

1.

## Introduction: An Overview of Trust and Some Key Epistemological Applications

Katherine Dormandy

[*Trust in Epistemology*, ed. Katherine Dormandy, Routledge 2020, 1-40.]

Trust in other people and ourselves is fundamental to accomplishing our aims. These include epistemic aims, such as the acquisition of knowledge, evidence, or understanding; the cultivation of intellectual virtue; and the overcoming of epistemic injustice. There are copious philosophical literatures on both trust and epistemology, yet scholars from each are only beginning to draw on the other. Much conversation is still needed, and this volume aims to promote and shape it.

The authors focus on the role that trust plays in epistemology. This is called for because, with some exceptions to be discussed below, it is not uncommon for epistemologists to regard trust as simply part of the philosophical wallpaper: as a precondition for testimonial knowledge (when directed toward others) and for rationality (when directed toward oneself), but of limited interest in its own right.

A précis of the contributions can be found in section 3. Before that, to facilitate the conversation between the philosophies of trust and knowledge, I give an overview of the trust literature and then of six central issues concerning epistemic trust. The survey of trust in section 1 zeroes in on the kinds of expectations that trust involves (1.1), trust's characteristic psychology (1.2), and what makes trust rational (1.3). The discussion of epistemic trust focuses on its role in testimony (2.1), the epistemic goods that we trust for (2.2), the significance of epistemic *trust* in contrast to reliance (2.3), what makes epistemic trust rational (2.4), and epistemic self-trust (2.5 and 2.6).

### 1. Theories of Trust

We can trust objects, such as shelves or software, as well as persons. My topic here is trust in persons. Accounts are too varied to systematize neatly, so I will list some salient features on the basis of which they may be grouped.

One feature that I will mention only to set aside concerns how many relata the trust relation has: three, two, or even one. On three-place trust, one person trusts another for some end, such as to be driven to the airport. On two-place trust, one person has a trusting attitude toward another even if there is no particular end for which he trusts her. And on one-place trust, he is simply a trusting person: he goes through life with a high degree of “basal security” (Jones 2004). There is ample room for research into two- and one-place trust,<sup>1</sup> but three-place will be the focus of this introduction. One reason is that epistemic trust certainly involves three-place trust, namely for certain epistemic goods (see section 2.1), and this holds regardless of whether three-place trust is basic or can be understood in terms of a lesser-place trust.

Three-place trust essentially involves the truster's *relying* on the trustee for something; this amounts roughly to working into his plans the supposition that she will do what he is relying on her for (Holton 1994, 68).

---

<sup>1</sup> See Jones (2004), Becker (1996), Faulkner (2015), Domenicucci and Holton (2017).

Beyond this common view that (three-place) trust entails reliance, there is widespread disagreement about what else it involves. I will look at three areas of debate.

### 1.1 Predictive and Normative Expectations

In a trust relationship, the truster is commonly taken to have certain expectations, either predictive or normative, toward the trustee. A *predictive* expectation is an expectation *that* someone will do something – for example, that the trustee will do what you are trusting her for. A *normative* expectation is an expectation *of* someone to do something – for example, the expectation *of* the trustee to come through. Predictive and normative expectations can come apart: one partner expect *that* the other will wash the dishes (because he usually washes them when she has worked late) without expecting *of* him that he'll do it (it may strictly speaking be her turn). And she can expect *of* him that he will wash them (it is his turn tonight), but not expect *that* he will do so (because he often forgets).

One way to group accounts of trust is by the types of expectation that they posit. There are four views. One says that predictive expectations alone are constitutive of trust; I'll call this the *mere-predictive account* (Dasgupta 1988; Hardin 1993; Coleman 1990, chapter 5). Another says that trust is constituted by predictive and normative expectations; we may call this the *combined account* (Baier 1986, 1991; Nickel 2007; McMyler 2011; Cogley 2012; Keren 2014, this volume). The third, which I'll call the *mere-normative account*, says that only normative expectations are constitutive (Holton 1994; Becker 1996; Jones 1996, 2004; Hinchman 2005; Faulkner 2007; Darwall 2017). There is conceptual space for a fourth account, on which trust does not necessarily come with any expectations at all, being characterized instead, for example, in terms the psychological profiles discussed in section 1.2, or on which there is no single account of trust (Simpson 2012). But this section limits discussion to the three accounts on which trust comes with expectations of one form or another. In considering how each type of expectation may feature in trust, we will see some plusses and minuses of each account.

Let's look first at predictive expectations. Accounts positing them differ over how strong the predictive expectations need to be; two alternative requirements have developed. What I'll call a *strong* predictive-expectations requirement says that the truster must *believe* that the trustee will come through (Hieronymi 2008; Keren 2014, this volume; McMyler 2011). What I'll call a *weak* requirement says, more generally, that he may harbor any kind of positive expectation, ranging from belief to a probability assignment just above  $\frac{1}{2}$  (Hardin 1993; Hawley 2014a).<sup>2</sup>

The strong predictive-expectations requirement is motivated by the following sort of case (Hieronymi 2008, 218): you accept your friend's invitation to dinner, but are subsequently invited to a clashing event. Unbeknownst to you, your friend learns of the conflict. As a result, he does not believe that you will make good on your dinner commitment, so when you turn up he expresses pleased surprise. Hieronymi argues that you have cause to be dismayed at his response, on the grounds that his disbelief displays a lack of trust in you.

Against the view that trust involves strong predictive expectations, however, is the observation that trust is voluntary, whereas beliefs are not (Holton 1994): a person can choose to trustingly rely on someone even if his confidence that she will come through amounts to less than full-on belief. Indeed, trust can be "therapeutic" (Horsburgh 1960, 346; Faulkner 2007, 886): you might trust someone simply for the sake of encouraging her to grow in trustworthiness, even if you do not believe that she will come through. The classic example is of a father, aiming to cultivate trustworthiness in his teenage daughter; he goes away for the

---

<sup>2</sup> Proponents of both requirements, however, tend to exclude the possibility that trust is compatible with certainty that the trustee will come through.

weekend, trusting her not to trash the house even though his confidence that she won't is too low to constitute belief (Jones 2004, 5). These observations support the claim that, if trust requires predictive expectations, the requirement is weak rather than strong.

The mere-predictive view of trust might posit either a strong or a weak requirement; what it insists on is that trust is constituted by predictive expectations alone, not normative ones. On this view, most commonly held in the social sciences, trusting someone is simply a matter of incorporating her anticipated actions into your plans. This view is argued to be insufficient to capture full-fledged trust (Baier 1986; Holton 1994; Nickel 2007; Ruokonen 2013; Kallestrup, this volume). The criticism is that making plans on the basis of a person's predicted behavior, by itself, is not trust, but mere *reliance*. To see what this distinction amounts to, consider the oft-cited example of Kant, whose habits were allegedly so regular that his neighbors could set their clocks to the time at which he left his house each day. Kant's neighbors could *rely* on him to be punctual, but they cannot be said to have *trusted* him for this. They would have no cause to feel betrayed, for example, should he have overslept one day (Baier 1986, 234). Yet trust legitimates such reactive attitudes: it is appropriate for the truster to feel grateful if the trustee comes through, and to feel betrayed or let down if she does not (see Dormandy, this volume, and Hinchman 2017 for discussions of betrayal).<sup>3</sup> Trust, unlike mere reliance, comes with norms that make reactive attitudes appropriate.

A closer look reveals that trust comes with other norms too. One is that the trustee ought to clearly signal her acceptance of trust (Jones 2017) and do her reasonable best to come through for the truster. Another is that the truster ought not micromanage the trustee's efforts (Baier 1991, 117). A third is that both parties ought to represent their intentions toward each other authentically (Frost-Arnold 2014a). None of these norms applies in the case of Kant's neighbors timing their clocks to his movements. The difference is that Kant and his neighbors are not engaged in a cooperative relationship, so their actions swing free from obligation or responsiveness to each other. In a trust relationship, by contrast, each party must take the other into account, at least with respect to accomplishing the end being trusted for (Greco, this volume, discusses the relationship between cooperation and trust).

This due consideration of the other party that characterizes trust over mere reliance can be understood in terms of *care* (Baier 1986, 1991; Becker 1996; Jones 1996; Cogley 2012; Hinchman 2017; Grasswick 2018). It is not that one party is expected to *feel caring* toward the other, but rather that the actions normatively expected in trust are actions characteristic of care. Each party grants the other a measure of discretion as to how to best care for him: the truster leaves it to the trustee to determine when and how to come through, and the trustee exercises discernment in how best to come through for the truster in his context. For example, if your neighbor trusts you to water his plants while he is away, then he leaves it to you to decide how to fit this activity into your day. As for you, you might fulfill his trust by doing something beyond just watering his plants: for example, in the surprising event that you also discover his tap running, you turn it off. His explicit trust in you is for the plants, but there is an area of discretion surrounding this task that may sometimes extend to other things (Baier 1986; Hinchman 2017). The norms of trust, in summary, preside over a cooperative relationship of care.

If trust, as opposed to mere reliance, comes with special norms, then it surely constitutively involves normative expectations, whether or not it also involves predictive ones. If this is so, the mere-predictive account is mistaken. It may capture one phenomenon that we sometimes call "trust", but it misses the normatively rich sort just described. The philosophical literature tends to reserve the term "trust" for the latter phenomenon. On this picture, the expectations of trust are characterized either by the combined account or the

---

<sup>3</sup> Holton (1994) *defines* trust in terms of a disposition to form reactive attitudes, but McLeod (2000, 474) and Hawley (2014, 8) argue that this account is insufficient.

mere-normative one; as for the mere-predictive account, this is thought to describe mere reliance. I will follow this usage here.

But which of the two normative accounts is preferable? The only difference between them concerns whether they think that trust involves only normative expectations, or predictive ones too. To adjudicate between them, let's look more closely at the role of predictive expectations in trust. Recall that there are two alternative predictive-expectations requirements: strong and weak, and that, if trust includes one of them, it is arguably the weak one, since trust seems compatible with greater doubt than full-fledged belief permits. However, it bears asking whether trust can withstand greater doubt still: perhaps it does not require any (positive) predictive expectations at all. This would suggest a mere-normative account over a combined one. Let's consider what such an account might look like.

A mere-normative account might be mild or extreme. A *mild* version says that, although trust does not require *positive* predictive expectations, is incompatible with negative ones. That is, you can count as trusting someone even if you suspend judgment about whether she will come through, but not if you think her *unlikely* to. I won't linger on this view here, because I want to explore the extreme form of the mere-normative account. This version says that trust *is* compatible with negative predictive expectations – that is, with the expectation that the trustee will let one down.

One might be tempted to reject this suggestion on conceptual grounds. The thought would be that the idea of trusting someone whom you think is more likely than not to fail you is incoherent. After all, isn't the whole point of trust to obtain what you are trusting for? It is difficult to engage with this objection without simply trading intuitions. But let me spell out a scenario in which trust – or what I will provisionally call trust – is compatible with negative predictive expectations. It is a situation in which the point is *not* to obtain what the person is trusting for, but some extrinsic end.

The example I have in mind is therapeutic trust. Recall the father trusting his teenage daughter with the house for the weekend: we may imagine that his primary aim is not to preserve his house in one piece, but to cultivate trustworthiness in his daughter. And this is compatible with thinking her more likely than not to trash the house. He presumably *hopes* that she will take good care of the house, but he may judge that even if she trashes it, the exercise will still cultivate trustworthiness in the long run (say, through her processing her emotions of feeling sorry). Indeed, the father might judge that there is no more effective way to teach his daughter trustworthiness, and that the risk to the house is therefore worth it. So the extreme version of the mere-normative account cannot so easily be rejected on the grounds that the point of trust is just is to achieve what you are trusting for.

The objector might simply deny that the father *trusts* his daughter, as opposed to simply acting as if he does (Keren 2014, 2610-2611). In cases of real trust, one might say, the purpose *is* to achieve what you are trusting for; and here you cannot count as trusting if you have a negative predictive expectation that the trustee will deliver. In response, we may envision another scenario that calls this stark conceptual claim into question: it is one in which you *can* trust – or do what I'll provisionally call trusting – even if your sole aim *is* to obtain the thing that you are trusting for. This may happen if there is unfortunately no more effective way to obtain it. If I am dangling from a cliff, I might trust my bitterest enemy to hoist me up, knowing that the only alternative is plummeting to certain death. I may think it more probable than not that she will let me fall; but I may desperately trust her in the blind hope that she will be responsive either to my plight or at least to what is normatively expected of her. And this cannot be chalked up to mere reliance, as long as there are norms in place and my attitude (albeit without much hope) presupposes them.

