiedness of a global or a local hinge, characteristic of testimonial justification, while disagreeing on whether the required justification for the local or global hinge is a priori or a posteriori. In contrast, local anti-reductionists about testimonial justification – similarly to liberals about perceptual justification – deny the need for any justification or even assumption of the relevant (local) hinges.

Coliva proposes a moderate solution for cases of testimonial justification such that for a subject to be justified in believing that *p* based on testimony it must be the case that the relevant (global or local) hinge is assumed, absent defeaters. To better

---

[2]The classical argument about perceptual justification goes like this: I) I have the experience of having two hands, II) Here are two hands, III) There is an external world. The conservatives argue that to be justified in believing II) on the basis of I) we need additional evidence and reasons for III). Wright notices that if that was the case we would fall in circularity because we would presuppose additional epistemic reasons for the conclusion, we are set out to prove. Wright's (2014) brand of conservatism claims that we are entitled to III), which gives us the epistemic right to move from I) to II).

understand the analogy, Coliva represents the relevant positions as follows:

Reductionist and global anti-reductionist / Conservatives

$J_{test}(p)$: no defeater+E+J(H) $J_{perc}(p)$: no defeater+E+J(H*)

Local anti-reductionist/Liberals

$J_{test}(p)$: no defeater+E $J_{perc}(p)$: no defeater+E

Moderates

$J_{test}(p)$: no defeater+E+Ass(H) $J_{perc}(p)$: no defeater+E+Ass(H*)

In the case of testimonial justification take E to be being told that p by a testifier T, take H to be the relevant hinge such that if it is a global hinge it could consist of "People are generally reliable informants" or, if it is local, in "T is generally a reliable informant", and take J as the a priori or a posteriori justification required for H. Conversely, for cases of perceptual justification E is a current sensory experience with content p, while H∗ is a hinge such as "There is an external world".

The advantage of the moderate version proposed by Coliva over the liberal view is similar to the one in the perceptual case. In fact, without assuming the relevant hinge we would need reasons that make it more probable that *p* rather than the skeptical counterpart p′ such as "I am being deceived into believing that *p*". As well as in the perceptual skeptical scenario, the same experience would raise both the probability of the good case and of its skeptical counterpart, likewise in cases of testimonial knowledge the same testimony would raise the probability of *p* and *p′*. By these means, Coliva highlights that the only reason why we tend to favour *p* over *p′* is that we inevitably assume H in the case of testimony – or H* for perceptual justification – which consists in favouring the good case scenario. Thereby, the moderate proposal is more advantageous because by acknowledging the inevitability of assuming hinges it

makes it a requirement for justification.

The benefit of the moderate view over the conservative ones – that is, reductionism and global anti-reductionism – is that it frees us from the problematic task of giving a justification (albeit sui generis) for hinges, especially global hinges as the one considered above. At this point Coliva specifies that by assuming hinges we are not just making arbitrary assumptions. Indeed, like in the perceptual case, the testimonial hinge assumed for testimonial justification will be constitutive of social epistemic rationality (2019, p. 62; n.d., p. 8). In other words, the testimonial hinge "T is generally a reliable informant"[3]– where "reliable" means sincere and knowledgeable with respect to the topic they are testifying to – is constitutive of the practice of forming, suspending or changing beliefs on the basis of testimony. This is the practice on which our social epistemic rationality depends. Here is Coliva's definition:

> **Social Extended Rationality**
>
> It is socially epistemically rational to believe a proposition on the basis of testimonial justification and to assume those unjustifiable presuppositions (hinges) that make the acquisition of testimonial justification possible in the first place and are therefore constitutive of them.

Therefore, for Coliva we are epistemically rationally mandated by social epistemic rationality to assume the relevant hinges. Thereby, we neither have, nor need to have a justification or a non-evidential warrant or a pragmatic one to assume H. Rather, to assume H is a "requirement of reason" itself (2019, p. 62). Furthermore, this requirement mandates to assume the truth of what it mandates while it does not corroborate it (Coliva, n.d.). This last aspect of hinges is important because it raises the question of exactly which attitude we have toward hinges. In other words, the rational

---

[3] Coliva argues that this local hinge "T is generally a reliable informant" is better than the more general one "People are generally reliable informants" as it avoids the problem of bootstrapping inferences and it does not lead to easy knowledge (2019, p. 63)

