

# 1 *Trust and Trustworthiness*

## 2 **1. Introduction**

3 What is the relationship between trust and trustworthiness? The question is a fraught one, not  
4 least because philosophers of *trust* have tended to focus on three-place trust ( $S$  trusts  $X$  with  $\phi$ ),  
5 whereas philosophers of *trustworthiness* have been primarily concerned with two-place trustwor-  
6 thiness – viz., ( $S$  is trustworthy).<sup>1</sup>

7 This mismatch in focus presents challenges for those who want their accounts of trust and  
8 trustworthiness to be mutually illuminating.<sup>2</sup> And it also prompts deeper methodological ques-  
9 tions, such as whether we ought to be trying to understand trust in terms of trustworthiness (as  
10 some philosophers have<sup>3</sup>) or trustworthiness in terms of trust (as others have)<sup>4</sup>?

11 Though there is little consensus here, a widespread underlying assumption is that central phe-  
12 nomenon of interest on the trustee's side is *dispositional* (viz., trustworthiness) whilst the central  
13 phenomenon of interest on the trustor's side is *non-dispositional* (viz., trust). A byproduct of this  
14 assumption is that the evaluative norms of principal interest on the trustor's side regulate trusting  
15 attitudes and performances whereas those on the trustee's side regulate dispositions to respond  
16 to trust.

17 The aim here will be to highlight some unnoticed problems with this asymmetrical picture  
18 and to show that a symmetrical, 'achievement-first' picture has important advantages. The view  
19 I develop is guided by a structural analogy with practical reasoning. Just as practical reasoning is  
20 working as it should only when there is realisation (knowledge and action) of states (belief and  
21 intention) with reverse directions of fits (mind-to-world and world-to-mind), likewise, cooper-  
22 ation between trustor and trustee is functioning as it should only when there is an analogous  
23 kind of realisation on both sides of the cooperative exchange – viz., when the trustor 'matches'

---

<sup>1</sup>As Hardin (1996) notes, a further complication is that 'Many discussions of trust run trust and trustworthiness together, with claims about trust that might well apply to trustworthiness but that seem off the mark for trust' (1996, 28).

<sup>2</sup>For example, to control for the mismatch in focus, should we try to 'translate' three-place trust in to two-place trust in order to best connect trust with (two-place) trustworthiness? Or, would it be better to do things the other way around, to 'translate' two-place trustworthiness in to three-place trustworthiness in order to connect trustworthiness with (three-place) trust?

<sup>3</sup>See, e.g., O'Neill (2018, 293), Ashraf et al. (2006, 204), McLeod (2020, sec. 1),

<sup>4</sup>For example, according to Wright (2010), trustworthiness requires that the trustee 'acknowledges the value of the trust that is invested in them ... and to use[sic.] this to help rationally decide how to act' (2010, sec. 3.b.). Other accounts of trustworthiness in terms of trust are found in Williams (2000) and Potter (2002, 205).

1 her achievement in trusting (an achievement in *fitting reliance to reciprocity*) with the trustee's  
2 achievement in responding to trust (an achievement in *fitting reciprocity to reliance*). An upshot  
3 of viewing cooperation between trustor and trustee as exhibiting achievement-theoretic struc-  
4 ture is that we will be better positioned to subsume trustworthiness (and its cognates on the  
5 trustee's side), like trust, under a wider suite of evaluative norms that regulate attempts, disposi-  
6 tions, and achievements symmetrically on both sides of a cooperative exchange, with 'matching  
7 achievements' as the gold standard.

8 Here is the plan. §1 clarifies and criticises the kind of asymmetric picture that is embraced  
9 in the philosophy of trust and trustworthiness, which privileges performances (and norms regu-  
10 lating them) on the trustor's side and dispositions (and norms regulating them) on the trustee's  
11 side. §2 develops an analogy between practical reasoning and cooperation in order to motivate  
12 an alternative picture, on which trusting and trustworthiness are better understood as having  
13 achievement-theoretic structure with reverse directions of fit. §3 builds on this picture in order  
14 to defend symmetrical evaluative norms – norms of *success*, *competence*, and *aptness* – on both  
15 sides.

## 16 **2. Trust and Trustworthiness: doing versus being?**

17 The distinction between trust and trustworthiness is almost invariably described as distinction  
18 between doing something (i.e., trusting) and being a certain way (i.e., being trustworthy) on  
19 account of having a dispositional property.<sup>5</sup>

20 For example, according to a family of views defended by Annette Baier (1986), Karen Jones  
21 (2012), and Zac Cogley (2012), trustworthiness is to be identified with a disposition to fulfil  
22 commitments, in conditions under which one has those commitments, and *in virtue of* goodwill  
23 towards the trustor. For Diego Gambetta (1988), the trustworthy person needn't be disposed  
24 to fulfill the commitments they have out of good will; they simply must be disposed to fulfil  
25 their commitments, whatever they are, 'willingly.' More minimalistically, Christoph Kelp and  
26 Mona Simion (2021) identify trustworthiness with the disposition to fulfill one's commitments  
27 *simpliciter*, and not necessarily through any distinctive motivation or accompanying attitude.

---

<sup>5</sup>Possessing a dispositional property is not itself a matter of being in a mental state or behaving some way. Rather, dispositional properties 'provide the possibility of some further specific state or behaviour' (see, e.g., Mumford 2016) in certain conditions. Accordingly, the various 'accounts of trustworthiness' on the market are not aiming to give an account of any mental state, attitude, or behaviour; they are aiming to characterise accurately the dispositional property that they take trustworthy people but not others to possess.

1 More weakly, for Katherine Hawley (2019), the relevant disposition ‘trustworthiness’ refers to  
2 is best framed negatively – viz., as a disposition to *avoid unfulfilled* commitments. By contrast  
3 with all of these views, Nancy Potter (2002), insists that the relevant disposition lining up with  
4 trustworthiness should be understood as a full-fledged moral virtue – one that consists in being  
5 disposed to respond to trust in appropriate ways.<sup>6</sup>

6 Despite their differences, these accounts all retain the fundamental idea trustworthiness and  
7 trust stand to each other as a (mere) ‘being’ to a ‘doing.’<sup>7</sup> And thus, to the extent that these  
8 accounts are conducive to theorising about trustworthiness and trust in terms of each other, it  
9 will be as a being illuminates a doing, (or vice versa).