So the extreme version of the mere-normative view cannot be so easily dismissed (and thus nor, presumably, can the mild one). But determining whether the two cases I have described are conceptually admissible examples of trust would take us too far afield.

A final question remains: what is the source of the norms of trust relationships? One possibility is that they are moral, so that doing right by someone in a trust relationship is a way of *doing right* (Baier 1986; Holton 1994; Nickel 2007; Faulkner 2011; McMyler 2011; Frost-Arnold 2013; Hinchman 2017; Darwall 2017). The idea is not simply that the actions performed in trust relationships are answerable to morality in general – for every action is so answerable, including those not involving trust. It is rather that trust relationships come with special moral norms of their own. Consider an example: it is a general moral truth that, if I steal your cake, I have done something wrong. However, if you *trust* me not to steal it and I accept your trust, then, if I steal it anyway, I have committed an additional moral infraction above and beyond stealing: I have betrayed your trust.<sup>4</sup>

Against the moral view stands the observation that some instances of trust and trustworthiness – such as between members of a mafia – are patently *immoral*, and that some instances of betrayal – say, to prevent a crime – are good (Baier 1986; Mullin 2005; Jones 2017). Jones (2017) rejects the moral view (as does Becker 1996), arguing that trust, though constrained by moral norms, has an intrinsic, non-moral, normativity of its own. Mafia partners, on Jones’s view, can instantiate good trust even as the actions that this involves are wrong. However, the observation that trust relationships can be immoral and betrayal good does not compel us to Jones’s view.<sup>5</sup> One could hold that the norms of trust are *pro tanto*, so that the goodness of upholding them in a given situation can be outweighed by other moral norms.<sup>6</sup>

In summary, there is widespread agreement that trust, as opposed to mere reliance, comes with normative expectations, but there is disagreement over what sorts of norms are at issue and over the necessity of positive predictive expectations.

## 1.2 The Psychology of Trust

Beyond predictive or normative expectations, there are other psychological features that many think are constitutive of trust. A second division among accounts concerns these. The views tend to focus on the psychology of the truster, but we can infer from them a corresponding psychology of the trustee. More specifically, we can infer a corresponding psychology of a trustee *who accepts the trust*. It is possible to be a trustee, in the sense of being trusted by someone, yet to not accept his trust: you might be unaware that he is trusting you, or have signaled unwillingness to do what he needs. The accounts to be discussed here concern trustees who do accept the trust. Note finally that accepting trust does not entail *being trustworthy*, so discussions like those of Daukas (2006), Jones (2012b), and Simpson (2013) that explore trustworthiness cannot automatically be imported into a discussion of the psychology characteristic of a trustee.

There are roughly three types of account of the psychology of trusting and accepting trust.<sup>7</sup> On the first, the truster relies on the trustee to be motivated to come through for him, where he takes this motivation of the trustee’s to be self-interest (Coleman 1990, chapter 5; Hardin 1993). The corresponding account of accepting trust would thus say that it is self-interest that motivates the trustee to come through.

---

<sup>4</sup> Some go as far as to say that the very *concepts* of trust and trustworthiness are moral; i.e., that merely predicating them of something implies that that thing is good (McLeod 2000; Stern 2017; Ruokonen 2013). Similarly, that the concept of betrayal is bad, implying that anything of which it is predicated is bad.

<sup>5</sup> Jones herself argues indirectly, by criticizing Scanlon’s (2003) moral view of the norms of trust. See also Hinchman (2017), who argues that the norms of trust are not necessarily moral.

<sup>6</sup> Thanks to Arnon Keren. This is arguably the view of (Baier 1986).

<sup>7</sup> One view that does not fit neatly into this division is that of Pettit (1995), on which the truster ascribes both goodwill and self-interest to the trustee. Another is that of Kappel (2014), on which trust is a disposition to rely on someone.

The self-interest account is standardly paired with one of the two views on which trust involves predictive expectations,<sup>8</sup> but there is no entailment between a self-interest view and a predictive-expectations view. You can rely on someone to come through out of self-interest, without expecting that she will be competent to do so (you might, for instance, be in the unfortunate situation, described in section 1.1, of lacking better options); and you can predict that someone will come through without relying on her self-interest (you may rely for instance on her good will; see below).

However, the self-interest account *does* seem to entail that trust does *not* come with normative expectations (Ruokonen 2013; Kallestrup, this volume). Why? Because it is hard to see how a relationship defined in terms of self-interest and reliance on self-interest could *also* be answerable to the special moral norms of trust. For example, in such a relationship, it does not seem legitimate for the truster to expect *of* the trustee that she perform any actions that would not serve the latter's interests – after all, this is not part of the terms of their cooperation; so failing to come through cannot legitimately be met with feelings of betrayal. Consider the signing of a real estate contract: when seller and buyer show up and sign, gratitude is less appropriate than relief; and a buyer who pulls out at the last minute might legitimately cause annoyance, but not feelings of betrayal. When self-interest is (and is accepted as) the motivating factor in a relationship, what we have is not trust but mere reliance.

One might think that there is a version of the self-interest view that leaves room for the special moral normativity of trust. On the “encapsulated-interest” view, the truster relies on the trustee to adopt his interests as her own (Hardin 1993). A trustee can adopt a truster's interests for unselfish reasons, such as because she cares about him – and relationships of care *are* governed by special moral norms. So it would seem that the encapsulated-interest account can characterize trust, in the richly normative sense, over and above reliance. But this conclusion is too quick. On the encapsulated-interest view, if special moral norms enter into a case of trust, they do so not as part of the trust itself. Instead, they are smuggled in by whatever aspect of the relationship, extrinsic to the trust, motivates the trustee to adopt the truster's interests in the first place. If she is motivated by her care for him, then the special normativity comes from the care; but if she is motivated by more self-interest, it remains unclear, on the encapsulated-interest view, where the special moral normativity of trust might come from. As long as trust itself is characterized only in terms of self-interest, it lacks the normativity that would make it anything more than reliance. The self-interest view, then, is more appropriately called a view of mere reliance.

The second account of the psychology of trust also turns on the truster's ascribing a certain motive to the trustee. But this motive, unlike self-interest, is other-directed. Views differ over what this other-directed motive is. Some say that it is good will toward the truster or toward some shared project (Baier 1986, 1991; Jones 1996; Becker 1996),<sup>9</sup> and others say that it is the trustee's awareness that the truster is depending on her (Hinchman 2005; Faulkner 2007; Jones 2004; McMyler 2011; Ruokonen 2013). Others yet say that the trustee is motivated by moral uprightness (McLeod 2000). We may infer a corresponding account of accepting trust, on which the trustee has the requisite other-directed motive.

How does the other-directed account of trust's psychology square with accounts of its expectations? The other-directed account does not entail, and is not entailed by, the claim that trust involves *predictive* expectations. Again, you can take a trustee to be motivated to come through for you without thinking that she will (she may be incompetent), and you can think

---

<sup>8</sup> So, either the mere-predictive account or the combined account (see section 1.1).

<sup>9</sup> I take it that Baier (e.g. 1986) is often misread as saying that the goodwill ascribed to the trustee be directed toward the truster himself (e.g., Holton 1994, p. 65), which gives rise to various counterexamples. As Mullin points out, however (2005, 317), however, this is too narrow a reading.

that she will come through for you without thinking that her reason for doing so is other-directed. A parallel point holds for normative expectations. The other-directed account does not entail that trust has any normative expectations: ascribing or having good will does not entail the existence of norms that mandate or legitimate certain sorts of behavior. Nor do normative accounts of trust's expectations entail an other-directed view of its motives: it might be legitimate to expect of someone that she do something even if she does not happen to be motivated by some other-directed concern to do it. That said, some other-directed views of trust's psychology add a mere-normative account of its expectations (Becker 1996; Jones 2004; Hinchman 2005; Faulkner 2007; Darwall 2017), and others add a combined account of its expectations (Baier 1986, 1991; McMyler 2011; Cogley 2012).

Motives feature prominently in explaining people's behavior, so motive-based accounts have the advantage of explaining why the trustee would accept the burden of trust to begin with. But one problem with such accounts is that they have trouble supplying a corresponding account of *distrust* (Hawley 2014b). The key to this argument is that the theoretical tools that feature in your account of trust should also be able to account for distrust, yet the tools of the motive-based account cannot do this.

Hawley argues this by listing and rejecting some ways in which motive-based accounts might construe distrust (I'll limit this discussion to her arguments for accounts in which the motive is other-directed). Such an account might say, first, that distrusting someone is simply a matter of not trusting her – that is, of declining to ascribe her an other-directed motive (and hence declining to rely on her on account of such an ascription). But this is not sufficient for distrust (Hawley *ibid.*). For example, most of us do not rely on our colleagues to buy us champagne, the reason being that we do not ascribe them an other-directed motive to do so. Yet this does not mean that we *distrust* our colleagues vis-à-vis the provision of champagne; rather, we neither trust nor distrust them. Second, an other-directed motive account might suggest that distrust is a matter of ascribing *negative* other-directed motives to the trustee, such as ill will. But such an ascription is not necessary for distrust: a putative trustee might bear you no ill will whatsoever (she may for instance just be interested in enriching herself), yet you might still distrust her (*ibid.*, 6). Finally, an other-directed motive account might suggest that distrust is a matter of ascribing negative motives, if not toward the (putative) truster himself, then toward the prospect of his obtaining the thing that he would trust for. But this is not sufficient for distrust: someone may be motivated to thwart your aims without meriting your distrust. For example, she may, in a perfectly honorably manner, be running a campaign for your political opponent (*ibid.*). Here too, you neither trust nor distrust her vis-à-vis your aim of getting elected.

The proponent of the motive-based view might respond by denying that distrust must be accounted for by the same mechanisms used to account for trust: although an account that does so may be more theoretically elegant, there is little empirical reason to think that the world will conform to our aesthetic preferences.

This suggestion has bite if no means can be found to account for trust and distrust simultaneously. But Hawley argues that there is such an account. This brings us to the third view of the psychology of trust: the *commitment* or *obligation* account (respectively Hawley 2014b; Nickel 2007).<sup>10</sup> Here, the truster ascribes to the trustee not a specific motive, but rather a commitment or an obligation to doing what she is trusted to do. This commitment is compatible with various motives, other-directed or self-directed.<sup>11</sup> Trusting someone, then, includes both thinking that she is committed (or obligated) to coming through, and relying on her to do so; and distrusting her is thinking that she is committed (or obligated) to coming

---

<sup>10</sup> Hawley (*ibid.*) argues that commitment is a better notion for an account of trust than obligation.

<sup>11</sup> Other views posit commitments *and* a specific motive on the part of the trustee (Hinchman 2005; Faulkner 2011; McMyler 2011).

through, yet *declining* to rely on her. From this psychology of the truster we may infer that someone who accepts trust is committed to coming through.<sup>12</sup>

Note that thinking that someone is committed to doing something is different from expecting that she will do it, so the commitment account does not entail that trust comes with predictive expectations. Nor does the claim that trust involves predictive expectations entail that the truster ascribes a commitment to the trustee (he may think that she will come through for reasons independent of a commitment). What about normative accounts of the truster's expectations? The commitment view does entail that the truster has normative expectations of the trustee. The reason is that commitment only makes sense against the backdrop of norms governing the keeping of commitments. A normative account of trust's expectations, however, does not entail the commitment account, since you can expect *of* a person that she come through for you, and rely on her to do so, without thinking that she is committed to doing it.

I have discussed two groupings of accounts of what trust is: one in terms of the sort of expectations it involves, and the other focusing on its psychology; I'll now turn to what makes trust rational.