mandate does not corroborate the truth of the hinge H, yet the hinge together with E and absent defeaters, corroborates the truth of p. Coliva characterizes the attitude of assuming the truth of what we are rationally mandated to assume, as a stance of trust. In OC Wittgenstein refers several times to the role and nature of trust (OC 23, 34, 125, 133, 137, 337, 599, 600, 603-5, 671-3) and the relation this stance has toward hinges. Coliva calls this stance "hinge trust" and explains how through the Wittgensteinian lenses trust is not merely a personal attitude toward a specific subject. Rather this hinge trust concerns the particular role that hinges play in our practices. Indeed, hinge trust is not reserved for trustworthy people or even testimonial knowledge. It is to be intended in broader terms: we trust our perceptual and cognitive faculties or our memory as well as testimony from people or objects (newspapers, books, historical artifacts, etc.). It follows for Coliva that such a broad notion of trust is essential for the participation in any epistemic practice, as it is what allows us to learn a language and methods of enquiry, and – once hinges are already in place – to raise doubts and start new enquiries. By these means, hinge trust appears to be our "natural" way of approaching reality. Coliva describes this feature of trust by considering how a child would trust adults, authorities and regularities in the world (n.d., p.10) and how such trust is crucial for the acquisition of a language and various methods of enquiry. As such, trust is not grounded in reasons, for reasons for or against a proposition can be formed only once one is able to participate in epistemic practices, which can only be acquired by trusting in those who pass them on to us, thereby coming to trust their constitutive hinges. That is, the epistemic reasons for our trust in authorities and our senses can be investigated only if we are already in a trusting relation with that which allows the acquisition of the methodological tools, and the appropriate hinges, needed for such investigation. Coliva defines hinge trust as a "basic stance of openness and reliance on something or someone

to do something (for us)" (n.d., p. 10), and she explains this definition in four points. First, hinge trust is a stance because it is antecedent to beliefs intended as propositional attitude of acceptance of a proposition based on reason, for one cannot have justified propositional attitudes without hinge trust. Second, hinge trust is a stance of openness because it allows to act and acquire information by taking for granted both its source and content without questioning them. Third, hinge trust is a stance of reliance since we rely on objects, people, and institutions to provide us with a language and methods of enquiry that we need to form judgments and beliefs. Fourth and last, hinge trust is a basic psychological stance which is part of our psychological constitution. Thereby, it is essential to entertain propositional content and the respective epistemic evaluation (Coliva, n.d.).

Coliva's characterization of hinge trust also considers how this stance deals with its phenomenological description as feeling secure or certain (OC 217-222). She argues that while the feeling might be in place, it is not what makes us certain or secure of a given phenomenon or event. Rather we are certain because of the role played in our system of judgement by something we basically trust in. Namely, it is hinge trust – and not the mere feeling of certainty – which enables the collection of evidence and obtaining justification and that also provides resources for doubt and inquiry with respect to ordinary empirical propositions. Furthermore, since hinge trust is a stance, it does not constitutively involve a doxastic propositional attitude (Coliva, n.d.,). Of course, hinge trust is compatible with forming beliefs when the relevant concepts to consider the issue are available and we have supporting evidence. Coliva exemplifies this through the case of "expectations": if one trusts someone to do something for her, and one has the necessary concepts to consider the issue and considers it, then one can form the belief that that person will do what one expects and one will support this belief

with evidence. However, this belief cannot be a necessary and sufficient condition for hinge trust since it is made possible by hinge trust itself. In other words, hinge trust is prior to the reasons that one may have support any belief, including those about the trustworthiness of a person (n.d., p. 12).

Coliva highlights the duality of hinge trust: on the one hand the feeling of certainty and, on the other, the constant possibility of being let down. She claims that this fragility is at the basis of the idea that the human species is a social species. For, on the one hand, at the heart of individual and collective success is this natural reliance on others that enables us to acquire the necessary tools for the formation and assessment of beliefs, hence trust is default for humans. On the other hand – since trust is so basic and automatic – for Coliva distrust is to be considered as an "illness" (n.d., p. 13). That is, distrust is rational only when the conditions for trust have been systematically violated. However, trust is always privileged ontologically and axiologically over distrust as it is our "normal" stance toward someone or something.

In other words, as well as normality is defined in terms of the conditions heathy subjects are in, so normality in social epistemic rationality is determined by the conditions of trust. Conversely, just as illness represents the abnormal condition of our health, so distrust is the abnormal condition of our testimonial practice. What is normal and abnormal is intended here in a normative sense, namely conformity to the norms that rules our practice[4]. Yet, sometimes organisms get ill and similarly subjects distrusts

---

[4] Normality could be intended in a statistical sense. However, I think normality in a statistical sense represents a measurement or an indicator that we then go on interpreting against the background of the norms that we already follow. In particular, this idea surfaces when we think about some instances of normality that seem to be more fixed than others: seeing well as well as trusting others when learning a language. For instance, even if the majority of the population presented some kind of eye defect, our everyday practices would still be governed by the assumption that we are not sick or that we should be able to see (maybe with the help of glasses). As well, our social epistemic assumptions are governed by hinge trust, while distrust would pose a concerning threat.

each other's. When is such distrust legitimate and therefore rational? When is it illegitimate and irrational instead? These are the questions I address in the coming sections.