10 While there is no doubt that being trustworthy corresponds with possessing *some* disposition,  
11 so likewise does *being a competent trustor*, e.g., being one who trusts in ways that don’t too often  
12 lead to betrayed trust.<sup>8</sup> And by the same token, just as trusting is itself not a disposition but an  
13 activity or performance, so likewise is the trustee’s *manifestation of trustworthiness* when taking  
14 care of things as entrusted<sup>9</sup>, viz., when actually reciprocating the trust placed in her (as opposed  
15 to merely ‘being the sort of person’ who would take care of things as entrusted).

16 Is there any good reason that would justify the asymmetrical focus on the dispositional prop-  
17 erty of trustworthiness and not on the trustee’s performance of manifesting trustworthiness through  
18 reciprocity?

19 There would be if the disposition of the trustee (rather than any performance on the trustee’s  
20 part) features essentially in a plausible specification of what one aims at in trusting, and thus, in  
21 explaining when trust is successful.

22 Such a line of thought is implicit in what Carolyn McLeod (2020) takes to be a platitude  
23 about trust, which is that ‘Trust is an attitude that we have towards people whom we hope will  
24 *be trustworthy*, where trustworthiness is a property not an attitude’<sup>10</sup>. Variations of this idea are

---

<sup>6</sup>According to Potter (2002), trustworthy persons “[...]give signs and assurances of their trustworthiness” and  
“They take their epistemic responsibilities seriously” (2002, 174–175; Cf., for criticism, Jones 2012, 75–76) and  
Kelp and Simion (2021). Note, however, that Jones (2012) is more sympathetic to the idea of trustworthiness  
as a virtue in the special case of what she calls ‘rich’ rather than ‘basic’ trustworthiness. See, e.g., (2012, 79).

<sup>7</sup>Or, alternatively, with reference to Vendler/Kenny classes – as an occurrence (trust) to a state (trustworthiness).  
See, e.g., Verkuyl (1989).

<sup>8</sup>For a defence of this way of thinking about competent trust, see Carter (2020).

<sup>9</sup>The locution ‘as entrusted’ is meant to encompass views on which the trustee counts as taking care of things  
as entrusted only if doing so in a particular way, including, e.g., out of goodwill (Baier 1986; Jones 1996) or  
in conjunction with a belief that one is so committed (e.g., Hawley 2014). The present discussion – which is  
theoretically neutral on this point – is compatible with opting for either such kind of gloss.

<sup>10</sup>See McLeod (2020, sec. 1, my italics).

1 seen in Elizabeth Fricker’s (2018) claim that ‘[...] One is not really trusting unless one adopts  
2 an attitude of optimism to the proposition that the trustee *is trustworthy*’ (2018, 6). Likewise,  
3 as Russell Hardin (2002) puts it, ‘trusting someone in some context is simply to be explained  
4 as merely the expectation that the person will most likely *be trustworthy*’ (2002, 31). And per-  
5 haps most directly, proponents of doxastic accounts of trust (Hieronymi 2008; McMyler 2011)  
6 straightforwardly identify trust with a belief that the trustee *is trustworthy*.<sup>11</sup>

7 Of course, we seek out a trustworthy person when initially deciding *whether* to trust or forbear  
8 from trusting; in this respect, Onora O’Neill is right that ‘where we aim [...] to place and refuse  
9 trust intelligently *we must link trust to trustworthiness*’ (2018, 293). But when we actually trust  
10 someone, the relevance of the trustee’s simply ‘being a certain way’ independent of their actually  
11 performing in a way that manifests how we perhaps hope or believe they are (i.e., trustworthy) –  
12 is not clear at all.

13 When I trust you to pay back the loan, I rely on you to pay it back, making myself vulnerable  
14 to your betrayal.<sup>12</sup> Suppose you *do* then pay it back. Is my trust successful? Not necessarily, says  
15 the proponent of the idea that trustworthiness is of special interest in understanding trust. In  
16 trusting, I aim *not just* that you take care of things any way, but take care of things *as entrusted*,  
17 which (on this line of thought) involves your ‘being trustworthy.’

18 This is partly right. But it gets an important thing wrong. Just as my trust isn’t thereby success-  
19 ful if you merely take care of things *any* old way (e.g., by attempting to betray me, but in doing  
20 so accidentally pay back the loan – then only my *reliance* would be successful), likewise, my trust  
21 misses the mark if you simply *are* trustworthy but (perhaps due to bad luck) *don’t* pay back the  
22 loan. But crucially – even more – there is a sense in which my trust *still* misses the mark if you (i)  
23 pay back the loan; (ii) are trustworthy; but (iii) your paying back the loan doesn’t manifest your  
24 trustworthiness (e.g., perhaps despite being trustworthy you pay back the loan on this occasion  
25 under threat or through some kind of manipulation by a third party.<sup>13</sup>)

---

<sup>11</sup>For a related though less standard kind of doxastic account, see, e.g., Keren (2014; 2019).

<sup>12</sup>For various expressions of the idea that trust essentially involves subjecting oneself to risk of betrayal, see, along with Hardin (1992), e.g., Baier (1986, 244), McLeod (2020, sec. 1), Nickel and Vaesen (2012, 861–862), Carter (2020, 2301, 2318–2319), Carter and Simion (2020, sec. 1.a), Becker (1996, 45, 49), Dasgupta (1988, 67–68), Dormandy (2020, 241–242), Kirton (forthcoming), O’Neill (2017, 70–72), Potter (2020, 244), and Hinchman (2017). Cf., Pettit (1995, 208).