### 1.3 Rational Trust

The third grouping of accounts of trust centers on how to construe its rationality. One thing that is plausibly involved in rational trust is evidence. Yet trust and evidence are often contrasted, with trust portrayed as constitutively opposed or unresponsive to evidence (Baker 1987, 1; Hardwig 1991, 693; Jones 1996, 15; Becker 1996; Faulkner 2007, 876; Hieronymi 2008, 222; McGeer 2008, 240; Grasswick, this volume). How does the idea that trust can be rational, and that rational trust involves evidence, fit with the supposed contrast between evidence and trust? The key lies in seeing what makes actions in general rational, and how sometimes even rational actions can go beyond one's evidence. We will then see how these considerations apply to the actions of reliance involved in trust.

What makes an action rational is that it is the best means available to the agent, at least as far as he can tell on the basis of his evidence, for achieving an end that he values. We may construe evidence broadly, as including any reason for belief, including experiential reasons and intellectual seemings. Evidence helps an agent determine how the world is and thus shape his actions accordingly: it tells him which actions are the best means available, all things considered, to achieving his valued end. To see what makes trust rational, then, we must look at the different sorts of end that it can promote, and at the role that evidence place in cases where rational trust goes beyond it.

One end that trust might promote is simply that of obtaining what the truster is trusting for: a lift from the airport, emotional safety in a relationship, and so forth. In many such cases, trust is rational simply when the truster has sufficient evidence that the trustee will deliver, and that there is no better way to achieve this aim. But not in all cases: as we saw in section 1.1, trust can sometimes be rational even when the truster *lacks* sufficient evidence that the trustee will come through.<sup>13</sup> This can happen when the thing that he is trusting her for is worth a risk and there is no better way to obtain it (think of my enemy pulling me up from the cliff edge). Here, the evidence that will make his trust rational pertains primarily to whether trusting her is the best way to attain the end in question (including the costs of the risk), even

---

<sup>12</sup> Views positing a commitment on the part of the trustee might differ over what she is committed *to*. She might be committed to some person (standardly but not necessarily the truster; Hawley 2014b) or simply to the norms of trust (Mullin 2005; Hinchman 2005; Faulkner 2011).

<sup>13</sup> In cases like this and the one considered in the following paragraph, accounts on which predictive expectations are essential to trust will be committed to saying that, if the evidence causes one's predictive expectation to drop below a certain level of confidence, the action is no longer one performed in *trust*.



if it is merely the best of an assortment of bad ways.<sup>14</sup> On the specific matter of whether the trustee will come through, the trust here goes beyond the truster's evidence – yet his trust, because it fits his evidence on these other matters, is still rational.

Another sort of end that trust might promote, as we saw in section 1.1, is extrinsic to the trust itself: it is something other than the objective for which the truster is trusting. Recall our example of the father therapeutically trusting his teenage daughter not to trash the house while he is away.<sup>15</sup> Here, the evidence needed to rationalize the father's trust need not primarily concern whether his daughter will keep the house in one piece, but – more importantly – the value of teaching her trustworthiness, and whether leaving her with the house (whether or not she takes good care of it) is the best way available him to do so, even if it is the best of a bad lot.

These considerations point to the falsehood of what we may call the *narrow evidential constraint*.<sup>16</sup> This constraint says that trusting someone to do something is only rational if your total evidence on balance supports the belief that she will do it (Hardin 1993; Coleman 1990, chapter 5). “Belief” here may be understood, in the terms given in section 1.1, as a strong, as opposed to a weak, predictive expectation. The narrow evidential constraint is surely true in some cases, such as when you cannot afford a risk, or when there may be better options for achieving your end than trusting (Simpson 2017). But as a general constraint on rational trust, the narrow evidential constraint is false (Frost-Arnold 2014b; Simpson *ibid.*). It is too narrow, in two respects. First as we have seen, the evidence that rationalizes trust concerns more than whether the trustee will come through, and second, trust is made rational not just by the truster's evidence, but also by his value of the end that he seeks to promote.

We may summarize this observation as what I'll call the *broad evidential constraint*. This says that trusting someone to do something is only rational if it is sensitive to (i) the values that you assign to particular outcomes, and (ii) the balance of your total evidence, not just about whether the trustee will come through, but also about the likelihood that particular actions will yield particular outcomes. The broad evidential constraint brings the two apparently conflicting ideas into harmony: that trust is made rational by evidence, but that trust – even when rational – often exceeds evidence, at least about whether the trustee will come through.

I have left many issues out of this brief overview of trust. One is the question whether there is a unified account of trust at all or whether it is a pluralistic phenomenon (Simpson 2012; Kappel 2014); we have already seen that two competing phenomena are sometimes called “trust”, and I have followed the literature in calling one “trust” and the other “mere reliance”. Another important topic is what makes for *trustworthiness* (Jones 2012b; Simpson 2013; Daukas 2006) and whether this is a virtue. Yet another is what betrayal of trust amounts to (see Hinchman 2017 and Dormandy, this volume), and another is what value lies in trusting others as opposed to relying on oneself (see Simpson's and Fricker's contributions to this volume). These issues must remain undiscussed here in the interest of turning to the way in which trust features in epistemic contexts.

## 2. Epistemic Trust

Epistemic trust is trust for epistemic goods, such as knowledge or recognition as a knower (see section 2.2). Our gateway to a discussion of epistemic trust will be a topic in which it features heavily: the justification of beliefs formed on testimony. Examining this context for

---

<sup>14</sup> Some argue that the fact *that* you trust someone is relevant evidence, because your trust can encourage her to come through (McGeer 2008; Faulkner 2007; Pettit 1995; Hinchman 2012).

<sup>15</sup> Frost-Arnold (2014b) discusses two other types of extrinsic end for which one might trust.

<sup>16</sup> Simpson (2017), who introduces it, just calls it the “evidential constraint”.

epistemic trust will help us isolate some of its general features, including two types of epistemic good that we may trust other people for (section 2.2), and the characteristics that make it *trust* instead of mere reliance (section 2.3). I'll then discuss what makes epistemic trust rational (section 2.4). I will end with a discussion of the extent to which these aspects of epistemic trust in others apply to epistemic trust in oneself (sections 2.5 and 2.6).

### 2.1 Epistemic Trust in Testimony

Testimony occurs when one person, a speaker, tells something to another, a hearer.<sup>17</sup> In testifying, the speaker gives the hearer the option of believing what she says, on the basis of the fact that she tells it to him. Some have resisted the idea that belief on testimony can ever be epistemically beneficial (in this volume, see Simpson's discussion of Locke and Keren's discussion of revisionism), but contemporary epistemologists largely agree that belief on testimony, done right, is an important source of knowledge.

But much debate concerns what believing rightly on testimony amounts to – how the hearer's belief, should he take up the speaker's offer, can be justified or constitute knowledge. This concern arises because the hearer's evidence, if he has any, pertains not to the domain under discussion, but rather to the speaker (E. Fricker 2006). After all, if the hearer had enough evidence about the domain, he could justifiably form the belief on his own. Most views of testimonial justification agree that the hearer in a testimonial scenario trusts the speaker, but they differ over what relevance this fact has to the hearer's justification; so I will look at some views of testimony through the lens of the role that they assign to trust (but see Greco, this volume, for a more detailed taxonomy of views).

To make matters more tractable I will limit discussion to what Hinchman has called "consummated" testimonial exchanges (2005, 567) – that is, those in which the hearer *believes the speaker*. These are exchanges in which speaker purports to be trustworthy (i.e., sincere and competent), and the hearer believes what she says because he takes her at face value.<sup>18</sup> The question for theories of testimonial justification, then, is what justifies the hearer in *believing the speaker*? And our question in this section is what role the various theories of testimonial justification assign to epistemic trust.

One view, *reductionism*, says that what justifies the hearer's belief is the balance of his total evidence<sup>19</sup> about how likely the speaker is to speak truly (E. Fricker 1994, 1995; Lyons 1997; Lipton 1998; Adler 2002). This includes anything that may be relevant to her trustworthiness, most saliently any empirical evidence that he may have about her and speakers relevantly like her. If the speaker and hearer have a relationship of trust, then this fact gets thrown in with the rest of the hearer's evidence, typically raising the probability that she will speak truly. But if they have no such relationship, then the hearer simply takes account of whatever other evidence he has, such as the fact that the speaker is competent or that her interest in this case lies in speaking truly. According to reductionism, then, facts about trust do not play any special justificatory role other than being one consideration among

---

<sup>17</sup> I am using the terms *speaker* and *hearer* as synonyms for the precise but more unwieldy terms *teller* and *audience*. For simplicity I will adopt the convention of calling the speaker "she" and the hearer "he".

<sup>18</sup> Not every case in which the hearer believes what the speaker says, even if he believes it *because* she says it, counts as consummated. For example, the hearer may think that the speaker intends to lie but that, whenever she intends this, a cognitive quirk ensures that she in fact tells the truth. In such cases, the hearer believes what the speaker says because she testifies it, but he does not take her at face value by believing *her*: it is not a consummated testimonial exchange.

<sup>19</sup> This is not the standard characterization of reductionism, which usually specifies that the evidence is inductive or inferential; characterizing it as a total-evidence view maintains neutrality between these possibilities while leaving open other types of evidence too.

others that contribute to the hearer's total evidence about the speaker's likelihood of speaking truly.<sup>20</sup>

Reductionism is so-called because it claims that testimonial justification can be understood reductively, in terms of the familiar sort of justification made available by evidence. *Anti-reductionism*, by contrast, says that there is something special about the testimonial act that delivers justification in a way that bypasses the need for the hearer to appeal to his total evidence. On this view, there is only one type of situation in which a hearer's total evidence is relevant: when that evidence includes defeaters – that is, evidence indicating that the speaker will not in this case speak truly.

Versions of anti-reductionism differ over what it is about testimony that enables this bypassing of the hearer's total evidence. I will group these views according to the role that they assign to trust. What I will call *classic anti-reductionism* says that hearers have a special entitlement to believe a speaker. On some views, this entitlement is simply a priori (Ross 1986; Coady 1992; Burge 1993; Webb 1993). On others, it derives from the reliability of testimony or the social practices within which it occurs (Goldman 1999, 126–30; Goldberg 2010; Kappel 2014), or from the proper functioning of a cognitive mechanism for uptaking testimony (Plantinga 1993). Trust, on classic anti-reductionism, is simply the state that the hearer is in when he believes the speaker. It does not play a significant role in justifying the hearer's belief.

Another form of anti-reductionism is what I'll call *perceptual anti-reductionism* (M. Fricker 2007). Here, the epistemically unique aspect of the testimonial exchange is the hearer's perceptual experience of the speaker testifying. This perception portrays the speaker either as possessing properties that inspire the hearer's trust, or delivers a holistic impression of her *as trustworthy*. This perceptual experience, a tiny subset of the hearer's total evidence, is all the evidence that he needs, though Fricker adds the virtue-theoretic proviso that this perception must be produced by the hearer's epistemic virtues. Perceptual anti-reductionism has two alternative roles for trust. First, the hearer's perception of the speaker under a trustworthy guise may *cause* him to trust and thus believe her. Second, the hearer may happen to trust the speaker already, in which case his trust itself helps cause his perception of the speaker as trustworthy. In the first case, the hearer's trust is a consequence of his perception, so it has no independent justificatory role; in the second, his trust helps bring about his perceptual evidence and thus plays an indirect justificatory role. The epistemic role of trust on perceptual anti-reductionism, then, depends on whether the hearer antecedently trusts the speaker.

A final anti-reductionist view, whose proponents give it various names but which I'll follow Hinchman (2012) in calling the *assurance view*, gives trust the most significant justificatory role.<sup>21</sup> Here, trust brings about the hearer's testimonial justification *directly*. It can do this because, on the assurance view, it has epistemic significance in its own right.

There are three versions of the assurance view. The first two say that the trust at issue is the *hearer's trust in the speaker*. Of these views, one is *internalist*, saying that the fact that the hearer trusts the speaker is itself evidence<sup>22</sup> that the speaker is trustworthy, and this evidence is what justifies the hearer's belief. This version of the assurance view resembles reductionism in its internalist appeal to evidence alone. Yet unlike reductionism, it does not consider facts about trust just one sort of evidence among many. On the contrary, the

---

<sup>20</sup> This also holds for Lackey's "dualism" (2006), which adds a reliability condition to reductionism.