2. Illegitimate Distrust and its Irrationality

Philosophy has recently picked an interest in distrust, as this attitude has usually been an afterthought of the debate on trust and trustworthiness. However, analyzing distrust is important because, as pointed out by Katherine Hawley (2014), distrust is not the mere lack of trust, and it is not simply nonreliance. Indeed, sometimes it is not appropriate either to trust or distrust someone. For instance, as Hawley explains, trust involves the expectation that we have toward someone to fulfill a certain commitment. Then, when the subject in question has not made any commitment and we do not have expectations, we lack trust in the subject while we don't distrust her either (Hawley, 2019, p. 4). Accordingly, distrust seems only appropriate when the subject has a commitment, but we don't trust her to fulfill it. Likewise, distrust is not just nonreliance. Indeed, a judgment of distrust involves both nonreliance and an expectation of ill will, while this latter is not required for mere nonreliance. That is, distrust has a normative dimension that nonreliance doesn't present. For instance, one could decide not to rely on a colleague or a friend for reasons that are unrelated to her trustworthiness, i.e. I know that she is very busy, or that she is going through a rough time etc. On the other hand, distrusting someone presupposes a judgment of untrustworthiness, which consist in a condemnation of the subject's skills and of her ill intentions in fulfilling a commitment. In Hawley's view then a judgment of untrustworthiness is sometimes tainted by moral criticism, as they may be inviting trust recklessly or be insincere (2014, p. 3). However understanding distrust according solely to its moral dimension runs the risk to

wrongfully attribute blame and to extend distrust to one's whole character and multiple domains of interaction even in cases where someone simply fails to give us a piece of information. Their mistake does not seem to reasonably grant such a wholesale distrust.

For this reason, distrust is extremely relevant for the epistemological debate, especially involving the rationality of trust and distrust, and its relationship with testimonial knowledge. The epistemological discussion primarily raises the normative question "Ought I to trust or to distrust?". That is, given the way things are, is it rational for me to trust? This sort of question is warranted in situations where one cannot take trust for granted anymore i.e., contexts in which trusting, or distrusting have relevant consequences for the subject and her interlocutors. As explained in Section 1, trust here is intended in a broad and basic sense of hinge trust. Namely there is a way in which we blindly trust in order to acquire a language and a method of inquiry. In this sense trust is crucially prior and different from distrust, which can only come into play once there is reason and evidence to doubt. In the case of the testimonial practice, this happens when we need to discriminate between trustworthy and untrustworthy informant, to avoid epistemic vices as gullibility, dogmatism or prejudices. In this situation we want to be able to evaluate whom to trust or to distrust on the basis of evidence of the interlocutor's competence and sincerity. However, sometimes our trusting relationships rest on affective or prudential reasons which are led by excess credibility bias. Judgments of trustworthiness based on such bias can make us believe untrustworthy people and form false beliefs[5]. Along similar lines, distrust must be based on evidence and epistemic

---

[5] We need to trust the members of our group in order to achieve a certain goal. Cases of wholesale faith in experts show that we sometimes cannot help but trust a certain category as we don't have the tools to evaluate their operate or their opinions. The consequences of irrationally trusting someone can still be detrimental, in particular in cases of deep disagreement. Latest events of the global Covid19 pandemic could provide an interesting case study on the effects of positive biases toward medical sciences and how this bias oscillates when the deep disagreement amongst experts becomes more evident to public opinion, and it constitutes a matter of political relevance.

reasons otherwise it is misplaced. Yet, in our ordinary lives many instances of distrust do not seem to rest on a confident prediction of misbehaviour or lack of competence. Rather, distrust often raises from a mere suspicion, which could easily be based on bias. I call this illegitimate distrust and I claim that in such cases distrust is irrational, as it spread to multiple domains of interaction and can cause some systematic damage to the testimonial practice. Consider a case of distrust based on a certain skepticism in someone's competence. A friend offers to fix my car and I am suspicious of her ability as a mechanic. I then refuse her offer based on my distrust of her ability in this area of expertise. I have no evidence or reasons to think that she is not good at fixing cars, but I am still suspicious of her abilities in fixing cars as I think she is too clumsy. My distrust here is a kind of skeptical attitude toward my friend's competence. However, this doubt can easily spread across domains – the quality of her will or her integrity. As Jason D'Cruz explains (Trust and Distrust, 2020, p.40), distrust based on a certain skepticism that the other person won't fulfil a certain commitment, is not just limited to this person's competence in the domain of said commitment. Instead, distrust extends across different domains such as the ability of the agent to assess her own trustworthiness or how responsibly or recklessly the agent invites trust, which in turn might be an indicator of the agent's indifference in case trust is disappointed.

Illegitimate distrust presents some particular features that enable the shift from local distrust to systematic distrust. Firstly, illegitimate distrust tends to be self-confirming: while it extends to multiple domains of interaction, it tends to alter signals of the interlocutor's trustworthiness by systematically interpreting them as evidence of untrustworthiness. I will start to distrust my friend more often when she offers help as I will interpret her offers as evidence that she recklessly invites trust, thus judging her untrustworthy even when there is evidence that she can be trustworthy.

Consequently, this distrusting attitude will feed my judgments of untrustworthiness. This also highlights how distrust self-perpetuates and could extend the initial suspicion to how I interpret and evaluate testimony from my friend.

Secondly, as we noticed earlier, trust is a sui generis kind of stance that does not correspond to a propositional attitude or a belief. The same can be said about illegitimate distrust, at least in the sense that it is not responsive to reasons and evidence, and it is not warranted. According to this distinction, a stance of illegitimate distrust is not always expressed as a belief of untrustworthiness. The belief that I express could be one of trustworthiness towards my friend, but cognitive and affective components could make me reluctant to the idea of actually depending on her[6].