<sup>13</sup>Coercion isn’t essential to making this kind of point; for example, the trustworthy person might be such that her success (in taking care of things as entrusted) doesn’t manifest her trustworthiness not because she lacked the opportunity to do so (as would be the case if she were coerced) but rather due to the abnormal presence of luck in accounting for the success. The underlying idea here – one that has been defended variously by Sosa

1 This suggests a revisionist picture of the original assumption.<sup>14</sup> Trust aims not at the trustee  
2 merely *being* a certain way – or even at the trustee doing a certain thing while at the same time  
3 being a certain way – but at the trustee *achieving* a certain thing, viz., succeeding in taking care  
4 of things *through* their trustworthiness.

5 This insight offers us a new vantage point for revisiting the relationship between trust and  
6 trustworthiness, and to appreciate some important performance-theoretic symmetries between  
7 the two that the characteristic focus on trustworthiness as a mere disposition has so far obscured.

### 8 **3. Structural analogies with practical reasoning**

9 Whereas mere *reliance* is successful just in case the person relied on takes care of things *any way*,  
10 the success conditions for trust are more demanding. Where we've got to so far is that *trust*  
11 as opposed to mere reliance is successful just in case the trustee *manifests her trustworthiness* in  
12 successfully taking care of what the trustor relies on her to do. And this involves, on the part of  
13 the trustee, a kind of *success through trustworthiness* – viz., an *achievement* in trustworthiness.

14 Let's now take this working idea – that the trustor aims in trusting at the trustee's *achievement*  
15 in trustworthiness – even further. Just consider that when the trustor *herself* attains this aim  
16 (i.e., the aim that trustee's taking care of things manifests her trustworthiness) – this might on  
17 some occasions of cooperation be down to dumb luck; they might foolishly trust the one trust-  
18 worthy person in the village of tricksters, but this lone trustworthy person might then manifest  
19 her trustworthiness full well in taking care of things.<sup>15</sup>

20 Trust is *successful* here. *And* the trustee exhibits an achievement in fulfilling the trust placed in  
21 her through trustworthiness. But, in this situation, there is no symmetrical achievement (success  
22 that manifests a trusting competence) on the *trustor's* side, even though trust is successful. And

---

(2007), Greco (2010, Ch. 5), Pritchard (2012), and Zagzebski (1996) – is that a success doesn't manifest one's reliable disposition (construed as an ability, virtue, or competence) if that success is unusually due to luck. How to unpack 'unusually' (alternatively: abnormally) is a contested point, one that features centrally in discussions in virtue epistemology of achievement, luck, and credit. See, e.g., Turri et al. (2019, sec. 5, §7).

<sup>14</sup>That is, it suggests we revise the widely shared assumption in the philosophy of trust that the theoretical focus on *trustworthiness* qua disposition (as opposed to, e.g., focusing on performances of trustworthiness) is justified in light of the importance of this disposition to understanding trust.

<sup>15</sup>In this example, we are to suppose that the trustor is not manifesting any trusting competence here, but rather, simply and naively trusting and just happens to be lucky. The structure of the case is importantly different from a case where the trustee *does* manifest competence, in a normal environment, but could have easily trusted someone who was mistaken in that environment. In the latter case, the structure is analogous to that of a 'fake barn case' – and that is a case which, at least within performance-theoretic epistemology, there is no barrier to attributing the success to the ability and thus to attributing achievement. For discussion, see, e.g., Sosa (2007, Ch. 2), Littlejohn (2014), Carter (2016), Pritchard (2009), Jarvis (2013), and Kallestrup and Pritchard (2014).

1 in this respect, there is an important sense in which the cooperation itself still falls short; the  
2 cooperation does not match ‘achievement to achievement,’ but matches merely success (by the  
3 trustor) to achievement (by the trustee).

4 Of course, the symmetry can be regained if we simply shore up the performance on the trustor’s  
5 side. Suppose it is *not* simply through good fortune but *through the trustor’s competence* (to trust  
6 *successfully* reliably enough) that the trustor trusts successfully. In such a case, the relevant trust  
7 is not just successful but ‘apt’ in that the successful trust manifests the trustor’s competence to  
8 bring that success about reliably. This *apt* (and not merely successful) trust, an achievement of  
9 trusting, on the trustor’s side would then match the trustee’s achievement in trustworthiness.  
10 And now *cooperation* is functioning well in that the cooperation between the two *falls short on*  
11 *neither side* of the cooperative exchange.

12 An analogy is useful here between (i) the symmetrical picture just described of cooperation  
13 working well; and (ii) Tim Williamson’s (2017) view of practical reasoning working well. Ac-  
14 cording to Williamson, a practical reasoning system is working well when and only when one acts  
15 on what one knows.<sup>16</sup> One is in a position to act on what one knows only if one ‘realises’ two  
16 kinds of states, with *reverse directions of fit* (mind-to-world and world-to-mind). Accordingly, on  
17 Williamson’s picture, practical reasoning is not functioning as it should if there is a defect on ei-  
18 ther on the mind-to-world side (i.e., mere belief rather than knowledge) or on the world-to-mind  
19 side (i.e., mere intention rather than action).<sup>17</sup>

20 The working analogy so far is this: just as practical reasoning is functioning well only when we  
21 have symmetrical realisation (knowledge and action) of states (belief and intention) with reverse  
22 directions of fits, likewise, cooperation between trustor and trustee is functioning well only when  
23 we have an analogous kind of *symmetrical realisation on both sides of the cooperative exchange* –

---

<sup>16</sup>This idea, originating in Williamson (2002), is given a sustained defence in his (2017) with some further up-  
dates in (2021). Whereas Williamson encourages us to view the idea that practical reasoning’s working well is  
a matter of acting on knowledge in the service of a wider criticism of the centrality of belief-desire psychology  
as explanatorily central, the core normative idea that, in practical reasoning, one should act only on what one  
knows has received defences by (along with Williamson) Hawthorne and Stanley (2008), Stanley (2005), Fantl  
and McGrath (2002). For an overview, see Benton (2014, sec. 2.a).

