Philos Stud (2020) 177:2299–2322
https://doi.org/10.1007/s11098-019-01311-2

On behalf of a bi-level account of trust

J. Adam Carter[1]

Published online: 17 June 2019

Abstract A bi-level account of trust is developed and defended, one with relevance in ethics as well as epistemology. The proposed account of trust—on which trusting is modelled within a virtue-theoretic framework as a *performance-type with an aim*—distinguishes between two distinct levels of trust, *apt* and *convictive*, that take us beyond previous assessments of its nature, value, and relationship to risk assessment. While Sosa (A virtue epistemology: apt belief and reflective knowledge, volume I, Oxford University Press, Oxford, 2009; Judgment and agency, Oxford University Press, Oxford, 2015; Epistemology, University Press, Princeton, 2017), in particular, has shown how a performance normativity model may be fruitfully applied to *belief,* my objective is to apply this kind of model in a novel and principled way to *trust*. I conclude by outlining some of the key advantages of the performance-theoretic bi-level account of trust defended over more traditional univocal proposals.

Keywords Virtue epistemology · Sosa · Trust · Competence · Virtue reliabilism

1.

A helpful—and thus far, unexplored—way to think about trust will be to draw from the resources of virtue epistemology, and in particular, from the bi-level virtue epistemology framework developed by Ernest Sosa (2009, 2015, 2017).

Sosa's distinction between animal knowledge and reflective knowledge (a distinction around which his epistemological project is centred) is

✉ J. Adam Carter
adam.carter@glasgow.ac.uk

[1] Philosophy, University of Glasgow, 67-69 Oakfield Avenue, Glasgow G12 8QQ, UK

controversial.[1] Nonetheless, his sophisticated framework for modelling the normative structure of performances with aims will be useful—regardless of what we say about *knowledge*—for theorising about two very distinctive *levels* of trust, levels that take us beyond previous assessments of its nature, value and relationship to risk assessment.

The bi-level account of trust I aim to propose is one in which the notion of *competence* plays a starring role. One may initially wonder why competence should play *any* role in an account of trust (as opposed to *trustworthiness*[2])? Competences, after all, are dispositions, but trust is paradigmatically an *attitude* or a hybrid of attitudes (e.g., optimism, hope, belief, etc.), one that involves actual, and not merely dispositional, vulnerability to betrayal.

In what follows, I'll focus (i) first on a simple auxiliary role for competence to play in an account of trust, and then (ii) on some more sophisticated constitutive roles. What results will be a novel performance-theoretic account of the structure of trust, one that will have applications in both ethics and epistemology as well as advantages over more traditional accounts in both areas.

2.

Beliefs *aim* at truth.[3] Sometimes, beliefs hit that aim through dumb luck. Such beliefs, while valuable to the extent that truth is valuable, are regarded as less valuable, less epistemologically praiseworthy, as *knowledge*—viz., what you get when belief hits its aim (truth) through cognitive skill. This is an idea that lies at the heart of contemporary virtue epistemology.

Does trust *aim* at anything? If it does, then a tempting line of thought goes as follows: when trust hits its aim—whatever that is—that is good to the extent that trust's aim is a worthy aim.[4] But, when trust hits its aim skilfully, that is something even better.

[1] See, for example, Kornblith (2009, 2010, 2012) for some notable criticisms of Sosa's animal/reflective knowledge distinction. Cf., Perrine (2014) and Carter and McKenna (2018).

[2] For a view of trustworthiness as a kind of disposition, see Potter (2002) and Hardin (1996). For criticism, see Jones (2012).