Thirdly, illegitimate distrust easily generalizes from one particularly salient case to an entire group of subjects. The association between social groups and attributes can be intended as an empirical inductive generalization, where one case becomes representative of an entire category. I could generalize the local skepticism I have toward my friend to those who have her social characteristics and develop a sexist bias such as, females are clumsy or not good at manual work.

In "In My Next Life, I'll Be White" (Thomas, 1990), Laurence Thomas describes the workings of illegitimate distrust based on racial bias and he deems them irrational. Indeed, in the report of his own experiences, we can notice all the features of

---

[6] Tamar S. Gendler (2008) explains this discrepancy by introducing the notion of belief discordant alief. Briefly, an alief is a mental state usually in tension with explicit beliefs, that presents representational, affective and behavioral content, and it is automatically activated either internally or environmentally. In this sense, and alief causes a belief-behavior mismatch. Coliva offers a different explanation of this synchronic intrapersonal disagreement (2015, 2019), by arguing for a two-genus understanding of the concept of belief, as disposition and as commitment. Someone holding a belief and a conflicting distrusting attitude, is holding two different kinds of belief - a belief as a commitment with content that p and a belief as a disposition with a content incompatible with p.

illegitimate distrust considered above. He describes himself as a "well-placed and well-groomed black man" (1990, p. 84) who, despite his social position and social appearance, still experiences the distrust of white people. This shows how illegitimate distrust based on racial bias is resistant to evidence, since the same kind of evidence is taken to be an indicator of trustworthiness in white men. He provides numerous contexts in which white women clutched their purses in his presence, or the police stopped him for inexistent reasons. We could imagine some of this people holding anti-racist belief, while acting (or reacting) with a distrusting attitude. Moreover, it is very clear form everyday examples that the individuals' distrust based on race is generalized to the entire social group regardless of social, cultural, and financial differences. Thomas notices that this "public distrust" allows a vicious circle: since illegitimate distrust is accompanied by a judgment of blameworthiness it may undermine the internal psychological motivation of the wrongly distrusted subject to be trustworthy. When trust is denied a priori the subject of distrust might as well give up trying to "re-earn it"[7]: "Thus the sear of distrust festers and becomes the fountainhead of low self-esteem and self-hate. [...] For it is rare for anyone to live morally without the right sort of moral and social affirmation." (1990, p. 84)

I think this characterization of illegitimate distrust as irrational fits perfectly within social hinge epistemology. In particular, distrust based on identity bias will not just be irrational, but it will end up jeopardizing the entire testimonial practice.

---

[7] It could seem that to re-earn trust would be in tension with the idea that hinge trust is basic, and thus being considered trustworthy is not something that should be earned. However, in this scenario hinge trust is violated by the defeater constituted by the racial bias which spreads across the domain of interaction. Therefore, while people of color should be considered trustworthy by default, they find themselves in a practice where the racial bias constitutes a quasi-permanent defeater which makes testimonial justification unavailable for those who hold it. To re-earn trust then is not in tension with hinge trust being basic, as it would be a practical move to make the defeater generated by the identity bias evident and thus to show how those who hold it are violating the mandate of social epistemic rationality.

Illegitimate distrust toward a specific individual's testimony (or an entire social category) consists in parting with the relevant testimonial hinge and thus adopting a stance of default-distrust. Indeed, illegitimate distrust targets the role of hinge trust in our epistemic practices. We said that hinge trust is the necessary stance we have toward hinges, and, in particular, towards the testimonial hinge "T is a reliable informant on this occasion". Furthermore, we said that hinges are unjustifiable assumptions that make the acquisition of testimonial propositional justification for a belief that *p* possible. It follows that the testimonial hinge is at the basis of our social epistemic practices and as such assuming it is a rational requirement. In other words, to rationally acquire testimonial knowledge – and thus participate in testimonial practices – we must assume the testimonial hinge and its truth[8]. Crucially, the way we assume the testimonial hinge is through hinge trust, which is part of the extension of rationality operated by Coliva. So, hinge trust will also be a stance of rationality, not because we have reasons to back it up, but because it enables us to assume a hinge and to abide by epistemic rationality's requirements. Along these lines, an attitude of distrust toward the testimonial hinge will have the consequence of hindering our rational participation in the testimonial practice, but more worryingly it will mean to transgress the mandate of social epistemic rationality (and the result would be a situation as the one described by Thomas, where hinge trust does not work as default for some). Just as distrusting the relevant hinge in perceptual cases of knowledge, namely deciding not to assume "There is an external world", would lead to radical skepticism and would impede the rational participation in many epistemic practices, distrust in the testimonial hinge would prevent the very

---

[8] Coliva clarifies that hinges can be considered truth-apt in a deflationary sense of truth. Hinges can be true in a minimal sense, as they can be semantically assessed as the content of proposition of acceptance or meaningful negation or of conditionals (Coliva and Palmira, 2020). The truth of hinges can be captured by the equivalence schema typical of deflationary truth (Horwich, 1998)

rational participation in the epistemic practice of acquiring knowledge and justified beliefs through testimony, which is itself constitutive of social epistemic rationality.