<sup>17</sup>I say ‘intention’ here rather than ‘desire’ as standing in for botched action to reflect Williamson’s updated (2017)  
structural analogies. In *Knowledge and its Limits* (2002), practical reason’s working well was also understood in  
terms of acting on knowledge. This picture was meant to replace belief-desire psychology as the centre of intel-  
ligent life. However, the original (2002) version of the analogy maintained that belief stood to knowledge (on  
the mind-to-world side) as desire stood to action (on the world-to-mind side). The updated picture assimilates  
desire to belief – i.e., belief about what is good – e.g., (see, e.g., Lewis 1988; Price 1989) – and replaces ‘desire’  
with ‘intention’ in the analogy. Thus, the updated picture holds that belief stands to knowledge as intention to  
action. For critical discussion, see Miracchi and Carter (2021).

1 viz., when the trustor ‘matches’ her achievement in trusting with the trustee’s achievement in  
2 responding to trust.

3 This working analogy can be extended further, by considering how the trustor’s and trustee’s  
4 matching achievements, when cooperation is working well, are themselves (like knowledge and  
5 action) *realisations of attempts with reverse direction of fit*.<sup>18</sup> To a first approximation: whereas the  
6 trustor aims not just to rely, but to *fit her reliance to the trustee’s reciprocity*, the trustee (as such)  
7 aims to *fit her reciprocity to the trustor’s reliance*.<sup>19</sup>

8 When the trustor attempts, but fails, to fit her reliance to reciprocity, what is residual is a kind  
9 of botched trust. When the trustee attempts, but fails, to fit her reciprocity to reliance, what is  
10 residual is a kind of botched reciprocity. (Compare with Williamson’s suggestion that belief is  
11 a kind of ‘botched knowledge’ and mere intention ‘botched’ action).

12 On this wider picture, then, in any two-way cooperative system, trust stands to apt trust as  
13 reciprocity to apt reciprocity (reciprocity that succeeds through trustworthiness) in a way that  
14 is broadly analogous to how (in a practical reasoning system, for Williamson) belief stands to  
15 knowledge (*viz.*, *apt belief*<sup>20</sup>) as intention to action (*i.e.*, *apt intention*<sup>21</sup>). And, further, just as

---

<sup>18</sup>The language of ‘direction of fit’ is originally usually credited to Anscombe (1957), as a way of characterising a distinction between theoretical and practical intentional mental states. Theoretical mental states aim at representing things as they are (e.g., beliefs) and practical mental states aim at getting things done (*i.e.*, desires). Realisation (*i.e.*, success), for a cognitive (or theoretical) intentional mental state involves fitting mind-to-world; realisation for a practical mental state (e.g., desire, intention, etc.) involves fitting world-to-mind. A central ‘lesson’ direction-of-fit theorists (e.g., Smith 1994; Velleman 2000; cf. Frost 2014) have taken from Anscombe’s initial discussion is that intentional mental states are characterisable along the mind-to-world or world-to-mind faultline. However, the kinds of things to which direction of fit talk is applicable are not limited to intentional mental states. For example, according to Searle (1979), statements and predictions have a *word-to-world* direction of fit, whereas commands and promises have a *world-to-word* direction of fit. It’s worth noting further that the very thought that things other than mental states can have directions of fit is actually perfectly compatible with Anscombe’s initial idea, which is that what *makes* intentional states like beliefs and desires have the directions of fit is that they have normative realisation conditions; beliefs *aim* (constitutively, not intentionally) at a certain kind of realisation, the same for desires. A similar normative reading is due to Platts (1980; for discussion see Frost 2014, 449–450). What this suggests, then, is that – at least in so far as we follow progenitors of DOF theory such as Anscombe and Searle, there is no barrier to viewing attempts more generally (including, e.g., trust and its reciprocation) with constitutive aims as admitting of directions of fit in so far as they have specifiable normative realisation conditions.

<sup>19</sup>I say ‘reverse’ direction of fit for ease of presentation, given that ‘reliance-to-reciprocity’ and ‘reciprocity-to-reliance’ are ostensibly reverse directions of fit. That said, it might have been more precise to describe the kind of direction of fit here as lining up even better with what Searle (1979) calls ‘double direction of fit.’ The reason here is that – in the unique case of cooperation – the realisation of one entails the realisation of the other.

<sup>20</sup>The idea that knowledge is type-identical with apt belief has advantages in epistemology; see Sosa (2007; 2010), Greco (2010, Chs. 5-6) and Zagzebski (1996) for some notable defences of this position. Although I find this view plausible, the identification of knowledge with apt belief – while it fits snugly with the proposal developed here – isn’t essential to it. For some criticism of the identification of knowledge with apt belief, see, e.g., Lackey (2007), Pritchard (2007), Kelp (2013), Kornblith (2004), and Kallestrup and Pritchard (2014).

<sup>21</sup>For defences of the view that action is fruitfully understood as apt intention, see Sosa (2015, Ch. 1) and Miracchi

1 practical reasoning’s working well requires a match between not merely belief and intention but  
 2 between knowledge and action; cooperation working well requires a match between not *mere*  
 3 but *apt* trust and reciprocity.

4 The tables below represent these key analogies:

**Table 1:** Practical reasoning: realisations and attempts

<b>Practical reasoning</b>	<b>Fitting mind-to-world</b>	<b>Fitting world-to-mind</b>
Functioning well	knowledge (realisation)	action (realisation)
	belief (attempt)	intention (attempt)

**Table 2:** Cooperation: realisations and attempts

<b>Cooperation</b>	<b>Fitting reliance-to-reciprocity</b>	<b>Fitting reciprocity-to-reliance</b>
Functioning well	trustor’s apt trust (realisation)	trustee’s apt reciprocity (realisation)
	trust (attempt)	reciprocity (attempt)

5 One important feature of Williamson’s ‘knowledge-action’ centric picture of practical reason-  
 6 ing is that it is meant to contrast with a competing picture (see, e.g., Humberstone 1992) that  
 7 takes attempts – belief and desire – *rather than their realisations* as the core explanatory mental  
 8 attitudes at the centre of intelligent life.<sup>22</sup> Attempts at knowledge and attempts at action retain  
 9 a place in this picture, but it is their realisations, rather than the attempts themselves, that are of  
 10 comparative theoretical importance.