[3] This idea is, at any rate, commonplace in epistemology—viz., the idea that truth is the standard of correctness for belief; a belief is correct if and only if true. A stronger, but more controversial, way of articulating this idea is embraced by normativists about belief, e.g., Shah and Velleman (2005) and Shah (2003). According to this view, belief *constitutively* aims at truth; what distinguishes an attitude as a belief is that it is governed by the truth norm in this way. Note that the ground-level idea that belief aims at truth needn't commit one to the stronger view held by normativists. For representative work on the topic, see Chan (2013). Note, finally, that some epistemologists maintain that belief aims not at truth, but at *knowledge*; this is the position embraced by proponents of the knowledge-first program e.g., Williamson (2000). I am for the present purposes setting this point aside.

[4] Some forms of trust may be morally bad. As Baier (1986, 232) notes, 'There are immoral as well as moral trust relationships, and trustbusting can be a morally proper goal'. Likewise, Kvanvig (2008) notes that some truths can be morally bad to possess, or for that matter pointless. The proposal developed here (like an account of knowledge, in epistemology) is compatible with this much; it may be that, on some occasions, moral and other values render a particular case of trusting all-things-considered disvaluable, just as knowledge may on occasion be all-things-considered disvaluable.

Plausibly, trust *does* have an aim, much like belief does. Even among those who disagree about the nature of trust, it is generally agreed that trust is an attitude we have towards people whom we hope will (in some suitably specified way) *take care of things as we have entrusted them*,[5] where the entrusting itself involves incurring some non-negligible level of risk.[6]

There is much discussion about what it would be for the trustee to take care of things *as* entrusted—viz., what additional attitudes and/or beliefs[7] this may involve, including attitudes and beliefs *about* the attitudes (e.g., goodwill) and beliefs of the trustee.[8] However these details are filled out, it remains that when one trusts another to take care of things as entrusted, there is a clear sense in which the trust *succeeds* only when the trustee *actually does* take care of things as they are entrusted (perhaps, compresently, with certain attitudes or beliefs).

I want to accordingly take as a starting point a distinction between (i) trust as a *success-apt attitude* and (ii) *successful trust*, as follows:

> *Trust attitude* (*Trust-A*): an attitude we have towards people whom we hope will take care of things as we have entrusted them,[9] and which involves incurring some non-negligible level of risk.
>
> *Successful trust*: when one's Trust-A in another is fulfilled—viz., when another has taken care of things as we have entrusted them.

3.

Just as one can fail to be trustworthy, one can fail to trust well, by trusting (and incurring some risk of betrayal) in ways that don't ordinary lead to successful trust.

[5] See, for example, Baier (1986, 235) and Jones (1996). Note that the term "trust" also has an impersonal use; for example, one might trust the rope they are climbing or their car's breaks. The kind of trust of interest in this paper is interpersonal trust. For discussion of the distinction between trust and mere reliance, see §11(e).

[6] For a helpful overview of this common ground, as well as points of contentions, see Simon (2013) and McLeod (2015).

[7] According to cognitive accounts of trust, trusting is a species of belief. See, for example, Gambetta (1988) and Coleman (1990). Doxastic accounts maintain similarly that trust entails or involves a belief to the effect that the trustee will be trustworthy.

[8] This may also involve entrusting them to take care of things out of goodwill e.g., Baier (1986, 234). On Baier's view, 'Where one depends on another's good will, one is necessarily vulnerable to the limits of that good will', and accordingly, to risk. The role of goodwill more generally in an account of trust is controversial. As Holton (1994, 65) notes: 'For instance, I can trust someone to take care of a third party or to themselves, without requiring that they have goodwill towards me.' Moreover, it might also involve further beliefs: as Hawley (2014, 10) puts it: 'To trust someone to do something is to believe that she has a commitment to doing it and to rely upon her to meet that commitment'. The ground-level idea that trust requires some kind of optimistic attitude is important in order to have a notion of trust that doesn't collapse into mere reliance. More will be said on this issue in §11. Thanks to Mona Simion for discussion on this point.

[9] See, for example, Baier (1986) and Jones (1996).

For example, one might be overly naive *or* overly cynical with respect to others' intentions or abilities.[10] In the former case, one fails to trust well if one (for instance), overestimates others' goodwill, or if one too easily perceives bad will when it is not present. In the latter case, one fails to trust well if (for instance) one, overly jaded, too easily perceives others as incapable of taking care of things that they would have easily done.