3. Two Definitions of Bias

In the foregoing section we have focused on how illegitimate distrust occurs in cases of bias, and it is especially perilous when its source is identity bias. Moreover, I have shown that illegitimate distrust is irrational both from our everyday social perspective and from the theoretical perspective of hinge epistemology. However, there is one more issue that needs to be addressed: it is not trivial that our testimonial practices constituted by hinge trust deems identity bias as a source of irrational illegitimate distrust. Indeed, this conclusion depends on the definition of identity bias that we adopt, which could allow for illegitimate distrust to be rational and to contradict the previous general claim of section 2. Two main interpretations of bias within the social hinge framework are proposed by Coliva (2019b) and Boncompagni (2021)[9]. In this section I want to consider both definition and evaluate which one is more beneficial for the purposes of understanding illegitimate distrust as irrational. Let us consider this example of identity bias and whether these two accounts can define bias while maintaining the irrationality of motivated distrust. Brenda is an architect, and she lives in Milan. Some of her friends from Rome are visiting her. They are taking a walk in Milan and they arrive in Porta Nuova district and they stop in front of the Vertical Forest buildings (Bosco Verticale). Brenda is showing them around and thanks to her expertise she is able to give them some insights about those particular towers. Ben – one of the friends – cuts her off. He

---

[9] Boncompagni focuses on prejudices to clarify their normative import within cases of epistemic injustice. I take prejudice and biases here to be somewhat similar, as the differences are not relevant to my argument. However, we can distinguish prejudice from bias. Prejudice is a judgment that targets the speaker in her ability as a knower based on her social identity, identity bias is an inclination or preference dictated by the identity prejudice.

starts explaining to Brenda the architectonical features of the towers and he also suggests to her a couple of books to be more informed next time. Ben likes architecture, but he is a novice and all the information he has comes from a couple of books he read in his free time. This is a case of mansplaining, namely when someone, generally a man, explains to a woman something obvious or something she is an expert in, assuming he always knows more than a woman and/or that she does not really understand the subject matter; at times, the further assumption to this behavior is that when a man speaks, the woman must be quiet and listen even though she might be an expert in the subject matter. It is also the case that Brenda and Ben's friends hold the same gender bias such as "Men are more reliable than women as informants". Intuitively, when Ben mansplains to Brenda her own expertise, and when the friends support Ben, it appears that they are doing something harmful and irrational (or at the very least their rationality in this case is dubious). However, there are communities were holding such a bias and shutting down women regardless of their level of expertise is considered perfectly reasonable.

Let's consider this example of mansplaining within Boncompagni's account of identity bias. Especially we want to see if her account maintains the intuition about Ben's irrationality or rather Ben is rational within his community. Boncompagni adopts Coliva's account of hinge epistemology where hinges are assumptions constitutive of our epistemic practice. For Boncompagni these assumptions provide propositional justification for perceptual judgments or for other kinds of judgments, namely moral, aesthetic, religious or political ones (2019, p. 174). For Boncompagni some of these hinges are local because they are assumptions shared by a community at a given time of

history[10]. In this sense, hinges sometimes can be influenced by biases held within the community. Boncompagni ascribes some characteristic of hinges to these biases: they are normative with regards to the epistemic practice, they operate as rules of evidential significance, they are unresponsive to reasons and socially epistemically mandated. Therefore, for Boncompagni, there can be biased hinges, or hinge prejudices that govern (unjust) testimonial practices within a certain community. Furthermore, Boncompagni adopts the distinction proposed by (Coliva and Palmira, 2020, footnote iv) between de jure and de facto hinges. Namely, the former are hinges that cannot be challenged without forsaking the entire epistemic practice, whereas the latter are unchallenged by society but could be challenged given for instance some scientific or social revolution. Hinge prejudices are de facto hinges (even if not necessarily local) that can change along with social change. The testimonial hinge is instead a de jure hinge which cannot be renounced without giving up the whole testimonial practice.

With this definition of identity bias, Ben's hinge prejudice is normative with respect to his practice, it shapes the evidence he takes into account, it is unresponsive to reason and socially epistemically mandated within his community (remember that his friends hold the same biased hinge). Then, while the gender bias in play here gives raise to illegitimate distrust, it also ensures the rationality of illegitimate distrust, because the distrusting attitude is mandated by the hinge prejudice that governs the practice. It thus turns out that Brenda's friends are not being irrational after all. Brenda seems to have no standing to rebuke them or disregard them as irrational. More specifically, since she shares no common hinge, relative to that practice, such that her complaints would be regarded as rational from her friends. Moreover, the de facto hinge in play here allows

---

[10] Boncompagni draws on the presence of local hinges in OC such as "No one has never been on the moon" (OC 286) or "A king can make rain (OC 132).

for illegitimate distrust to thrive in the forms that we have seen in section 2, and it replace as a matter of fact the de jure hinge that constitutes the entire testimonial practice.