11 The picture of cooperation suggested here likewise gives primacy to realisations over their at-  
 12 tempts. That is, the present picture rejects the the trustee’s performance (trust), a mere attempt at  
 13 realisation by fitting reliance to reciprocity, and the trustee’s disposition (trustworthiness) to fit

and Carter (2021).

<sup>22</sup>On the kind of view embraced by Humberstone (1992), it is also possible to accept the structural analogy on which belief stands to knowledge as desire to action. However, such a structural analogy would (on the belief-desire centred picture) begin with belief and desire as ‘direction of fit mirror images,’ from which we would then ‘solve upward’ in the analogy to get the result that belief stands to knowledge as desire to action. Resisting this picture is the central argument in Williamson (2017), who suggests we begin with knowledge and action as direction of fit mirror images and then solve ‘downward,’ filling in the relevant attempts. For a criticism of the role of ‘mirrors’ in both Williamson and Humberstone’s approaches, see Miracchi and Carter (2021).

1 reciprocity to reliance are the most theoretically important notions in a wider picture of cooper-  
2 ation. Rather, we should think of the importance of the trustor and trustee's matching achieve-  
3 ments of trust and trustworthiness in cooperation as broadly analogous to the importance of  
4 action and knowledge (as opposed to mere belief and desire) in practical reasoning.

#### 5 **4. Symmetric evaluative normativity: trustor and trustee**

6 In the good case where cooperation is working as it should, the trustor matches her achievement  
7 in trusting with the trustee's achievement in responding to trust. In both of these achievements  
8 (of apt trust and apt reciprocity) *competence* is manifested in success.

9 Cooperation doesn't always go so well. It falls short – at least to some extent – if we have  
10 anything short of achievement on either the trustor or trustee's side. In some cases, cooper-  
11 ation doesn't fall short by much, as when the trustor matches successful and competent but inapt  
12 (i.e., Gettiered) success to the trustee's achievement.<sup>23</sup> The trustor could do far worse. Success-  
13 ful but incompetent trust falls short of Gettiered trust on the trustor's side, as does competent  
14 but unsuccessful trust.<sup>24</sup> On the bottom rung on the trustor's side, we have trust that is neither  
15 competent nor successful, e.g., the betrayal of the gullible.

16 Likewise, on the trustee's side, falling just short of achievement is a kind of *Gettiered reciprocity*;  
17 suppose the trustee manifests her trustworthiness in assiduously entering the correct bank details  
18 online to pay back the loan she was entrusted to pay back, but succeeds only because a fortuitous  
19 electronic glitch (good luck) *corrects for an initial glitch* (bad luck) that would have diverted the  
20 funds to the wrong account.<sup>25</sup>

21 The trustee could do far worse. For one thing, she could have *not* manifested trustworthiness  
22 in responding to the trust placed in her, but succeeded just by luck. In such a case, suppose she  
23 *intends* to wire the money to the wrong account but only accidentally wires it to the right one.

24 Whereas the first loan case is a case of Gettiered reciprocity, the second is successful but in-  
25 competent reciprocity. Two remaining categories, lower down the rung on the trustee's side are:

---

<sup>23</sup>Performances that are successful and competent but inapt have a 'Gettier' structure, where the success is disconnected from the good method used. For discussion, see Sosa (2007, Ch. 2; 2010, 467, 474–475) and Greco (2009, 19–21; 2010, 73–76, 94–99). Cf., Pritchard (2012, 251, 264–268).

<sup>24</sup>The performance-theoretic analogy with virtue epistemology holds that successful but incompetent trust and competent but unsuccessful trust fall short of apt trust in a way that is analogous to how unjustified true beliefs and justified false beliefs both fall short of knowledge. See, for discussion, Sosa (2007; 2010; Sosa 2015).

<sup>25</sup>For discussion of this kind of 'double luck' structure in relation to Gettier cases, see, e.g., Zagzebski (1994); see also Pritchard (2007) on what he calls 'intervening' veritic luck.

1 unsuccessful and competent reciprocity (i.e., exactly like the Gettiered reciprocity case *without*  
 2 the second stroke of good luck), and – at the very bottom rung – incompetent and unsuccess-  
 3 ful reciprocity (e.g., the trustee intends to wire money to the wrong account, and – failing in  
 4 reciprocity – succeeds in betrayal.)

5 The above picture shows not only the many ways that cooperation can be defective (by less or  
 6 greater degree) by matching anywhere from *just* less to *much* less than achievement on either the  
 7 trustor’s or trustee’s side. But it also reveals an important *normative symmetry* on both sides.

8 By ‘normative symmetry’ what I mean is that the relevant *attempts* on each side (fitting reliance  
 9 to reciprocity on the trustor’s side, fitting reciprocity to reliance on the trustee’s side) are such  
 10 that we can evaluate each for the very same three things: (i) *success*; (ii) *competence*; and (iii)  
 11 *aptness*. And, moreover, it is specifically by *failing* to satisfy combinations of these norms that  
 12 performances on the trustor and trustee’s side fall short of achievement to whatever extent that  
 13 they do.

14 The symmetrical picture of evaluative norms on each side is accordingly as follows:

	On the trustor’s side	On the trustee’s side
Direction of fit attempt	Reliance-to-reciprocity (trust) by means of reliance	Reciprocity-to-reliance reciprocity (by means of responding to trust)
success norm	<i>S</i> ’s trusting <i>X</i> with $\phi$ is better if successfully reciprocated than if not; <i>S</i> ’s trusting <i>X</i> with $\phi$ is successfully reciprocated iff takes care of $\phi$ as entrusted.	<i>X</i> ’s reciprocating <i>S</i> ’s trust with $\phi$ is better if <i>X</i> successfully reciprocates <i>S</i> ’s trust with $\phi$ than if not; <i>X</i> successfully reciprocates <i>S</i> ’s trust with $\phi$ iff <i>X</i> takes care of $\phi$ as entrusted.