<sup>21</sup> Faulkner (2011) and McMyler (2011) resist the label "anti-reductionism", but they construe it more narrowly than I do, as something logically in between reductionism and anti-reductionism. On my broader usage, anti-reductionism is just the denial of reductionism.

<sup>22</sup> Proponents of the assurance view use the term "evidence" more narrowly than I do. On their view, this term is reserved for any reasons for belief *other than* the one supplied by the hearer's trust in the speaker. On my broader usage, this reason too counts as evidence.

internalist version of the assurance view says that the fact that the hearer trusts the speaker *makes any other evidence epistemically unnecessary*, indeed inappropriate. Why? Because, if the hearer were to consider additional evidence, this would, by definition, destroy the very trust that makes testimony justificatory (Faulkner 2011; McMyler 2011; Zagzebski 2012; Keren 2014; see also section 2.4 below). For trust, on this view, is incompatible with considering one's evidence concerning whether the trustee will come through. According to the internalist assurance view, then, the hearer can bring about his own justification simply by trusting the speaker.

The second version of the assurance view criticizes this move, on the grounds that it licenses a vicious bootstrapping (Hinchman 2012). But it maintains the idea that the hearer's trust in the speaker plays a direct justificatory role. It says that the hearer's trust, instead of providing the hearer with *evidence* that the speaker will testify knowledgeably, simply *causes* her to do so, or at least causes it to be likely that she will. Why? Because trusting someone sets in gear a response, on her part, to meet you in your dependence. This version of the assurance view, then, is *externalist*, positing a reliable process – trusting the speaker – that tends, when conditions are right, to elicit knowledgeable testimony (Hinchman 2012; cf. McGeer 2008).

The first two versions of the assurance view, internalism and externalism, have in common that the *hearer's trust in the speaker* is what plays the direct justificatory role. A third version of the assurance view is possible, on which the epistemically relevant trust is *the speaker's trust in the hearer*. The idea is that the speaker, in testifying, puts her epistemic reputation on the line: she entrusts it to the hearer, giving him leverage over it should she fail to testify knowledgeably; in this case he will think less of her and perhaps bring others to do so as well. The speaker's epistemic reputation, then, is like a drivers' license offered as collateral for an audio-tour headset: entrusted to the hearer, serves as a guarantee that she will not let down the trust that he, in turn, has invested in her for knowledge.<sup>23</sup> On this third version of the assurance view, then, the speaker's trust in the hearer directly justifies the hearer's belief on her testimony.

In summary, the prevailing views of testimonial justification all feature trust, but they do so in different ways. Reductionism says that trust can be one among many things that supplies evidence about the speaker's trustworthiness. Classic anti-reductionism construes trust as a background condition without any significant theoretical role; perceptual anti-reductionism says that the hearer's trust can play an indirect justificatory role; and the assurance view ascribes trust, on some accounts the hearer's and on others the speaker's, a direct justificatory role. Greco and McCraw, in their respective contributions to this volume, expand this discussion by arguing that there is a commonality across all cases of testimonial justification, whatever account you endorse. Greco argues that this commonality is the fact that trust is an essential constituent, and McCraw argues that it arises from the fact that the hearer exercises the virtue of proper epistemic trust.

The epistemology of testimony, with the exception of the third version of the assurance view, is concerned primarily with the hearer's trust in the speaker (a concern that Kallestrup, this volume, extends to *group* hearers and speakers). In order to achieve a more complete understanding of epistemic trust, we must redress this imbalance by contrasting it with the

---

<sup>23</sup> This possible version of the assurance view is similar to that of Moran (2005), who argues that the speaker's testimony functions as a guarantee, or assurance, to the hearer. Moran, however, does not focus on either party's trust in the other. Yet doing so, I suggest, would enable him to counter an important objection to his view. The objection comes from Faulkner (2007), who argues that the hearer, on Moran's view, has no way to know whether the speaker's guarantee is deceitful. If Moran were to understand the speaker's guarantee in terms her entrusting him with the collateral good of her epistemic reputation, this would supply the hearer with a reason to believe that she is being sincere and doing her best to testify competently.

trust that a speaker exhibits in a hearer. The next section discusses both forms of epistemic trust.

## 2.2 Epistemic Goods that We Trust for: Representational and Recognitional

To understand epistemic trust, we must understand the epistemic goods that we trust for. In the case of testimony there are two types of epistemic good: one that the hearer trusts the speaker for and another that the speaker trusts the hearer for.

When we think of trust for epistemic goods, it is the former that standardly come to mind: goods for which the hearer trusts the speaker, such as knowledge, evidence, or true beliefs. Grasswick (this volume) adds *inquiry*, which someone else can do on one's behalf. I'll call these *representational* epistemic goods, for they either produce, or are aimed at producing, cognitive states that accurately represent reality. In trusting a speaker for representational goods, a hearer makes himself vulnerable (to practical mishaps, false beliefs, and the like) should she inquire irresponsibly or be dishonest and thus testify falsely, unfoundedly, or unreliably; and he does so in the context of norms that hold the speaker responsible for her words and license the hearer to feel betrayed should she fail him.

Epistemic trust is typically defined *just as* trust for representational goods (e.g. McCraw, this volume; Keren, this volume). But speakers characteristically trust hearers too, for another kind of epistemic good that I'll call *recognitional*. To see this, consider that testifying is a way of claiming epistemic status as a knower of what you testify, standardly trusting the hearer to recognize you as having that status. Epistemic recognition comes in various forms. One way of according a speaker recognition is by believing her. Another is by giving her appropriate credit as the source of the information, granting her the final say in how her words are to be interpreted, and keeping confidence if she requests it.

A speaker who trusts a hearer for epistemic recognition thereby makes herself vulnerable (Dormandy, this volume). Should she deserve recognition and the hearer decline to accord it, she may be subject to epistemic harm (M. Fricker 2007; Dotson 2011; Peet 2017). For example, not receiving the epistemic recognition that she deserves may negatively impact her epistemic self-trust (see section 2.6 and Mikkola, this volume), and if this happens in front of third parties it may damage her social standing as an epistemic agent. The speaker may shy away from sharing knowledge in the future, which will in turn cut her off from leaving her mark on the community's store of knowledge, and will isolate her from discourse that might otherwise help her refine her own knowledge.

These harms may result from a speaker's receiving *less* epistemic recognition from hearers than she deserves. But some speakers – think of celebrities or authority figures – receive *more*, and this too can be epistemically harmful. For example, undeserved epistemic recognition can “spoil” a speaker, desensitizing her to her epistemic shortcomings and thus hampering her epistemic development in the long term (Medina 2013). She might place too much trust in her own reasoning or ignore her intellectual debts to others; even if she is competent in her domain, she may develop the bad intellectual habit of testifying too often, crowding out equally or more deserving speakers (Frost-Arnold 2014a).

Recognition is an epistemic good in its own right. We have just seen why this is so for speakers who merit it. But according deserving speakers epistemic recognition is also epistemically good for the hearer. Not only does this put him in a position to acquire testimonial knowledge and other representational goods from speakers, but recognition of deserving speakers *as* deserving is a form of epistemic success.

In summary, there are at least two kinds of epistemic good for which trust may be extended: representational goods, for which a hearer may trust a speaker, and recognitional goods, for which a speaker may trust a hearer.

Let's look more closely at epistemic trust, bearing in mind that it can be trust either for representational or recognitional goods.

### *2.3 Epistemic Trust: More than Mere Reliance*

We have been discussing how hearers and speakers can trust each other for epistemic goods. But they might also merely rely on each other. This occurs when one or both parties doubt that the other will abide by the norms of trust or have the requisite other-directed attitude. A hearer merely relies on a speaker for the truth (for example) if he thinks that, given the chance, she would lie or testify carelessly, and that it is pure self-interest that keeps her from doing so. The speaker may for instance be witnessing in a criminal investigation for the sake of reducing her own sentence. Similarly, a speaker can merely rely on a hearer for epistemic recognition if, for example, she thinks that he does not want to accord it. In this case she may testify because she has little choice – she may for instance have been subpoenaed to testify before a jury that she knows is biased against her.

Given that the exchange of epistemic goods can be powered by trust or mere reliance, we may ask which of these is standard or at least more prevalent. There is reason to think that trust, rather than mere reliance, is the norm (see Grasswick, this volume, and Keren, this volume, footnote 6). For epistemic contexts standardly legitimate reactive attitudes. A hearer may (and arguably should) feel betrayed if a speaker lies or fails to do diligence to her evidence; and the speaker may (and should) feel similarly if a hearer, for no reason, accords her less credibility than she deserves. This is certainly true of ongoing epistemic relationships (Grasswick, this volume); but even in the relationally pared-down case of a local stranger directing a tourist to the train station, it is appropriate for the tourist to feel grateful for good directions and betrayed by bad ones, and for the local to be offended if the tourist, for no apparent reason, disbelieves her.

One might think that science is an exception – that it is not trust, but mere reliance, that governs the exchange of epistemic goods. Because scientists attach such value to perceived epistemic prowess, it is surely naïve to expect that the norms of trust, let alone other-directed motives, will be enough to keep them epistemically on the straight and narrow. Indeed, is this not precisely why science comes with such stringent professional norms, which sanction shoddy data gathering or the usurping of credit due to others? These norms, and the sanctions that accompany them, aim to ensure that it is in scientists' self-interest to be epistemically reliable collaborators.

Yet closer inspection points to trust, not mere reliance, as the epistemic motor of science. One reason is that a culpable failure to come through in science licenses reactive attitudes: falsified data and the usurpation of credit are not just annoying, but personally offensive. Such reactions would not be legitimate in relationships of mere reliance. A second reason is that the norms of science are arguably inadequate to deter epistemic betrayal; the prevalence of trust explains why science can function in spite of this. When it comes to the exchange of representational goods (e.g. knowledge or evidence), Hardwig (1991) argues that professional norms are inadequate to deter dishonesty on the part of speakers, such as the falsification of data, one reason is that the chances of getting caught are slim, since science is too fragmented for experts in one field to check up on the experts in another on whom they depend. Yet scientists continue to depend on each other for knowledge. And when it comes to the exchange of recognitional goods, Frost-Arnold (2013) argues that the professional norms of science are inadequate to deter hearers from maliciously declining to give epistemic recognition to deserving speakers, as when senior scientists usurp credit for the for results achieved by their junior colleagues. Yet scientists do depend on each other for due credit. If Hardwig and Frost-Arnold's arguments are on target, it would seem that a large dose of *trust*, as opposed to mere reliance, is needed. What scientists depend on in colleagues from other

fields, and what junior scientists seek and often find in senior colleagues, are personal commitment, other-directed motives, and moral sensibilities (Hardwig *ibid.*, Frost-Arnold *ibid.*). And if science, which due to its professionalism is among the least “touchy-feely” of domains, is marked by epistemic trust rather than mere reliance, then all the more so will other sorts of more casual epistemic relationship be.

If we take seriously the idea that the exchange of epistemic goods is standardly powered by trust instead of mere reliance, then another aspect of epistemic trust becomes apparent. Trust, as we saw, comes with norms governing a cooperative relationship of care between persons; this means that we can expect epistemic trust relationships to involve a species of *epistemic care* (Hinchman 2012; Grasswick 2018 and this volume). What does epistemic care look like in the case of the two types of epistemic good?

When it comes to the hearer’s trust in the speaker for representational goods, Hinchman (2012, 2017) and Grasswick (2018 and this volume) give an answer. Hinchman notes that people’s needs, including their epistemic ones, differ from one context to another. Accordingly, he argues, a hearer trusts a speaker for more than just knowledge; he trusts her for knowledge *tailored to his specific epistemic needs*. Hinchman gives the example of a hearer with a nut allergy who asks about the contents of a snack bowl (*ibid.*): for this hearer, much more hangs on whether the bowl contains nuts than it would for a hearer without such an allergy. Fulfilling this hearer’s trust requires a speaker to bear in mind the dire consequences that false information would have for him, and thus to hold herself to especially high evidential standards. Another example, not discussed by Hinchman, is that of our tourist asking for directions to the train station. If there are road closures, this hearer has different needs than he would if he asked for directions when the roads were open. A local speaker, advising him, should bear the closures in mind. While she is at it, she does well to envision how he might not notice the turnoff, which is easy to miss (it is landmarked by a red postbox). And she should *not* send him on an indirect route via the café that she owns. Cases like this show that what hearers trust speakers for is not just information in the abstract, but epistemic care in the form of information tailored to their needs. This fact is all the more evident in cases of the sort that Grasswick discusses (2018 and this volume), in which we are not trusting someone to give us knowledge, but to carry out inquiry on a topic about which we may never, ourselves, come to know, such as medicine or national security, but that is nonetheless important to us.