Crucially this hinge account differs from Coliva's in how one thinks of hinges and bias. Namely, in Coliva's constitutivist account of social hinge epistemology, hinges are just those that are constitutive of the relevant epistemic practice – that is, the de jure ones – and biases work merely as defeaters. In the first section we framed testimonial justification as available when the subject assumes the hinge that the testifier is reliable. A speaker is reliable when she is sincere and knowledgeable regarding what she is saying. Hence, the testimonial justification for the belief that p is available in the abstract space of reasons. At this point, the constitutivist account shows that if the hearer holds a negative bias toward the speaker T then the propositional justification to believe that p based on T's testimony will not be rationally available to the hearer (Coliva, 2019b, p. 64). Furthermore, the bias held by the hearer will only be the result of a doxastic attitude and thus it will constitute a collateral belief. This does not "destroy" the propositional justification for the belief that p, rather biases are defeaters preventing the hearer from availing herself of the relevant justification and thus they make this latter rationally unavailable. Now, in Brenda's situation her friends have a negative gender bias that makes them discriminate against Brenda, qua member of a given social category or group. Moreover, this gender bias is a defeater as it prevents them from participating in the social epistemic practice of acquiring testimonial justification and/or knowledge from a reliable informant such as Brenda. That is hearers would be propositionally justified to believe her testimony by assuming – i.e. hinge trust – the relevant testimonial hinge "T is a reliable informant on this occasion". However, due to their bias, Brenda's friends will miss out on this piece of testimonial knowledge as it

will not be rationally available to them. The friends' distrust toward Brenda is irrational for two reasons: it is generated by a bias that constitutes a defeater toward the acquisition of knowledge and it contravenes the mandate of social epistemic rationality as it disregards the assumption of the relevant hinge.

Considering the two possible definitions of bias within social hinge epistemology, it thus appears that understanding bias as a defeater, rather than as itself a hinge, allows us to maintain the irrationality of illegitimate distrust due to identity prejudice.

4. Legitimate Distrust

In the previous sections I have displayed the distinctive features of illegitimate distrust and how it is irrational to hold such a stance within a testimonial practice. At this stage, what is left to understand is whether there can be instances of legitimate distrust and what is the difference between these two stances.

Let us consider this example. I am worried about my health because I am noticing odd spots on my skin, and I am not sure what is causing this condition and how to cure it. A friend notices the spots on my skin, and she suggests applying a specific medicament. My friend is not a doctor or a dermatologist and thus she is not in the position to recognize the symptoms of a specific illness or to advise me on a cure. Given that in such a scenario my friend is not a knowledgeable informant on the matter, I can distrust her since she is not transmitting any knowledge or justified belief.

Considering the example, distrust is legitimate when the informant is not reliable on a certain occasion and thus no knowledge is transmitted, and no testimonial

justification is required. In cases of legitimate distrust, the informant testimony is not disregarded because of a bias-generated defeater but because of the lack of one of the conditions of the testimonial practice – the informant must be sincere and knowledgeable on the subject matter to be considered reliable. If we compare Brenda and Ben's example of illegitimate distrust with the legitimate distrust I have toward my friend, it is possible to see the difference between the two kinds of distrust. On one hand, we have Brenda who is a reliable informant on that occasion, and she is transmitting testimonial knowledge (Brenda is an architect and she is talking about architecture). However, Ben distrust her because of a gender bias which constitutes a defeater that makes testimonial justification unavailable for him. On the other hand, I have no bias towards my friend when she is giving me medical information. Rather, I have good reasons to think she is not a reliable informant on that occasion. Thus I can disregard her testimony because no knowledge or justified belief is transmitted in the first place. What makes distrust illegitimate is the bias that constitutes a defeater to the acquisition of knowledge and availability of justification. However, in cases of legitimate distrust there is no testimonial justification since the hinge "T is a reliable informant" cannot be assumed.

As opposed to illegitimate distrust, legitimate distrust is rational as it does not violate the mandate of social epistemic rationality. The hearer is allowed to not assume the testimonial hinge: when my friend who is not a doctor or a dermatologist gives me medical advice there is no proper testimonial exchange, for this reason I am allowed to not assume the testimonial hinge and I can distrust what she is saying without violating the mandate of social epistemic rationality.

Of course, one possible worry is that legitimate distrust could present a similar

risk of proliferation as illegitimate distrust. We have seen in section 2 how distrust could arise from a mere suspicion, and it could systematically spread in other domains of interaction and perhaps extend to those that are in a similar social category. However, that is the case only for illegitimate distrust generated from a bias since it is a stance that the hearer would hold regardless of the speaker's expertise or reliability. Contrarily, legitimate distrust is only allowed within a particular domain of interaction where no testimonial knowledge or justified belief is transmitted in the first place. Thus, legitimate distrust is rational because it does not violate the mandate of social epistemic rationality: the testimonial hinge is not assumed only when the testifier is not a reliable informant in a particular occasion.