	On the truster's side	On the trustee's side
competence norm	$S$ 's trusting $X$ with $\phi$ is better if $S$ trusts $X$ with $\phi$ competently than if $S$ does not.	$X$ 's reciprocating $S$ 's trusting $X$ with $\phi$ is better if $X$ reciprocates $S$ 's trust with $\phi$ competently than if $X$ does not.
aptness norm	$S$ 's trusting $X$ with $\phi$ is better if $S$ trusts $X$ with $\phi$ aptly than if $S$ does not.	$X$ 's reciprocating $S$ 's trusting $X$ with $\phi$ is better if $X$ reciprocates $S$ 's trust with $\phi$ aptly than if $X$ does not.

1 This symmetrical picture offers us a number of advantages. For one thing, our guiding idea  
2 that cooperation between trustor and trustee is working as it should when both sides match  
3 achievement to achievement can now be restated as an *aptness norm on cooperation*, one that  
4 is formulated *in terms of* trustor and trustee satisfying respective evaluative norms of aptness: a  
5 cooperative trust exchange  $E$  between trustor and trustee is better than it would be otherwise if  
6  $E$  is apt;  $E$  is apt iff trustor and trustee satisfy their respective aptness norms.<sup>26</sup>

7 Secondly – and this bring us back to where we started – it should now be even more evident  
8 why focusing principally on a disposition (trustworthiness) on the trustee's side but not on the  
9 trustor's side (and vice versa for performance) is going to be arbitrary. From a wider view that  
10 takes in and evaluates the trust exchange in full, neither has any special status, even though both  
11 are essential to cooperation going well. They are, in a bit more detail, essential to cooperation  
12 going well in a way that is roughly analogous to how our beliefs and intentions (or: dispositions  
13 to form intentions) are important to practical reasoning going well. Both deserve attention, but  
14 should be appreciated as attempts *at* realisations, where the realisations of those attempts are

<sup>26</sup>The idea that cooperation itself admits of an aptness norm suggests that cooperation is a kind of multi-agent performance itself. A natural way of thinking of this is as an irreducibly collective property of cooperators engaged jointly in a trust exchange. While I am sympathetic to this kind of gloss, I want to stress that we needn't be committed to it. The crux of the idea – viz., that cooperation is apt iff its individual cooperators perform aptly – is also compatible with a 'summativist' gloss, on which the cooperation has the relevant property (i.e., aptness) iff all its individual members have that property. For relevant recent discussion of these points, see Lackey (2021) and Broncano-Berrocal and Carter (2021). For a discussion of aptness as applicable to groups, see Kallestrup (2016).

1 what's needed in good practical reasoning as well as (*mutatis mutandis*) in good cooperation.

2 Thirdly, by transitioning to a symmetrical picture of the evaluative normativity of trust – with  
3 achievement matching achievement as the gold standard – we are better positioned to see the  
4 importance of questions that have been so far obscured. Perhaps most conspicuously here are  
5 questions about the competence norm of trust. After all, we have a grip on *apt trust* only by un-  
6 derstanding competent trust, and this involves a clear view of those dispositions of the trustor  
7 that lead them to trust *successfully* reliably enough. Other questions invited by the symmetri-  
8 cal picture involve the evaluative normativity of cooperation generally. Even if ‘aptness on both  
9 sides’ of the trustor/trustee divide implies that the cooperative exchange itself is apt, it remains an  
10 open question how to evaluate certain cooperation permutations that involve at least one norm  
11 violation on one side. For example: is cooperation working better if the trustor matches success  
12 without competence to the trustee’s achievement or competence without success to the trustee’s  
13 achievement?

14 Fourthly, given that *distrust* no less than trust can be successful, competent and apt, the nor-  
15 mative symmetry we find on the trustor and trustee’s side invites us to consider – analogically –  
16 what stands to distrust as as trust to reciprocity, and to consider how to best characterise paral-  
17 lel symmetrical norms that would regulate – symmetrically with successful, competent and apt  
18 distrust (on the side of the trustor) – also forbearance on the side of the trustee.

## 19 **5. Concluding Remarks**

20 The aim here has been to motivate and defend a new way of theorising about trust and trustwor-  
21 thiness – and their relationship to each other – by locating both within a broader picture that  
22 captures largely overlooked symmetries on both the trustor’s and trustee’s side of a cooperative  
23 exchange. The view I’ve defended here takes good cooperation as a theoretical starting point;  
24 on the view proposed, cooperation between trustor and trustee is working well when achieve-  
25 ments in trust and responding to trust are matched on both sides of the trust exchange. In a bit  
26 more detail, the trustor ‘matches’ her achievement in trusting (an achievement in fitting reliance  
27 to reciprocity) with the trustee’s achievement in responding to trust (an achievement in fitting  
28 reciprocity to reliance). From this starting point, we can then appreciate *symmetrical* ways that  
29 the trustor and trustee can (respectively) fall short, by violating what I’ve shown are symmetri-  
30 cal evaluative norms – of success, competence and aptness – that regulate the attempts made by

1 both trustor and trustee. The overall picture was shown to have important advantages over the  
2 received way of theorising about how trust stands to trustworthiness, and it clears the way – by  
3 identifying key questions that have been obscured – to making further progress.