Likewise, one might have (i) a distorted introspection of personal risk; or (ii) a distorted perception of gains of betrayal to trustee. In the former case, one fails to trust well if one (for instance) fails to appreciate what is at stake for oneself—viz., the costs to oneself of betrayal by the trustee, *regardless* of what the trustee stands to gain through betrayal. In the latter case, one fails to trust well if one (for instance) fails to appreciate the extent to which the trustee would benefit by betraying, *regardless* of the costs to the truster of betrayal.[11]

4.

Successful trust is compatible with failing to trust well. One might, by exhibiting any of the above (or other) defects in trusting—trust in a way that easily would not lead to successful trust, but which does anyhow on a particular occasion.

That you can trust badly and still trust successfully is analogous to what happens in other domains of performance.[12] Even the poorest chess player's strategy might on occasion win a game.

If trusting well were entailed by trusting successfully, there would be no obvious need for an account of what it is to trust competently. But it is not. And accordingly, to the extent that we value trusting well and not *just* trusting successfully, we need to be able to say what *trusting well* involves.

Here an analogy to the value of knowledge literature is helpful: epistemologists care about more than mere true (i.e., successful) belief; unsurprisingly, the value of *knowledge* (roughly: a belief whose correctness is in some way through ability[13]) is widely taken to exceed the value of mere true belief. Not implausibly, there is a

[10] A point emphasised by Nietzsche, cited also in Baier (1986, 246) is that the power to have one's promises accepted isn't a power that is possessed equally by everyone in relation to everyone else; it is oftentimes possessed only by those with a certain social status and can (in such structures) be a de facto privilege of the elite. In the present context, an appreciation of this point is instructive in that it indicates how certain social structures may themselves have a deleterious effect on the preconditions for trusting well—a point wielded famously in the service of a political conclusion by Hobbes. For a recent and helpful discussion on the moral and epistemic features of communities of trust, see Alfano (2016).

[11] Perhaps, trustors should also expect the very act of trusting will affect trustees in certain ways; if this is the case, then a bad trustor would be insensitive to such influences. For discussion of such a view, see Faulkner (2011).

[12] See Sosa (2009, 31).

[13] This is, at least, a common template proposal for analysing knowledge accepted by most contemporary virtue epistemologists. See, along with Sosa, Greco (2003, 2010), Turri (2011) and Zagzebski (1996). Two of the principal challenges to the material adequacy of this template formula in the recent literature concern environmental epistemic luck and testimony. For discussion, see Pritchard (2012) and Lackey (2007).

species of trust the value of which is greater than merely successful trust. If the value of knowledge debate is any guide,[14] we'll locate such a value by first looking at (i) what trusting well, viz., trusting *competently*, involves, and then to (ii) successful trust that is connected to competent trusting in the right kind of way.

5.

A *competence* is, in short, a disposition to perform well in a given domain.[15] What counts as performing well, or reliably enough, depends on the domain of performance.[16] A baseball-hitting competence may require just that one hit the ball safely about 30% of the times one tries. Though a competence to ride a bike may demand a greater degree of reliability—viz., falling off the bike more than 50% of the time one tries (in normal conditions) betrays a lack of a competence to ride a bike.

On Sosa's view, competences have a 'triple-S' constitution—*seat, shape* and *situation*—with reference to which three kinds of dispositions can be distinguished: the *innermost competence* (*seat*), the *inner competence* (*seat* + *shape*), and the *complete competence* (*seat* + *shape* + *situation*).