A further crucial characteristic of instances of legitimate distrust is that they are not in tension with hinge trust. Legitimate distrust rests on the absence of knowledge transmission and the hearer is not required to assume a testimonial hinge as the speaker is not a reliable informant on that occasion. Thus, legitimate distrust is not violating the mandate of social epistemic rationality as well as hinge trust. This is an important distinction because as we previously noticed illegitimate distrust caused by bias – especially identity bias – risks to place a quasi-permanent defeater that makes testimonial justification unavailable whenever testimonial knowledge is transmitted by members of a certain social category. Especially, such cases of systematic bias might result in social dynamics such as the ones we saw with the racial bias case offered by L. Thomas or the gender bias depicted in Brenda and Ben's example. Furthermore, in such scenarios things get more complicated if we also consider all the emotional and psychological factors that would make those targeted by bias feel the need to continuously reaffirm their expertise within the testimonial practice. At this point the issue that remains open assumes a more practical aspect: which techniques and

strategies we can put into place to contrast irrational distrust and change people's mind. This, however, is a topic for a separate discussion. In this section we only considered the epistemic dimensions of trust and distrust.

Conclusion

The philosophical debate about distrust – not only as an afterthought of trust – is quite recent and therefore there are still few considerations about the differences between legitimate and illegitimate distrust especially in the epistemic domain. It may seem intuitive that illegitimate distrust would be considered irrational, especially from the prospective of current events: in the public opinion distrust toward certain groups based on race or gender is seen as irrational as well as distrust toward experts (take the flat-earthers supporters who distrust entire fields of science). As intuitive as it may seem from the perspective of current events, it is not trivial that illegitimate bias-based distrust is irrational.

Rather, some accounts of bias and prejudice seem to run the risk of making it rational, after all. Relying on our daily experiences of negative bias and then arguing for illegitimate distrust's irrationality within the framework of social hinge epistemology developed by Coliva, I have provided an account of the irrationality of illegitimate distrust based on identity prejudice. Key to that account is the idea that identity-bias works as a defeater of the testimonial hinge that is constitutive of social epistemic rationality and makes the hearer irrational inasmuch as they would not be able to avail themselves of the propositional testimonial justification there is, for a given proposition p, that is the content of a speaker's assertion. I then proposed to distinguish between legitimate and illegitimate distrust. On one hand, illegitimate distrust makes testimonial

justification unavailable because of the defeater generated by the hearer's bias. On the other hand, legitimate distrust is such because the speaker is not a reliable informant on that specific occasion and thus no knowledge or justified belief is transmitted. By these means, legitimate distrust is also rational because the hearer is not required to assume the testimonial hinge and by distrusting the speaker, she is not violating the mandate of social epistemic rationality as well as hinge trust.

References

Baier, A. (1986). "Trust and Antitrust". In: Ethics 96.2, pp. 231–260. issn: 00141704, 1539297X.

Baker, J. (1987). "Trust and Rationality". en. In: Pacific Philosophical Quarterly 68.1, pp. 1–13. issn: 1468-0114. doi: 10.1111/j.1468- 0114. 1987.tb00280.x.

Boncompagni, A. (2019). "Hinges, Prejudices, and Radical Doubters". In: Wittgenstein-Studien 10.1, pp. 165–181. doi: 10.1515/witt-2019-0010.

— (2021). "Prejudice in Testimonial Justification: A Hinge Account". en. In: Episteme, pp. 1–18. issn: 1742-3600, 1750-0117. doi: 10 . 1017 / epi .2021.40.

Coliva, A. (2015). Extended rationality: a hinge epistemology. eng. Palgrave innovations in philosophy. Houndmills, Basingstoke, Hampshire New York: Palgrave Macmillan. isbn: 978-1-137-50189-9 978-1-137-50188-2.

— (2019a). "Disagreeing with Myself: Doxastic Commitments and Intrapersonal Disagreement". In: American Philosophical Quarterly 56.1, pp. 1–14.

— (2019b). "Testimonial hinges". en. In: Philosophical Issues 29.1, pp. 53–68 issn: 1758-2237. doi: 10.1111/phis.12140.

— (n.d.). "Social Hinge Epistemology". in J. Lackey & A. McGlynn (eds.) Oxford Handbook of Social Epistemology (accepted)

Coliva, A & Palmira, M. (2020). "Hinge Disagreement". In: M Kusch (ed.) Social epistemology and epistemic relativism, Routledge, 2020.

D'Cruz, J. (2020). "Trust and Distrust". In: The Routledge Handbook of Trust and Philosophy. Routledge. isbn: 978-1-315-54229-4.

Faulkner, P. (2007). "On Telling and Trusting". In: Mind 116.464, pp. 875–902. issn: 0026-4423.

— (2015). "The attitude of trust is basic". In: Analysis 75.3, pp. 424–429. issn: 0003-2638. doi: 10.1093/analys/anv037.

— (2017). "The Problem of Trust". In: The Philosophy of Trust. Ed. By Paul Faulkner and Thomas Simpson. Oxford University Press, pp. 109–128. isbn: 978-0-19-873254-9. doi: 10.1093/acprof:oso/9780198732549.003. 0007.

— (2020). "Trust and Testimony". In: The Routledge Handbook of Trust and Philosophy. Routledge. isbn: 978-1-315-54229-4.