A parallel point applies to trust for recognitional goods. The speaker, in trusting the hearer for these, trusts him too for a kind of epistemic care suited to her context. This includes trusting him to accord her neither too little nor too much credibility, to bear in mind any practical consequences that she might face should her testimony be leaked, to triangulate her need to be heard with that of others whom she does not wish to inadvertently speak over, to understand her testimony well enough to represent it accurately to third parties (and ask follow-up questions otherwise), to competently determine which third parties can be told and which cannot, and so forth. The contextual factors that delineate a speaker’s epistemic needs are legion, and a hearer who accepts her trust must be alive to them.

In summary, taking seriously the fact that epistemic trust is *trust* instead of mere reliance helps us see that what we epistemically trust others for is epistemic care: not just information or recognition, but these goods tailored to our epistemic needs in our contexts.

Having looked at some descriptive features of epistemic trust, I will now return to the question of its normativity.

#### 2.4 Rational Epistemic Trust

Epistemic trust, like trust in general, can be rational or irrational. Consider first a speaker’s trust in a hearer for epistemic recognition. This trust is rational in the same way that trust in

general is rational: that is, just in case it is the best available means to achieve whatever aim the speaker has in mind in trusting the hearer with her words. This end might simply be to obtain recognition, but it might also be something else, such as furthering her message at all costs, or teaching the hearer to accord epistemic recognition appropriately. Thus if trust in general, as I argued in section 1.3, is subject to the broad evidential constraint, then so is trust for epistemic recognition. That is, a speaker's trust in a hearer for recognition is rational only if the speaker takes into account (i) the values that she assigns to particular outcomes of trusting as opposed to not trusting, and (ii) the balance of her total evidence: about whether the hearer will in fact accord her recognition, but also about whether trusting him is the best available way to achieve her valued outcomes (including any costs incurred by the trust itself).

Matters are more complicated when the rationality in question is that of a hearer trusting a speaker, where what he trusts her for are representational goods such as knowledge or evidence. In this case, there are two kinds of rationality to consider. One is the decision-theoretic rationality that I have been discussing so far. Here, the rationality of the hearer's trust depends, in the familiar way, on (i) the value that he assigns to the outcome of trusting as opposed to not trusting the speaker, and (ii) the balance of his total evidence about whether trusting the speaker is the best way to achieve his valued outcomes.

When it comes to decision-theoretic rationality, the hearer trusting the speaker for representational goods might have a variety of aims, non-epistemic or epistemic. First, the hearer's aim might not be epistemic. It might instead be to achieve some other end that believing the speaker would further. This might for instance be to avoid conflict with the speaker, or, if she is testifying optimistically that he will recover from an illness, to further his prospects for doing so by believing her. In such cases it could be rational, in the decision-theoretic sense, to trust the speaker even if his total evidence does not favor her trustworthiness (at least, as long as he can bring himself to believe her despite insufficient evidence).

Second, the hearer's aim may be epistemic: to obtain the representational good for which he is considering trusting the speaker (e.g., knowledge, evidence, the truth). If his total evidence indicates that she is trustworthy and that no alternative means would secure his aim more effectively, then it is straightforwardly rational to trust her for it. But what if the hearer's total evidence does not indicate that the speaker is trustworthy? He might still, in some circumstances, be rational in trusting her in the current, decision-theoretic, sense. For example, suppose that he values the prospect of forming a true belief more than he disvalues the prospect of forming a false one or none at all, and suppose that there is no better way available to form a true belief than to trust this speaker. (His friend may claim to be innocent of a crime that she is accused of, and he wants to believe truly – but places greater value on a true belief that she is innocent than on a false belief that she is guilty, since the latter would ruin their friendship.) In this case it could be decision-theoretically rational to form the belief – betting, so to speak, on its truth.

In summary, the first form of rational trust for representational goods is strictly decision-theoretic, depending (among other things) on the hearer's aims in trusting.

There are some sorts of epistemic aim, however, that mere decision-theoretic rationality cannot achieve for the hearer. We saw that this sort of rationality can help him form true beliefs in certain circumstances. But if his aim is to obtain something more normatively laden, such as knowledge or evidence, this sort of rationality is not enough. For even if he wound up forming a true belief, it would not count as knowledge *unless it were also epistemically rational* – that is, unless it were appropriately related to something like evidence, a virtuous epistemic character, or the like. Epistemic rationality, the second type of rationality governing



trust for representational goods, we have seen already; I have been calling it *epistemic justification* (section 2.1).<sup>24</sup>

On the face of it, epistemic justification looks like decision-theoretic rationality, just applied to cases in which the agent aims for the truth. But this conclusion is too quick. First, a belief can be decision-theoretically rational with respect to this aim, and yet fail to be epistemically justified. A case in point is the friend scenario above, in which the hearer's evidence does not support the trustworthiness of his friend about her innocence but other considerations, such as the value he places on believing truly compared to the possibility of believing falsely, make it rational to trust her anyway.<sup>25</sup> Second, a belief can be *irrational* in the decision-theoretic sense above and yet epistemically justified. For example, a person may aim to keep himself in the dark about the truth of some proposition, such as the end of a film that he has not yet seen – but if his friend inadvertently reveals it to him, he would be epistemically unjustified not to believe it, even though doing so goes against his aim (Kelly 2003). So when it comes to epistemic justification, a person's values play a smaller role than they do in determining decision-theoretic rationality.<sup>26</sup> Epistemic justification is an *epistemic good*, whether or not any particular agent desires it or any other end it might promote.

Epistemic justification, then, differs from decision-theoretic rationality: it is not subject to the broad evidential constraint. For this constraint, as we saw in Section 1.3, takes on board not just an agent's total evidence about whether the trustee will come through, but also the truster's values, and the likelihood that trusting (as opposed to doing something else) will lead to his desired outcome. Yet the latter two considerations, the truster's values and aims, do not factor in epistemic justification, at least not as saliently (see footnote 26). The only factor that is a candidate for relevance to epistemic justification is the first: the hearer's total evidence about whether the speaker will come through. But this is just the narrow evidential constraint. Recall that I rejected this constraint as being too narrow to apply to rational trust in general – but perhaps this makes it just the constraint to apply to epistemic justification. In the context of epistemic justification, the narrow evidential constraint becomes the claim that trusting a speaker for knowledge (or the truth, etc.) is justified only if your total evidence on balance supports the belief that she is trustworthy.<sup>27</sup>

We have seen this claim before (section 2.1): it is reductionism about testimonial justification. The various forms of anti-reductionism, by contrast, deny that a hearer is only justified in believing the speaker when his total evidence supports her trustworthiness. And some forms of the assurance view's version of anti-reductionism, as we saw, go as far as to say that the hearer should *not* employ any evidence beyond the speaker's testimony itself. Epistemically rational hearer trust, then, is a controversial matter – but whatever it amounts to, it presents a different kind of rationality than the decision-theoretic sort.

I'll finish this section with a brief discussion of the extreme claim propounded by versions of the assurance view, that epistemically justified testimonial belief involves *not* making recourse to any evidence beyond the testimony itself. What would recommend this view? One answer is that the speaker, being better situated than the hearer with respect to the truth, is likely to base her belief on better evidence than he is – and that for this reason the hearer would do well to disregard his other evidence (Keren 2007, 2014, this volume; cf. Zagzebski

---

<sup>24</sup> Some authors keep these notions distinct, but for simplicity I will use them interchangeably.

<sup>25</sup> That said, some argue that beliefs can be epistemically justified in some cases even when evidence does not support them, as long as the evidence does not actively speak against them (e.g. James 1921). But I would argue that this view conflates epistemic justification with decision-theoretic rationality.

<sup>26</sup> This is not to say that the agent's aims play *no* role determining what is epistemically rational. For agents themselves have a measure of control over the evidence that they are exposed to. And some argue that it is up to an agent whether to adopt an epistemically risky policy (forming more beliefs even if some might turn out false) or an epistemically cautious one (forming fewer beliefs in order to avoid falsehoods at all costs) (James 1921).

<sup>27</sup> This need not entail that the hearer must explicitly hold the belief that the speaker is trustworthy.

2012; Sandis 2011; Constantin and Grundmann forthcoming). Another answer is that the norms of the trust relationship are excellent guarantors of the speaker's trustworthiness – and that a hearer who seeks additional evidence violates them, thus risking this relationship (Faulkner 2007; Moran 2005; Hinchman 2012; McMyler 2011).

The assurance view is surely on to something in claiming that the trust relationship itself is epistemically significant. But it must still address why it is so significant as to upstage appeals to additional evidence, especially if such evidence happens to be easily available (or even possessed by the hearer already). That is, we need a response to the main argument supporting the narrow evidential constraint for epistemically justified testimonial belief. This argument is that a hearer's total evidence about the speaker is likelier to provide a more complete, and thus more accurate, sampling of reality than the small subset constituted by the fact of her testimony (Dormandy 2018). The answer that Keren provides (this volume) is that the hearer's trust in the speaker enables him to base his belief, vicariously, on *her* evidence, which is likely to be better than his. Whether this idea pans out depends on whether one person can really be said to base his belief not just on the say-so, but on the evidence (to which he himself may have no access) of another person.

In summary, epistemic trust for recognitional goods is subject to the norms of decision-theoretic rationality just like trust in general. Epistemic trust for representational goods is subject to this kind of rationality in certain cases, but is more properly governed by *epistemic* rationality (or epistemic justification), which swings freer from many of the agent's own aims and values. Whether the narrow evidential constraint applies to it, however, is controversial.

## 2.5 Epistemic Self-Trust as Rationally Necessary

It is not just other people in whom we invest epistemic trust – it is also ourselves. Epistemic self-trust is of great intrinsic value (Fricker, this volume). More than this, it is arguably a causal and rational precondition for all of our epistemic endeavors. A body of literature has developed that aims to show why this is so. This section sketches three prominent views of trusting oneself for representational goods, and section 2.6 considers whether epistemic self-trust is also possible for recognitional ones.

Each argument for the rational necessity of epistemic self-trust works from a different construal of epistemic (as opposed to decision-theoretic) rationality. On Foley's view (2001), to be epistemically rational is to be able to defend yourself against epistemic criticism. His view might be thought to cast us all as irrational, for there is an important epistemic critique, launched by philosophers such as Descartes and Locke, against which we seem unable to defend ourselves: the fact that we cannot non-circularly prove the reliability of our cognitive faculties. For the belief that they are reliable relies on the assumption that they were reliable in bringing us to hold that very belief. But Foley responds that we can count as rational on his view, namely by exercising epistemic self-trust. He grants that this "leap of intellectual faith", as he styles it (*ibid.*, 19-20), is our only option – but he insists nonetheless that it is a rational one. More than this, he insists that self-trust is rational *precisely because* we have no alternative. The idea seems to be that what is rational must, on pain of violating the maxim that *ought implies can*, be located within the range of things that we are able to do. Since all that we are able to do is exercise epistemic self-trust, it is rational to do so.

The first response is that it is not obvious that we can do nothing other than exercise epistemic self-trust. There are two ways in which declining to trust yourself for knowledge and other representational goods seems possible. First, it seems possible to decline to rely on yourself for knowledge and related representational goods, at least in isolated cases when you can muddle along without a belief on a particular matter – though Foley is right that a global failure to rely on yourself would be paralyzing. Second, even in the cases where you must rely on yourself, this reliance need not amount to trust. For you might lack the requisite other-

directed attitudes (such as good will or commitment) toward yourself, and you might not normatively expect yourself to have them. This is arguably not an epistemically healthy way to be, but it is possible. And if this is so, then self-trust is not rational simply because it is the only option. At best, self-reliance on a global scale, whether or not this also amounts to self-trust, is the only option – so at best it is this, not self-trust, that Foley shows to be rational.