To bring this idea into sharp relief, just consider the illustrative example Sosa offers concerning one's (complete) competence to drive a car:

> With regard to one's competence in driving, for example, we can distinguish between (a) the innermost driving competence that is seated in one's brain, nervous system, and body, which one retains even while asleep or drunk; (b) a fuller inner competence, which requires also that one be in proper shape, that is, awake, sober, alert, and so on; and (c) complete competence or ability to drive well and safely (on a given road or in a certain area), which requires also that one be well situated, with appropriate road conditions pertaining to the surface, the lighting, etc. The complete competence is thus an SSS (or an SeShSi) competence (2017, 191–92).

The idea that one can retain one's innermost competence to drive a car even while drunk or on slick roads comports well with the familiar idea—defended variously by Tony Honoré (1964), Anthony Kenny (1976) and Mele (2003, 447–70)—that one can retain a *general ability* to do something, ϕ, even when one lacks a *specific ability* to ϕ.

Consider, for example, a competent prosecutor, Akira, who is alert and in possession of the defendant's deposition (and other relevant legal documents),

[14] See, for example, Kvanvig (2003), Haddock et al. (2010), and Pritchard (2011). For an overview, see Pritchard et al. (2018).

[15] For what is perhaps Sosa's clearest exposition of this idea, see Sosa (2010).

[16] As Sosa (2015, 73) remarks 'What sets such a threshold? This will vary from domain to domain. It may be conventional and formalized, as in some professional contexts, or it may be less formal, more intuitive, as in the domain of a hunt. In each case, the threshold will be set by considerations distinctive of the domain and the proper basic aims of performances in it'.

standing in a courtroom. If we assume further that the evidence in Akira's possession sufficiently implicates the defendant, there is nothing standing between Akira and a successful prosecution. In such a circumstance, as John Maier (2018, §1.3) puts it, *every prerequisite* for Akira's successfully prosecuting the defendant is met. In such a case, Akira has the *specific ability* to prosecute the defendant. But now imagine an internally identical lawyer—Akira*—who is just like Akira in all respects, except that Akira* has been drugged and deprived of the defendant's deposition and is locked in the courthouse basement. Akira* lacks the specific ability to prosecute the defendant—she is certain to fail if she tried—though she retains the *general ability* to do so[17]; she is, after all, very different from the hopeless Zakira, who, suppose, is also drugged and deprived of the defendant's deposition and locked in the basement, but who (if not drugged, and equipped with all relevant documents, and situated in the courtroom in front of the judge) would, unlike Akira*, have no clue as to how to legally proceed, and so would fail to successfully prosecute the defendant.

How might we test for a general ability? What the foregoing suggests is that we should not ask whether one would succeed *simpliciter*, if they tried, but rather, if they would succeed if they tried *while in conditions appropriate* to the relevant performance-type.

Sosa's own view offers a clear and helpful way to model this idea: according to Sosa, we test for an innermost competence with *trigger-manifestation conditionals*[18]; we ask, of the individual: would they perform reliably enough if they tried (i) *in proper shape* and (ii) *properly situated*? where what counts as 'proper' shape and situation is relative to the domain of performance.[19] Because our competence-discerning judgments need to keep track of who would perform well in situations where (as Sosa puts it) 'human accomplishment is prized (or otherwise of special interest),'[20] it is our own human interests and needs that—as we should expect—play a role in fixing the limits of proper shape and proper situation that circumscribe a given competence-type.[21]

[17] This is a modification of an example from Maier (2018, §1.3).

[18] Sosa (2010, 466).

[19] Consider, for example, that we don't test for one's driving competence by asking: would the driver perform reliably enough (make it to the destination safely, avoid accidents, etc.) if deprived of oxygen and placed on abnormally slick roads; driving poorly in those conditions doesn't count against one's possessing a competence to drive reliably enough when in proper shape and properly situated—viz., in normal driving conditions. The same goes for more mundane competences, like visual-perceptual competences: one possesses the (innermost) visual-perceptual competence if one's visual-perceptual beliefs are reliably enough correct when one is in proper shape (i.e., awake, alert) and properly situated (not in the dark, in thick fog, etc.).

[20] Sosa (2017, 205; 2010, 466.).