Gendler, T. S. (2008). "Alief and Belief". In: The Journal of Philosophy 105.10, pp. 634–663. issn: 0022-362X.

Govier, T. (1992), "Distrust as a Practical Problem". Journal of Social Philosophy, 23: 52 63. https://doi.org/10.1111/j.1467-9833.1992.tb00484.x

Hawley, K. (2014). "Trust, Distrust and Commitment". en. In: Noûs 48.1, pp. 1–20. issn: 1468-0068. doi: 10.1111/nous.12000.

— (2019). How to be trustworthy. First edition. Oxford ; New York, NY: Oxford University Press. isbn: 978-0-19-884390-0.

Hardin, R. (2002). Trust and Trustworthiness. Russell Sage Foundation. http://www.jstor.org/stable/10.7758/9781610442718

Hertzberg, L. (1988). "On the attitude of trust". In: Inquiry 31.3, pp. 307–322. issn: 0020-174X. doi: 10.1080/00201748808602157.

Hieronymi, P. (2008). "The reasons of trust". en. In: Australasian Journal of Philosophy 86.2, pp. 213–236. issn: 0004-8402, 1471-6828. doi: 10.1080/00048400801886496.

Holton, R. (1994). "Deciding to trust, coming to believe". en. In: Australasian Journal of Philosophy 72.1, pp. 63–76. issn: 0004-8402, 1471- 6828. doi: 10.1080/00048409412345881.

Horwich, P..(1998). "Truth". eng. In: Meaning. Oxford: Oxford University Press. isbn: 978-0-19-823824-9. doi: 10.1093/019823824X.003.0004.

Jones, K. (1996). "Trust as an Affective Attitude". In: Ethics 107.1, pp. 4–25. issn: 0014-1704.

Lahno, B. (2017). "Trust and Collective Agency". In: The Philosophy of Trust. Ed. by Paul Faulkner and Thomas Simpson. Oxford University Press, pp. 129–148. isbn: 978-0-19-873254-9. doi: 10.1093/acprof:oso/ 9780198732549.003.0008.

McGeer, V. & Pettit, P. (2017). "The Empowering Theory of Trust". In: The Philosophy of Trust. Ed. by Paul Faulkner and Thomas Simpson. Oxford University Press, pp. 14–34. isbn: 978-0-19-873254-9. doi: 10.1093/acprof:oso/9780198732549.003.0002.

McLeod, C. (2002). Self-trust and reproductive autonomy. Basic bioethics. Cambridge, Mass: MIT Press. isbn: 978-0-262-13408-8.

Moran, R. (2006). "Getting Told and Being Believed". In: The Epistemology of Testimony. Ed. by Jennifer Lackey and Ernest Sosa. Oxford University Press, pp. 272–303. isbn: 978-0-19-927601-1. doi: 10.1093/acprof: oso/9780199276011.003.0013.

Nguyen, C. T. (2020). "Echo Chambers and Epistemic Bubbles". en. In: Episteme 17.2, pp. 141–161. issn: 1742-3600, 1750-0117. doi: 10.1017/epi. 2018.32.

— (2022) "Trust as an Unquestioning Attitude", in Tamar Szabó Gendler, John Hawthorne, and Julianne Chung (eds), Oxford Studies in Epistemology Volume 7, Oxford Studies in Epistemology

Pryor, J. (2000). "The Skeptic and the Dogmatist". In: Noˆus 34.4, pp. 517– 549. issn: 0029-4624.

Potter, N. N. (2002). How Can I Be Trusted?: A Virtue Theory of Trustworthiness. Rowman & Littlefield Publishers.

— (2020). "Interpersonal Trust". In: The Routledge Handbook of Trust and Philosophy. Routledge. isbn: 978-1-315-54229-4.

Simpson, T. W. (2012). "What Is Trust?" en. In: Pacific Philosophical Quarterly 93.4, pp. 550–569. issn: 1468-0114. doi: 10.1111/j.1468-0114. 2012.01438.x.

Stern, R. (2017). "Trust is Basic". In: The Philosophy of Trust. Ed. by Paul Faulkner and Thomas Simpson. Oxford University Press, pp. 272– 294. isbn: 978-0-19-873254-9. doi: 10.1093/acprof:oso/9780198732549. 003.0016.

Thomas, L. (Aug. 1990). "Opinion — Next Life, I'll Be White". en-US. In: The New York Times. issn: 0362-4331.

Wright, C. (2014) "On Epistemic Entitlement (II): Welfare State Epistemology." In: Dodd D & Zardini E (eds.) Scepticism and Perceptual Justification. Oxford: Oxford University Press, pp. 213-247. https://global.oup.com/academic/product/scepticism-and-perceptual-justification-9780199658343?cc=gb⟨ =en&

Wright, C. (2004). "Warrant for Nothing (and Foundations for Free)?" en. In: Aristotelian Society Supplementary Volume 78.1, pp. 167–212. issn: 1467-8349. doi: 10.1111/j.0309-7013.2004.00121.x.

Wittgenstein, L. Anscombe, G. E. M. & von Wright G. H. (1969). On Certainty. Oxford: Blackwell. isbn: 978-0-631-12000-1.