Another worry is that Foley's claim that self-trust (or self-reliance) is by default rational might rationalize a dogmatic adherence to whatever our faculties deliver, for better or worse. Foley resists this consequence by arguing that, even though trusting (or relying on) himself is all that an agent can do, some ways of doing this are epistemically preferable to others. We should not, for example, rely on each individual faculty by default, but should use our other faculties to critically reflect on it. Even though doing this requires trusting those other faculties, we can test them in turn against each other and against the first faculty. Epistemic self-trust (or self-reliance), then, is rational on Foley's view because we can do no other – but only when combined with critical self-reflection.

Foley views epistemic self-trust negatively, as a fallback position given our lack of a non-circular proof of cognitive reliability. This attitude makes sense if rationality, as he claims, is a matter of defense against epistemic criticism: a non-circular proof would make for a stronger defense. But two other views, Lehrer's (1997) and Zagzebski's (2012), do not see the lack of such a proof as an epistemic liability. For they locate epistemic rationality in coherence, and on coherence views, proofs *just are* circular. Where Lehrer's and Zagzebski's views part ways is in how they construe coherence and thus rational epistemic self-trust.

On Lehrer's view (1997), epistemic rationality – or “reasonableness”, as he calls it – amounts to coherence among a person's beliefs. This means that two beliefs can epistemically support each other, either directly or by supporting intermediary beliefs. (Opposed to a coherentist view of epistemic support is a foundationalist view, on which epistemic support only goes one way.) What holds a coherent belief system together, on Lehrer's view, is epistemic self-trust. More specifically, what holds it together is the agent's acceptance of Lehrer's “keystone premise”: that she is trustworthy concerning the beliefs that she accepts. What makes acceptance of the keystone premise rational? Like any self-respecting coherentist, Lehrer argues for it circularly – indeed with two (circular) arguments. The first says that the keystone premise, a fortiori, implies the further claim that the agent is epistemically trustworthy in accepting the keystone premise itself. This further claim, in turn, makes her acceptance of the keystone premise rational – because it is rational to accept propositions with respect to which you are trustworthy. Lehrer's second argument says that the keystone premise supports everything else that the agent believes, and that a subset of those things in turn supports the keystone premise. In this way epistemic self-trust is rationally necessary, but also epistemically justified.

A similar dogmatism worry arises for Lehrer's view as for Foley's. The keystone premise does not exclude the agent's taking herself to be *infallibly* trustworthy in the beliefs that she accepts, but this would by her lights lend infallible support to the keystone premise (Cohen 1999). Lehrer might respond by claiming that the other beliefs that the agent accepts do *not* support the infallibility version of the keystone premise, preventing her from being rational in accepting it. But this response only holds for agents whose other beliefs are like this. In principle an agent could accept only beliefs that support her own infallibility (*ibid.*), and come out justified on Lehrer's view in so doing.

Whereas Lehrer locates rationality in the coherence among an agent's beliefs, Zagzebski (2012) takes this notion to apply to broader whole that also includes his affective states and willings. Being epistemically rational, she says, amounts to conscientiously eliminating any cognitive dissonances that there may be among all of these aspects of oneself. Epistemic self-trust emerges on her view as the most effective means of doing so: a person should examine his cognitive, affective, and volitional states, and resolve dissonances in favor of whichever

he trusts the most. One might worry, however, that there are some cognitive dissonances that not even self-trust can resolve, such as those in which his self-trust itself conflicts with some other state that delivers reasons against trusting oneself. For example, an agent might think that skeptical worries about his ability to know rationally undermines the self-trust that he exercises in forming perceptual beliefs. How, on Zagzebski's view, can self-trust overcome this obstacle to coherence? Her answer is that epistemic self-trust always rationally wins out, even over cognitive states that yield reasons to question it. Why? Because we naturally trust *our self-trust itself* more deeply than any of our other states (*ibid.*, 45). So on Zagzebski's view, epistemic self-trust is rationally mandated for creatures like us – or at least for those of us who do invest greater trust in our self-trust itself than in other states. (This leaves open the possibility that there are people for whom self-trust is *not* rational, namely those of whom this empirical claim is false.) Zagzebski's position, like the other two, might seem to legitimate an uncritical dogmatism. Her response is similar to Foley's: she mandates that self-trust be exercised conscientiously.

Yet it is not clear that appeals to internalist epistemic abilities like conscientiousness or the capacity for self-criticism can play the rationally legitimating role that Zagzebski and Foley claim for them. As Mikkola (this volume) shows, human cognition is subject to many adverse influences at the implicit level that make many of us prone to overestimate or underestimate our cognitive status and abilities. She argues that these factors can sabotage the abilities of certain people – for example, the marginalized in society who are targets of negative epistemic stereotypes – in their attempts to exercise conscientious epistemic self-trust, because they will be overly cautious or self-critical. She argues that epistemically beneficial self-trust cannot be promoted by prescribing individual measures such as conscientiousness, but requires social education that aims to uproot these pernicious biases.

The views discussed in this section aim to establish the rational necessity of epistemic self-trust. What epistemic trust *is*, however, gets less discussion. The next section turns to this question.

## 2.6 Epistemic Self-Trust is More Than Reliance

The three views just discussed differ about how and why epistemic self-trust is rationally necessary, but they share an assumption about what it is. Epistemic self-trust, they seem to agree, is not trust as such, but mere reliance. To see this reading, note that the attitude that the authors have in mind is not directed at *oneself*, but rather at one's "opinions and faculties" (Foley, 2001). Zagzebski (2012) for her part says, "I trust my epistemic faculties to get me to the truth. Trusting myself in this sense is like trusting my car" (36 cf. E. Fricker 2016, 154, who says explicitly that epistemic self-trust is a matter of (mere) reliance on one's faculties).<sup>28</sup> Yet opinions, faculties, and cars are things, not persons, even though they belong to and are steered by persons. As such they can be relied on, but are not proper objects of trust in the morally normative sense at issue here. The importance of this point must not be underestimated. If epistemic self-trust is no more than reliance, then it is a fundamentally different sort of thing than (most of) our epistemic trust in others.<sup>29</sup>

One clue that there may be more to epistemic self-trust, however, comes from Zagzebski herself. She makes a point of characterizing a self as being "conscious of itself" and "the inner world of a person" (29), neither of which applies to a person's faculties or opinions taken on their own. This rich talk of selves leads us to suspect that Zagzebski, in later

---

<sup>28</sup> Lehrer does not the object of self-trust one way or another, but reliance on one's faculties suffices for his argumentative purposes.

<sup>29</sup> Tanesini (this volume) is one of the few who discuss this matter explicitly. She does not claim wholesale that epistemic self-trust is a matter of mere reliance. Rather, epistemic self-trust is a complex state involving mere reliance on one's faculties, on the one hand, and trust in the normative sense in one's own *will*, on the other.

comparing the object of our self-trust to a car, sells self-trust short even by her own lights. There is a similar incongruence in Hinchman (2012). As we saw above, he takes epistemic trust to be trust for “a distinctively epistemic species of care” (80) that can only arise in a personal relationship. Yet in discussing how trust “is realized intrapersonally” (77), he says that “you rely on your epistemic faculties” (*ibid.*). Faculties, however, cannot care for you – so it would seem that epistemic self-trust on Hinchman’s view cannot be a type of trust, as opposed to mere reliance, after all.<sup>30</sup>

The deflationary view on which epistemic self-trust is mere reliance on one’s faculties should be rejected. Although we do rely on our faculties, we also extend epistemic trust to ourselves *qua* persons – and this claim gives a better account of how we secure epistemic goods. Before arguing this, it is worth establishing that trust in oneself is a psychological possibility – that is, that a person can direct the normative expectations and psychological attitudes characteristic of interpersonal trust toward herself, and that she can accept her own self-trust accordingly.<sup>31</sup>

To see this, consider normative expectations. We can let ourselves down, a situation that normatively licenses reactive attitudes such as disappointment in ourselves. And we can and sometimes should be grateful to ourselves: although we do not tend to speak this way, teachers of self-coaching will be quick to assure us that it is a coherent sentiment and will encourage us to thank ourselves or pat ourselves on the back for job well done. And they may explicitly encourage us to hold ourselves to high expectations on the grounds that “you know you can do it” – where the second “you” refers to the addressee, not (just) his faculties. So we do seem to have, and respond to, normative expectations of ourselves that look very much like those accompanying trust. As for the relationship-responsive attitudes, a person can have good will toward herself, can respond to her own needs, and can be committed to living up to her own expectations. To trust herself she must simply rely on herself to respond to these attitudes.

But granting that trust in oneself is a live possibility, why think that it is this, rather than mere self-reliance, that secures us epistemic goods? There are two questions here. One is: why think, when a person *relies* on himself for epistemic ends, that this reliance is in fact on his self and not just his faculties? The second is: supposing that the object of his reliance is his self as opposed to his faculties, why think that, in the epistemic case, this self-reliance has the additional features that make for self-trust? I’ll answer each question for self-trust for representational epistemic goods, then turn to recognitional ones.

So consider representational goods, such as knowledge, true beliefs, evidence, and inquiry. In answer to the first question (i.e., why think that it is a person’s self that he relies on for these, not just his faculties): although a person does rely on his faculties for these goods, he himself occupies his faculties’ “control center”. It is he who determines how and when to rely on them, and who is thus epistemically responsible for much of their output. This means that he does not rely only on his faculties, but also on *himself* to manage them responsibly. Indeed, some go as far as to make epistemic responsibility – and thus reliance on oneself to exercise it – a necessary condition for representational goods such as knowledge (Code 1987; Montmarquet 1993; Zagzebski 1996; McCraw, this volume). Foley and Zagzebski come close to agreeing recall that both, in response to the objection that their views of epistemic self-trust seem to license dogmatism, appeal to epistemic responsibility, alternatively called critical self-reflection or conscientiousness, as the antidote. Yet epistemic responsibility presupposes reliance on oneself, *qua* person, to exercise it. Reliance on one’s faculties is not enough.

---

<sup>30</sup> Jones (2012), by contrast, seems more sympathetic to a view of epistemic self-trust as directed at oneself *qua* person, rather than simply at one’s faculties, speaking of epistemic self-trust as a normatively rich phenomenon.

<sup>31</sup> Govier (1993, 106-109) makes a similar argument for non-epistemic trust in oneself.

Let's turn to the second question, why this reliance on oneself for representational goods is best construed as *trust* in oneself. The answer is that self-trust is more epistemically effective than mere self-reliance. For the person in a trust relationship with herself has more factors in place that promote epistemic success than the person who merely relies on herself. One factor is the internal psychological arrangement characteristic of trust: the self-truster holds herself to normative expectations, including to not let herself down epistemically; and she is inclined to respond positively to her own epistemic needs. The (mere) self-reliant, by contrast, relies on herself to do what she would do irrespective of considering her own needs – an arrangement apt to be subject to the impulses of the moment and inimical to cultivating virtuous epistemic habits.

A second factor favoring the success of the self-truster is this. Trusting someone signals that you construe her as trustworthy: as someone who will respond to your normative expectations, have a positive attitude toward you or your needs, and who will competently come through for you. And people tend, all else equal, to unconsciously conform their self-concept, and thus their behavior, to the way in which others construe them (Hinchman 2012; Alfano 2012, chap. 5; D'Cruz 2019). Merely relying on a person, by contrast, signals just that you construe her as someone who will behave in the usual predictable ways, irrespective of her relationship to you. This gives her no positive vision to conform to, and possibly even a negative one. If trust, as opposed to mere reliance, can exert this sort of psychological influence from one person to another, it is apt to do so all the more forcefully within the psyche of a single person. These two factors put in place by trust, over and above mere reliance, make the self-truster likelier to enjoy epistemic success than the (mere) self-reliant. So not only do we rely on ourselves rather than just our faculties, but this reliance, when epistemically effective, has the marks of *trust* in oneself.<sup>32</sup>

Let's turn to recognitional epistemic goods, to which a similar line of thought applies: self-trust for these cannot adequately be construed as mere reliance. Taking our two questions in order: first, why is the reliance at issue reliance on oneself and not just on one's faculties? The reason here is ontological: epistemic recognition is not the kind of thing that a mere object can accord. Second, why is our self-reliance for recognitional goods best construed as self-*trust* as opposed to mere self-reliance? The answer is that, although one *could* merely rely on oneself for epistemic recognition, trusting oneself is more effective at securing it. Just as with representational epistemic goods, in the case of self-trust for recognitional ones, there are psychological factors in place that promote appropriate ascriptions of epistemic recognition to oneself, whereas these factors are not in place in the case of (mere) self-reliance.