[21] For further discussion, see Sosa (2010, Sects. 1–3 and 2017, 205).

6.

What would it take, then, to have an (innermost) competence to specifically *trust* well? To a first approximation, a promising answer will take the following form:

> *Trust Competence* (*Trust-C*): A competence to Trust-A successfully reliably enough, when in proper shape and properly situated.

What is it to be in proper shape and properly situated, *vis-à-vis* trusting; in what circumstances is trusting well of special human interest? Consider, by way of a comparison, how this question might be approached in the case of other dispositions: (a) the flammability of a match[22] and (b) an archery competence. In each case, the relevant trigger-manifestation conditional corresponds with shape and situational parameters that track the conditions under which the relevant success (viz., hitting the aim) is of primary interest to us.

	Trigger-manifestation conditional	Shape	Situation
Flammability of a match	If it were struck it would likely light	Dry	Plenty of oxygen
Archery competence	If you shoot, you'd likely hit	Awake, sober, alert	Plenty of light, low winds
Trust competence	If you trust, you'd likely trust successfully	?	?

Consider first the *shape* relevant to Trust-C. Presumably, this will involve *at least* certain healthy levels cognitive functioning (e.g., the sort relevant to risk assessment) which preclude various kinds of mental incapacitation.

Moreover, it's plausible that one is not in proper shape to trust if one is the subject of some form of *ex ante* manipulation or coercion. Here a distinction is needed between (i) manipulation *ex ante* into trusting (e.g., as when one is tricked into trusting[23]); and (ii) manipulation post hoc by the trustee. One may be in proper shape to trust competently *even if* one's trustee ultimately happens to betray one's trust—viz., manipulation post hoc.[24]

However, just as we don't test for a driving competence by checking one's reliability in conditions where (for instance) one is misled about the correct speed limit—for example, if pranksters swapped out a 20 mph sign for a 50 mph sign—

[22] See Sosa (2010, 465–466).

[23] For instance, suppose you are duped into trusting the medical advice of someone who presents as a doctor, but under conditions in which this deception would be undetectable even by the most cautious.

[24] This is just a corollary of the more general idea that a performance's being competent does not entail that it is successful. Compare: Even when every prerequisite is in place for a basketball player's making a free-throw, the shooter may still miss on occasion, despite attempting a competent shot. A shooter's competence, after all, is (as noted in §5) a competence to hit the target reliably enough via one's method exercised in proper shape and when properly situated. Whereas, in baseball, such a method need only be 30% reliable to qualify as a competence, an archery competence may require a more reliable method, though not an infallible method. For discussion see Carter et al. (2015) and Carter (2019).

nor, in the case of a competence to trust well, do we check whether one performs reliably enough in conditions in which (for instance) one is manipulated *into* trusting, as when one is the subject of an elaborate prank or deception (i.e., a surprise party). These are *abnormal* circumstances, not the circumstances in which we generally value the accomplishment of trusting well.

Moreover, just as we don't test whether one is a competent truster by asking whether they trust successfully reliably enough when drugged, coerced or manipulated *ex ante* (i.e., in improper shape), nor do we do so in situations where normal bounds of *risk*, *effort* and *skill* are not present. In a bit more detail: it doesn't count against someone's having a competence to trust well if the truster would not trust-A successfully reliably enough in conditions where the (a) risk to the truster is excessively high and gains of betrayal are enormous[25]; or where the level of (b) *effort* or (c) skill that would be required by the trustee to take care of things as entrusted is abnormally high.[26]

The foregoing then suggests we fill out the above table in the following way. We test for a trust competence by asking: would the truster trust successfully reliably enough when in normal mental shape and not manipulated *ex ante* into trusting (i.e., in proper shape), and in a trust situation within normal bounds of risk, effort and skill (i.e., when properly situated).[27]

7.