#### 4 **References**

- 5 Anscombe, G. 1957. *Intention*. Harvard University Press.
- 6 Ashraf, Nava, Iris Bohnet, and Nikita Piankov. 2006. “Decomposing Trust and Trustworthiness.”  
7 *Experimental Economics* 9 (3). Springer: 193–208.
- 8 Baier, Annette. 1986. “Trust and Antitrust.” *Ethics* 96 (2): 231–260.
- 9 Becker, Lawrence C. 1996. “Trust as Noncognitive Security about Motives.” *Ethics* 107 (1). Univer-  
10 sity of Chicago Press: 43–61.
- 11 Benton, Matthew A. 2014. “Knowledge Norms.” *Internet Encyclopedia of Philosophy*.
- 12 Broncano-Berrocal, Fernando, and J. Adam Carter. 2021. *The Philosophy of Group Polarization: Epis-*  
13 *temology, Metaphysics, Psychology*. Routledge.
- 14 Carter, J. Adam. 2016. “Robust Virtue Epistemology as Anti-Luck Epistemology: A New Solution.”  
15 *Pacific Philosophical Quarterly* 97 (1). Wiley Online Library: 140–155.
- 16 Carter, J. Adam. 2020. “On Behalf of a Bi-Level Account of Trust.” *Philosophical Studies* 177 (8):  
17 2299–2322. doi:10.1007/s11098-019-01311-2.
- 18 Carter, J. Adam, and Mona Simion. 2020. “The Ethics and Epistemology of Trust.” *Internet Encyclo-*  
19 *pedia of Philosophy*. Internet Encyclopedia of Philosophy.
- 20 Cogley, Zac. 2012. “Trust and the Trickster Problem.”
- 21 Dasgupta, Partha. 1988. “Trust as a Commodity.” In *Trust: Making and Breaking Cooperative Rela-*  
22 *tions*, edited by Diego Gambetta, 49–72. Blackwell.
- 23 Dormandy, Katherine. 2020. “Exploitative Epistemic Trust.” In *Trust in Epistemology*, edited by  
24 Katherine Dormandy, 241–264.
- 25 Fantl, Jeremy, and Matthew McGrath. 2002. “Evidence, Pragmatics, and Justification.” *The Philosoph-*  
26 *ical Review* 111 (1): 67–94.
- 27 Fricker, Elizabeth. 2018. “Trust and Testimonial Justification.” *Manuscript*.
- 28 Frost, Kim. 2014. “On the Very Idea of Direction of Fit.” *The Philosophical Review* 123 (4): 429–484.  
29 doi:10.1215/00318108-2749720.
- 30 Gambetta, Diego. 1988. *Trust: Making and Breaking Cooperative Relations*. Blackwell.

- 1 Greco, John. 2009. "Knowledge and Success from Ability." *Philosophical Studies* 142 (1). Springer:  
2 17–26.
- 3 Greco, John. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cam-  
4 bridge University Press.
- 5 Hardin, Russell. 1992. "The Street-Level Epistemology of Trust." *Analyse & Kritik* 14 (2). De Gruyter  
6 Oldenbourg: 152–176.
- 7 Hardin, Russell. 1996. "Trustworthiness." *Ethics* 107 (1). University of Chicago Press: 26–42.
- 8 Hardin, Russell. 2002. *Trust and Trustworthiness*. Russell Sage Foundation.
- 9 Hawley, Katherine. 2014. "Trust, Distrust and Commitment." *Noûs* 48 (1). JSTOR: 1–20.
- 10 Hawley, Katherine. 2019. *How to Be Trustworthy*. Oxford University Press, USA.
- 11 Hawthorne, John, and Jason Stanley. 2008. "Knowledge and Action." *The Journal of Philosophy* 105  
12 (10). JSTOR: 571–590.
- 13 Hieronymi, Pamela. 2008. "The Reasons of Trust." *Australasian Journal of Philosophy* 86 (2). Taylor  
14 & Francis: 213–236.
- 15 Hinchman, Edward. 2017. "On the Risks of Resting Assured: An Assurance Theory of Trust." In *The*  
16 *Philosophy of Trust*, edited by Paul Faulkner and Thomas W. Simpson. Oxford: Oxford University  
17 Press.
- 18 Humberstone, I. Lloyd. 1992. "Direction of Fit." *Mind* 101 (401). JSTOR: 59–83.
- 19 Jarvis, Benjamin. 2013. "Knowledge, Cognitive Achievement, and Environmental Luck." *Pacific*  
20 *Philosophical Quarterly* 94 (4). Wiley Online Library: 529–551.
- 21 Jones, Karen. 1996. "Trust as an Affective Attitude." *Ethics* 107 (1). University of Chicago Press:  
22 4–25.
- 23 Jones, Karen. 2012. "Trustworthiness." *Ethics* 123 (1). University of Chicago Press Chicago, IL:  
24 61–85.
- 25 Kallestrup, Jesper. 2016. "Group Virtue Epistemology." *Synthese*. Springer, 1–19.
- 26 Kallestrup, Jesper, and Duncan Pritchard. 2014. "Virtue Epistemology and Epistemic Twin Earth."  
27 *European Journal of Philosophy* 22 (3). Wiley Online Library: 335–357.
- 28 Kelp, Christoph. 2013. "Knowledge: The Safe-Apt View." *Australasian Journal of Philosophy* 91 (2).  
29 Taylor & Francis Group: 265–278.
- 30 Kelp, Christoph, and Mona Simion. 2021. "What Is Trustworthiness?" *Manuscript*.
- 31 Keren, Arnon. 2014. "Trust and Belief: A Preemptive Reasons Account." *Synthese* 191 (12). Springer:  
32 2593–2615.
- 33 Keren, Arnon. 2019. "Trust, Preemption, and Knowledge." *Trust in Epistemology*. Routledge.