To see this, note that according oneself appropriate epistemic recognition is not always easy; it often requires virtues such as courage, honesty, and discernment (Tanesini, this volume). It involves (for example) sticking up for what you know when you come under pressure to self-censor, yet being unafraid to admit when you do not know; it involves insisting on the epistemic credit you deserve when others belittle you or try to usurp it; it involves ceding the floor to others who are equally epistemically deserving but less influential than you are; it involves believing the conclusions that you draw on the basis of good reasoning, yet declining to believe conclusions that you draw for poor reasons; and so forth. These activities are not for the faint-hearted, but require courage, honesty, and discernment.

The self-truster is in a better position to exercise these virtues than the (mere) self-reliant, for reasons we have seen. First, the self-truster relates to himself a person, with the accompanying normative expectations and counting on himself to meet his own needs. That is, he holds himself to normative expectations to accord himself appropriate epistemic recognition, and he counts on himself to be responsive to his own need for such recognition.

---

<sup>32</sup> McCraw (this volume) goes as far as to argue that epistemic self-trust is an epistemic character virtue.

Someone whom you treat as a person – including yourself – is apt to be moved by the norms and the personal attitudes to do things in the context of the relationship that he might not do otherwise. The mere self-reliant, by contrast, lacks this internal arrangement with himself, relying only on what he would do anyway. Second, the self-truster, *in* trusting himself for epistemic recognition, thereby signals that he construes himself as trustworthy for it, thus encouraging himself to act in a trustworthy manner by according it in appropriate measure. The mere self-reliant, by contrast, lacks this positive vision. So epistemic trust, for recognitional no less than representational goods, is best made sense of as a form of self-trust, as opposed to reliance either on one’s faculties or (mere) reliance on oneself.

Section 2 has looked at epistemic trust in others through the lens of testimony, and applied some of the resulting insights to the case of epistemic trust in oneself. I do not pretend to have exhausted the epistemic phenomena in which trust plays a role. And I have limited this discussion to trust between (or within) individuals, though an important area for continued research is the way in which epistemic trust, both for recognitional and representational goods, features between groups (see Kallestrup’s contribution in this volume) and epistemic communities (Grasswick 2018). But I hope that this exploration of epistemic trust through the lens of testimony and self-trust has laid some groundwork for future discussion. The contributions to this volume aim to advance many of them, and I turn to them now.

### 3. Chapter Overview

#### 3.1 *The Value of Trust in Others and in Oneself*

In “Locke on Trust,” Thomas W. Simpson discusses the value of trust, starting with an apparent puzzle in the work of John Locke and finishing with his own alternative view. On the one hand, Locke’s political philosophy places a high positive value on trust. On the other hand, in Locke’s epistemology trust is presented as the ultimate epistemic sin. The key to resolving this apparent disunity in Locke’s views, Simpson argues, is to recognize that Locke thinks that the only possible value that trust can have is instrumental. This means that, in the case of political trust, trust is valuable because it sets up norms according to which the state is responsible to its citizens. In the case of epistemic trust, by contrast, Locke sees only disvalue, because he takes it merely to discourage individuals from thinking autonomously. So Locke’s view is coherent given the premise that the only value to trust is instrumental. But Simpson rejects this premise, arguing that trust has great *intrinsic* value, whether in politics or epistemology.

Value is also the focus of Elizabeth Fricker’s “Epistemic and Practical Dependence and the Value of Skills or: Satnavs, Good or Bad?” In counterpoint to Simpson, who explores the non-instrumental value of epistemic and practical trust, Fricker focuses on the non-instrumental value of *not needing to trust others*. Fricker acknowledges that there is great positive value in trusting others for knowledge and other things, but she explores the idea that there is non-instrumental value in being skilled to know and do things for ourselves. What things, and under what circumstances? Fricker argues that, of metaphysical necessity, each one of us has reason to acquire a certain set of skills the exercising of which will bring him pleasure. She then argues that there is non-instrumental value in having some skills (as opposed to being entirely unskilled), which derives from the fact that, given the sorts of beings we are and the sort of world we inhabit, we derive self-respect from having skills. Finally – and most controversially – Fricker argues that there is a specific set of core skills that every person, given the sorts of beings we are and the sort of world we inhabit, has reason to acquire. Taken together, Simpson’s and Fricker’s chapters yield a complementary picture

of the non-instrumental value both of trusting others and of having the skills on the basis of which to trust oneself.

### 3.2 *Trust in Testimony*

We move from the value of trust across epistemic and non-epistemic domains to a discussion of its specifically epistemic value, in Greco's chapter, "The Role of Trust in Testimonial Knowledge". Against the idea that interpersonal trust cannot be epistemically interesting in its own right because, unlike such notions as evidence and reliability, it does not help constitute epistemic aims, Greco argues that *trust is essential to testimonial knowledge*. He offers two arguments, both of which depend on the premise that testimonial knowledge is transmitted knowledge. But Greco then argues that the transmission of knowledge essentially involves joint agency – which in turn essentially involves interpersonal trust. Along the way Greco casts new light on social-epistemological categories, distinguishing two forms of reductionism (source reductionism and transmission reductionism) and situating so-called "trust theories" of testimonial justification, such as the assurance view, within this framework. His chapter yields a new, finer-grained, way to think about the epistemic value of trust, and through this about epistemic value in general.

In "Trust, Preemption and Knowledge", Arnon Keren also discusses trust in testimonial knowledge. Whereas Greco focuses on the necessity of epistemic trust for testimonial knowledge, Keren aims to give an account of epistemic trust itself. He claims that trust is a matter of declining to take precautions against the trustee's failing to come through, and that this amounts in the epistemic case to declining rely on evidence for the testified proposition, instead relying solely on the testifier. But if this is so, how can trust play a positive epistemic role, rather than a negative one? The key, Keren argues, lies in recognizing that trust is *preemptive*: Trusting someone entails believing that she is trustworthy, and this belief *preempts* any other evidence about whether she is trustworthy. In other words, this belief gives the agent a good reason to desist from relying on any evidence other than the trustee's word. But if trust is preemptive, how is it compatible with epistemic responsibility, which seems to involve relying on your own evidence? Because, Keren claims, preempting your own evidence in favor of the testifier's say-so enables your belief to be supported by *her* evidence – which, we may assume, is superior to your own. Far from abandoning epistemic responsibility, then, epistemic trust on the preemptive account gives you justificatory access to a swathe of evidence that you would not otherwise have had.

The contributions thus far focus on epistemic trust between individuals, but Jesper Kallestrup extends the discussion to epistemic trust between groups. In "Groups, Trust and Testimony", he argues that, if groups can have knowledge and can testify in ways that cannot be understood reductively in terms of the knowledge or testimony of their individual members, then groups can also engage in relations of epistemic trust. He offers some considerations supporting the antecedent, that groups can irreducibly engage in these epistemic activities. But he notes that one might resist the consequent, that groups can engage in relations of epistemic trust, on the grounds that trust – given its normative and affective dimension – seems the kind of attitude that only individuals can enjoy. But Kallestrup appeals to a variety of cases to argue that, because groups can enter into testimonial relations of trust over and above any such relations between their individual members, group trust is equally irreducible to trust between individuals. Kallestrup's chapter synthesizes two cutting-edge conversations: one on group epistemic attitudes, and another on the role of trust in testimonial knowledge.

### 3.3 *Epistemic Trust and Epistemic Responsibility*



So there is an apparent tension between trust and evidence (highlighted for example in the abovementioned contributions by Simpson and Keren); yet trust also has an important epistemic role to play (as all of the contributions have emphasized). These two observations make it relevant to ask how trust can be *epistemically responsible*. For epistemic responsibility includes, among other things, responsibility to one's evidence. Heidi Grasswick, in "Reconciling Epistemic Trust and Responsibility", aims to do just that. Grasswick sets up a broad-lens framework. She distinguishes two types of epistemic trust, trust in testimony (the kind discussed in the contributions so far), and trust in inquiry, which involves trusting others to research and safeguard important knowledge even if they do not wind up testifying it. And she focuses not mainly on responsibly trusting a testifier for a one-off belief, but rather on the responsible development of epistemic-trust relationships that extend over time. This framework enables her to pinpoint three layers of epistemic responsibility that make for "healthy" relationships of epistemic trust: responsibility toward the evidence we have at a given time, responsibility in developing and maintaining communities of epistemic trust, and responsibility in critically evaluating the epistemic norms in play in our communities, for example by listening to the epistemically marginalized who are not served well by them.

In "Proper Epistemic Trust as a Responsibilist Virtue", Benjamin McCraw also seeks to elucidate the relationship between epistemic trust and epistemic responsibility – here, by construing proper epistemic trust itself as a responsibilist epistemic virtue. He develops an account of epistemic trust and then, drawing on contemporary psychology and Linda Zagzebski's account of epistemic virtue, presents an account of what makes it proper. McCraw argues for the existence of this virtue on the basis of its theoretical fruits. First, it fills a theoretical space between the epistemic vices of suspiciousness and gullibility. Second, it fits neatly into two different types of virtue theory: proper epistemic trust contributes to the epistemic good life, and the paradigmatically rational or virtuous epistemic agent will be motivated to display it. An important feature of proper epistemic trust is its context-sensitivity. McCraw draws on this to argue, third, that his account successfully arbitrates a number of epistemological debates, including about low-grade knowledge, reductionism vs. anti-reductionism in the epistemology of testimony, and the rational response to peer disagreement. Far from conflicting with epistemic responsibility, epistemic trust – at least when proper – instantiates it.

Whereas the discussions thus far have centered mainly on epistemic trust in others, Alessandra Tanesini, in "Virtuous and Vicious Intellectual Self-Trust", turns her attention to epistemic *self*-trust. She construes this as a set of dispositions to rely on one's faculties (so, as a form of what I have called "mere reliance"), together with positive epistemic feelings and confidence in one's willpower. Tanesini distinguishes one epistemically beneficial form of epistemic self-trust, confident optimism, and three pathological ones, respectively arrogant, timid, and servile self-trust. Drawing extensively on psychological research, she argues that each form of epistemic self-trust arises from a corresponding form of self-esteem. Confident optimism has its roots in healthy self-esteem, which frees the agent to be motivated in her inquiry by the prospect of achieving epistemic aims. Arrogant, timid, and servile self-trust derive from warped forms of self-esteem that focus the agent not on epistemic goods but rather on her own vulnerability. Confident optimism about one's faculties, then, turns out epistemically virtuous, whereas the three pathological forms of self-trust turn out to be epistemically vicious. What Tanesini gives us is a psychologically robust account of epistemic self-trust within the framework of epistemic virtue responsibilism.

### 3.4 *The Vulnerabilities of Trust*

One running theme of the contributions thus far is that trust incurs vulnerability (Simpson, Fricker, Keren); another is that our relationships of trust are deeply embedded in social structures (Grasswick, Kallestrup). Katherine Dormandy's contribution explores some ramifications of both ideas for epistemic trust in others, Mari Mikkola's does so for epistemic self-trust.

In "Exploiting Epistemic Trust", Dormandy looks at vulnerabilities arising in relationships of epistemic trust in others, specifically in testimony. She starts with a discussion of how trust in general can be exploited; a key observation is that trust incurs vulnerabilities not just for the party doing the trusting, but also for the trustee (after all, trust can be burdensome), so either party can exploit the other. She applies these considerations to epistemic trust, specifically in testimonial relationships. There, we standardly think of a hearer trusting a speaker. But we miss an important aspect of this relationship unless we consider too that the speaker standardly trusts the hearer, for example for epistemic recognition. Given this mutual trust, and given that both trustees and trusters can exploit each other, we have four possibilities for exploitation in epistemic-trust relationships: a speaker exploiting a hearer (a) by accepting the hearer's trust or (b) by imposing her trust on him, and a hearer exploiting a speaker (c) by accepting the speaker's trust or (d) by imposing his trust on her. One result is that you do not need to *betray* someone to exploit her – you can exploit her just as easily by doing what she trusts you for. Through exploring these four forms of exploitation, Dormandy seeks to better understand interpersonal epistemic trust itself as well as its embeddedness in social power structures.

In "Self-Trust and Discriminatory Speech", Mari Mikkola also considers the interplay between social power and trust, focusing on epistemic self-trust. Discriminatory speech (as opposed to hate speech) is legally protected on the grounds that, even if false, it can at least be argued against. But Mikkola argues that, contrary to this common view, discriminatory speech *weakens* the ability of its targets to argue against it, because it damages their epistemic self-trust. Drawing on the recently burgeoning literature about the cognitive role of social stereotypes, Mikkola singles out three types of effect – implicit bias, stereotype threat, and striking generics – that operate at a less than fully conscious level and are exacerbated by discriminatory speech. She argues that, because of the operation of these mechanisms, such speech bypasses rational thought and directly inhibits targets' ability to trust themselves epistemically. Epistemic self-trust, then, is a fragile phenomenon that is vulnerable to the social structures and norms that yield our social stereotypes. Rather than ban discriminatory speech, however, Mikkola advocates education to counter the pernicious stereotypes themselves.

There is much more to say about trust in epistemology than the chapters of this volume can say. Each makes a significant contribution to the ongoing conversation; but I encourage readers not to simply take my word for it.

## References

- Adler, Jonathan. 2002. *Belief's Own Ethics*. Cambridge, MA: MIT Press.
- Alfano, Mark. 2012. "Expanding the Situationist Challenge to Responsibilist Virtue Epistemology." *Philosophical Quarterly* 62 (247): 223–49.
- Baier, Annette. 1986. "Trust and Antitrust." *Ethics* 96 (2): 231–60.
- . 1991. "Trust and Its Vulnerabilities, and Sustaining Trust." In *Tanner Lectures on*

- Human Values*, 13:107–74. Salt Lake City, UT: University of Utah Press.
- Baker, Judith. 1987. “Trust and Rationality.” *Pacific Philosophical Quarterly* 68: 1–13.
- Becker, Lawrence C. 1996. “Trust as Noncognitive Security about Motives.” *Ethics* 107 (1): 43–61.
- Burge, Tyler. 1993. “Content Preservation.” *The Philosophical Review* 102 (4): 457–88.
- Coady, C. A. J. 1992. *Testimony: A Philosophical Study*. Oxford: Clarendon.
- Code, Lorraine. 1987. *Epistemic Responsibility*. Hanover, NH: University Press of New England and Brown University Press.
- Cogley, Zac. 2012. “Trust and the Trickster Problem.” *Analytic Philosophy* 53 (1): 30–47.
- Cohen, Stewart. 1999. “Review: Lehrer on Coherence and Self-Trust.” *Philosophy and Phenomenological Research* 59 (4): 1043–48.
- Coleman, James. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.
- Constantin, Jan, and Thomas Grundmann. 2018. “Epistemic Authority: Preemption through Source Sensitive Defeat.” *Synthese*. Springer Netherlands.
- D’Cruz, Jason. 2019. “Humble Trust.” *Philosophical Studies*. Springer Netherlands.
- Darwall, Stephen. 2017. “Trust as a Second-Personal Attitude of the Heart.” In *The Philosophy of Trust*, edited by Paul Faulkner and Thomas Simpson, 35–50. Oxford: Oxford University Press.
- Dasgupta, Partha. 1988. “Trust as a Commodity.” In *Trust: Making and Breaking Cooperative Relations*, edited by Diego Gambetta, 49–72. Oxford: Basil Blackwell.
- Daukas, Nancy. 2006. “Epistemic Trust and Social Location.” *Episteme* 3 (1–2): 109–24.
- Domenicucci, Jacopo, and Richard Holton. 2017. “Trust as a Two-Place Relation.” In *The Philosophy of Trust*, edited by Paul Faulkner and Thomas Simpson, 149–60. Oxford: Oxford University Press.
- Dormandy, Katherine. 2018. “Epistemic Authority: Preemption or Proper Basing?” *Erkenntnis* 83 (4). Springer Netherlands: 773–91.
- Dotson, Kristie. 2011. “Tracking Epistemic Violence, Tracking Practices of Silencing.” *Hypatia* 26 (2): 236–57.
- Faulkner, Paul. 2007. “On Telling and Trusting.” *Mind* 116 (464): 875–902.
- . 2011. *Knowledge on Trust*. Oxford: Oxford University Press.
- . 2015. “The Attitude of Trust Is Basic.” *Analysis* 75 (3): 424–29.
- Foley, Richard. 2001. *Intellectual Trust in Oneself and Others*. Cambridge: Cambridge University Press.
- Fricker, Elizabeth. 1994. “Against Gullibility.” In *Knowing from Words*, edited by Bimal K. Matilal and A. Chakrabarti, 125–61. Dordrecht: Springer.
- . 1995. “Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony” 104 (414): 393–411.
- . 2006. “Second-Hand Knowledge.” *Philosophy and Phenomenological Research* LXXIII (3): 592–618.
- . 2016. “Doing (Better) What Comes Naturally: Zagzebski on Rationality and Epistemic Self-Trust.” *Episteme*, 1–16.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Frost-Arnold, Karen. 2013. “Moral Trust and Scientific Collaboration.” *Studies in History and Philosophy of Science Part A* 44 (3). Elsevier Ltd: 301–10.
- . 2014a. “Imposters, Tricksters, and Trustworthiness as an Epistemic Virtue.” *Hypatia* 29 (4): 790–807.
- . 2014b. “The Cognitive Attitude of Rational Trust.” *Synthese* 191 (9): 1957–74.
- Goldberg, Sanford. 2010. *Relying on Others*. Oxford: Oxford University Press.
- Goldman, Alvin. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.

- Govier, Trudy. 1993. "Self-Trust, Autonomy, and Self-Esteem." *Hypatia* 8 (1): 99–120.
- Grasswick, Heidi. 2018. "Understanding Epistemic Trust Injustices and Their Harms." *Royal Institute of Philosophy Supplement* 84: 69–91.
- Hardin, Russell. 1993. "The Street-Level Epistemology of Trust." *Politics and Society* 21 (4): 505–29.
- Hardwig, John. 1991. "The Role of Trust in Knowledge." *Journal of Philosophy* 88 (12): 693–708.
- Hawley, Katherine. 2014a. "Partiality and Prejudice in Trusting." *Synthese* 191: 2029–45.
- . 2014b. "Trust, Distrust and Commitment." *Noûs* 48 (1): 1–20.
- Hieronymi, Pamela. 2008. "The Reasons of Trust." *Australasian Journal of Philosophy* 86 (2): 213–36.
- Hinchman, Edward. 2005. "Telling as Inviting to Trust." *Philosophy and Phenomenological Research* 70 (3): 562–87.
- . 2012. "Can Trust Itself Ground a Reason to Believe the Trusted?" *Abstracta* 6 (Special Issue VI): 47–83.
- . 2017. "On the Risks of Resting Assured: An Assurance Theory of Trust." In *The Philosophy of Trust*, edited by Paul Faulkner and Thomas Simpson, 51–69. Oxford: Oxford University Press.
- Holton, Richard. 1994. "Deciding to Trust, Coming to Believe." *Australasian Journal of Philosophy* 72 (1): 63–76.
- Horsburgh, H. J. N. 1960. "The Ethics of Trust." *The Philosophical Quarterly* 10 (41): 343–54.
- James, William. 1921. *The Will to Believe, Human Immortality, and Other Essays in Popular Philosophy*. Longmans, Green, and Co.
- Jones, Karen. 1996. "Trust as an Affective Attitude." *Ethics* 107 (1): 4–25.
- . 2004. "Trust and Terror." In *Moral Psychology: Feminist Ethics and Social Theory*, edited by Peggy DesAutels and Margaret Urban Walker, 3–18. Lanham, MD: Rowman and Littlefield.
- . 2012a. "The Politics of Intellectual Self-Trust." *Social Epistemology* 26 (2): 237–51.
- . 2012b. "Trustworthiness." *Ethics* 123 (October): 61–85.
- . 2017. "'But I Was Counting on You!'" In *The Philosophy of Trust*, edited by Paul Faulkner and Thomas Simpson, 90–108. Oxford: Oxford University Press.
- Kappel, Klemens. 2014. "Believing on Trust." *Synthese* 191 (9): 2009–28.
- Kelly, Thomas. 2003. "Epistemic Rationality as Instrumental Rationality: A Critique." *Philosophy and Phenomenological Research* LXVI (3): 612–40.
- Keren, Arnon. 2007. "Epistemic Authority, Testimony and the Transmission of Knowledge." *Episteme*, 368–81.
- . 2014. "Trust and Belief: A Preemptive Reasons Account." *Synthese* 191 (12): 2593–2615.
- Lackey, Jennifer. 2006. "It Takes Two to Tango: Beyond Reductionism and Non-Reductionism in the Epistemology of Testimony." In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 160–91. Oxford: Oxford University Press.
- Lehrer, Keith. 1997. *Self-Trust: A Study of Reason, Knowledge, and Autonomy*. Oxford: Oxford University Press.
- Lipton, Peter. 1998. "The Epistemology of Testimony." *Studies in History and Philosophy of Science* 29: 1–31.
- Lyons, Jack. 1997. "Testimony, Induction and Folk Psychology." *Australasian Journal of Philosophy* 75: 163–78.
- McGeer, Victoria. 2008. "Trust, Hope and Empowerment." *Australasian Journal of Philosophy* 86 (2): 237–54.
- McLeod, Carolyn. 2000. "Our Attitude Towards the Motivation of Those We Trust."

- Southern Journal of Philosophy* 38 (3): 465–79.
- McMyler, Benjamin. 2011. *Testimony, Trust, and Authority*. Oxford: Oxford University Press.
- Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. Oxford: Oxford University Press.
- Montmarquet, James. 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman and Littlefield.
- Moran, Richard. 2005. “Getting Told and Being Believed.” *Philosophers’ Imprint* 5 (5): 1–29.
- Mullin, Amy. 2005. “Trust, Social Norms, and Motherhood.” *Journal of Social Philosophy* 36 (3): 316–30.
- Nickel, Philip J. 2007. “Trust and Obligation-Ascription.” *Ethical Theory and Moral Practice* 10 (3): 309–19.
- Pettit, Philip. 1995. “The Cunning of Trust.” *Philosophy and Public Affairs* 24 (3): 202–25.
- Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford: Oxford University Press.
- Ross, Angus. 1986. “Why Do We Believe What We Are Told?” *Ratio*.
- Ruokonen, Floora. 2013. “Trust, Trustworthiness, and Responsibility.” In *Trust: Analytic and Applied Perspectives*, edited by Pekka Makela and Cynthia Townley, 1–14. Amsterdam and New York: Rodopi.
- Sandis, Constantine. 2011. “Issue Introduction” 12 (1).
- Scanlon, T. M. 2003. “Promises and Contracts.” In *The Difficulty of Tolerance*, edited by T. M. Scanlon, 234–69. Cambridge: Cambridge University Press.
- Simpson, Thomas W. 2012. “What Is Trust?” *Pacific Philosophical Quarterly* 93 (4): 550–69.
- . 2013. “Trustworthiness and Moral Character.” *Ethical Theory and Moral Practice* 16 (3): 543–57.
- . 2018. “Trust, Belief, and the Second-Personal.” *Australasian Journal of Philosophy* 96 (3). Taylor & Francis: 447–59.
- Simpson, Thomas W. 2017. “Trust and Evidence.” In *The Philosophy of Trust*, edited by Paul Faulkner and Thomas Simpson, 177–94. Oxford: Oxford University Press.
- Stern, Robert. 2017. “‘Trust Is Basic’: Løgstrup on the Priority of Trust.” In *The Philosophy of Trust*, edited by Paul Faulkner and Thomas Simpson, 272–93. Oxford: Oxford University Press.
- Webb, Mark Owen. 1993. “Why I Know About as Much as You: A Reply to Hardwig.” *The Journal of Philosophy* 90 (5): 260–70.
- Zagzebski, Linda. 1996. *Virtues of the Mind*. Oxford: Oxford University Press.
- . 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford: Oxford University Press.