- 1 Kirton, Andrew. forthcoming. "Matters of Trust as Matters of Attachment Security." *International*  
2 *Journal of Philosophical Studies*, forthcoming, 1–20. doi:10.1080/09672559.2020.1802971.
- 3 Kornblith, Hilary. 2004. "Sosa on Human and Animal Knowledge." *Ernest Sosa and His Critics*.  
4 Wiley Online Library, 126–134.
- 5 Lackey, Jennifer. 2007. "Why We Don't Deserve Credit for Everything We Know." *Synthese* 158 (3).  
6 Springer: 345–361.
- 7 Lackey, Jennifer. 2021. *The Epistemology of Groups*. Oxford University Press, USA.
- 8 Lewis, David. 1988. "Desire as Belief." *Mind* 97 (387). JSTOR: 323–332.
- 9 Littlejohn, Clayton. 2014. "Fake Barns and False Dilemmas."
- 10 McLeod, Carolyn. 2020. "Trust." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N.  
11 Zalta, Fall 2020. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2020/entries/rust/>.
- 12
- 13 McMyler, Benjamin. 2011. *Testimony, Trust, and Authority*. OUP USA.
- 14 Miracchi, Lisa, and J. Adam Carter. 2021. "Refitting the Mirrors: On Structural Analogies in Episte-  
15 mology and Action Theory." *Manuscript*.
- 16 Mumford, Stephen. 2016. "Dispositions." In *Routledge Encyclopedia of Philosophy*, 1st ed. London:  
17 Routledge. doi:10.4324/9780415249126-N116-1.
- 18 Nickel, Philip J., and Krist Vaesen. 2012. "Risk and Trust." In *Handbook of Risk Theory*, edited by  
19 Sabine Roeser, Rafaela Hillerbrand, Martin Peterson, and Per Sandin. Springer.
- 20 O'Neil, Collin. 2017. "Betraying Trust." In *The Philosophy of Trust*, edited by Paul Faulkner and  
21 Thomas W. Simpson, 70–89. Oxford, UK: Oxford University Press.
- 22 O'Neill, Onora. 2018. "Linking Trust to Trustworthiness." *International Journal of Philosophical*  
23 *Studies* 26 (2). Routledge: 293–300. doi:10.1080/09672559.2018.1454637.
- 24 Pettit, Philip. 1995. "The Cunning of Trust." *Philosophy and Public Affairs* 24 (3). Wiley-Blackwell:  
25 202–225. doi:10.1111/j.1088-4963.1995.tb00029.x.
- 26 Platts, Mark. 1980. "Ways of Meaning." *Mind* 89 (355). Oxford University Press: 454–456.
- 27 Potter, Nancy Nyquist. 2002. *How Can I Be Trusted?: A Virtue Theory of Trustworthiness*. Rowman  
28 & Littlefield.
- 29 Potter, Nancy Nyquist. 2020. "Interpersonal Trust." In *The Routledge Handbook of Trust and Philos-*  
30 *ophy*, edited by Judith Simon, 243–255. Routledge.
- 31 Price, Huw. 1989. "Defending Desire-as-Belief." *Mind* 98 (389). [Oxford University Press, Mind  
32 Association]: 119–127. <https://www.jstor.org/stable/2255064>.
- 33 Pritchard, Duncan. 2007. "Anti-Luck Epistemology." *Synthese* 158 (3). Springer: 277–297.

- 1 Pritchard, Duncan. 2009. "Knowledge, Understanding and Epistemic Value." *Royal Institute of Philos-*  
2 *ophy Supplement* 64. Cambridge: Cambridge University Press: 19–43. doi:10.1017/s1358246109000046.
- 3 Pritchard, Duncan. 2012. "Anti-Luck Virtue Epistemology." *The Journal of Philosophy* 109 (3): 247–  
4 279.
- 5 Searle, J. R. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts* Cambridge University  
6 Press. Cambridge.
- 7 Smith, Michael. 1994. *The Moral Problem*. Blackwell.
- 8 Sosa, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford  
9 University Press.
- 10 Sosa, Ernest. 2010. "How Competence Matters in Epistemology." *Philosophical Perspectives* 24 (1):  
11 465–475. doi:10.1111/j.1520-8583.2010.00200.x.
- 12 Sosa, Ernest. 2015. *Judgment & Agency*. Oxford University Press UK.
- 13 Stanley, Jason. 2005. *Knowledge and Practical Interests*. Oxford University Press.
- 14 Turri, John, Mark Alfano, and John Greco. 2019. "Virtue Epistemology." In *The Stanford Encyclope-*  
15 *dia of Philosophy*, edited by Edward N. Zalta, Fall 2019. Metaphysics Research Lab, Stanford Uni-  
16 versity. [https://plato.stanford.edu/archives/fall2019/entries/epistemology-](https://plato.stanford.edu/archives/fall2019/entries/epistemology-virtue/)  
17 [virtue/](https://plato.stanford.edu/archives/fall2019/entries/epistemology-virtue/).
- 18 Velleman, David. 2000. *The Possibility of Practical Reason*. Oxford University Press.
- 19 Verkuyl, Henk J. 1989. "Aspectual Classes and Aspectual Composition." *Linguistics and Philosophy*  
20 12 (1). Springer: 39–94.
- 21 Williams, Bernard. 2000. "Formal Structures and Social Reality." *Trust: Making and Breaking Coop-*  
22 *erative Relations* 1. Citescer: 3–13.
- 23 Williamson, Timothy. 2002. *Knowledge and Its Limits*. Oxford University Press on Demand.
- 24 Williamson, Timothy. 2017. "Acting on Knowledge." In *Knowledge First: Approaches in Epistemol-*  
25 *ogy and Mind*, edited by J. Adam Carter, Emma C. Gordon, and Benjamin W. Jarvis, 163–181.  
26 Oxford: Oxford University Press.
- 27 Williamson, Timothy. 2021. "Epistemological Ambivalence." In *Epistemic Dilemmas*, edited by Nick  
28 Hughes. Oxford: Oxford University Press.
- 29 Wright, Stephen. 2010. "Trust and Trustworthiness." *Philosophia* 38 (3). Springer Netherlands: 615–  
30 627. doi:10.1007/s11406-009-9218-0.
- 31 Zagzebski, Linda. 1994. "The Inescapability of Gettier Problems." *The Philosophical Quarterly* (1950-  
32 ) 44 (174). JSTOR: 65–73.

- 1 Zagzebski, Linda T. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical*  
2 *Foundations of Knowledge*. Cambridge University Press.