Consider now the following case:

> *Gettier trust*: Inspector Pazzi is attempting to catch an art thief, and to do so, he relies on a museum curator, Dr. Fell, whom Pazzi knows has always been trustworthy (unlike some of the potential experts he could have relied on), to assist him. Dr. Fell, it turns out, easily could have betrayed Pazzi in this particular situation (the art thief, it turns out, is his lover) but Fell decides ultimately by flipping a coin not to betray Pazzi.[28] Here Inspector Pazzi trusts competently *and* successfully, but that he trusts successfully isn't *because* of the competence he has, it's because Dr. Fell's coin landed the way it did.

[25] A point of clarification. There can of course be *some* risk as well as some gains for betrayal present in ordinary contexts of trusting, such that unreliability even in the presence of such risks (and gains for betrayal) would count against competence. The claim is that *abnormal* levels of both risk and gains of betrayal present conditions under which higher-than-usual failure needn't count against one's competence to trust well in the kinds of conditions under which reliable trusting is valued. Thanks to an anonymous referee for discussion on this point.

[26] Compare: we don't test for the competence to play poker in conditions under which buy-in and cash-out structures are dramatically altered.

[27] Correlatively, a *complete* trust competence is exercised only when one trusts while in proper shape and properly situated, as circumscribed.

[28] For those who take issue with the possibility of voluntary belief, the coin-flip can be changed, with no loss, to an involuntary but superstitiously grounded belief.

The above case is, no doubt, a case of *lucky* trust,[29] one where the luck in play taints our assessment of Pazzi's trusting successfully. But the luck does not resemble (in epistemology) a completely wild guess[30] or (in an athletic performance) a shot performed blindly, and this is because the truster here exercises a competence to trust well. We have, to put it simply, a kind of 'justified, successful trust' where the source the justification (or—more generally—what accounts for why the trusting is competent) has very little to do with why the trusting is successful.[31]

What this all suggests is that even when one trusts in a way that manifests competence and *also* trusts successfully, something still might fall short. One's trusting successfully might still not manifest one's competence to trust well.[32]

On Sosa's performance normativity model, when a performance's success issues (non-deviantly[33]) from a complete competence, the performance is not only successful, and competent, but *apt*.[34] This points to there being something better than trusting (merely) successfully and competently: trusting *aptly*, where one's trusting successfully manifests one's complete Trust-C competence—something Inspector Pazzi clearly lacked.

Apt trust: S trusts aptly if, and only if, S's successful Trust-A'ing manifests S's complete Trust-C competence.

[29] The luck at play here is analogous to what Pritchard (2005, 146–149) calls 'veritic luck'—viz., as when it is a matter of luck, given the way the success was attempted, that it was successful. Cf., Engel (1992). Note that Sosa (2010, 467) himself thinks that 'the Gettier phenomenon thus generalizes beyond the case of belief … A performance of whatever sort is Gettiered if it is both accurate and adroit without being apt'. These are just the structural features that are in place in the above case.

[30] Guessing (unlike ordinary beliefs) involves affirmation in a way that takes a chance on gaining truth with little to no weight given to risk of error. For further discussion of such cases—viz., where one guesses in an eye exam—see (along with §11) Sosa (2015, 74–75). Cf., Carter (2016).

[31] This kind of luck is what Pritchard (2009, Chs. 2–3) calls *intervening* luck—where luck intervenes between the subject and the success. This is distinct from *environmental* luck, where the unsafety of the success is down to the subject's being in a bad environment (one where error possibilities are modally close). For further discussion, see Pritchard and Whittington 2015). Note that the intervening/environmental distinction is a distinction that falls within the wider class of veritic (i.e., malignant) luck. See fn. 29.

[32] I take the above to be a performative analogue, in the case of trusting, to an example that is often used to illustrate 'Gettiered' performances—viz., Sosa's 'double-gust-of-wind' case. In this case, suppose a shot is released skilfully, deflected by a fluke gust of wind, and then (at the last moment) brought back toward the target by a fortuitous second gust of wind. In such a case, the shot is successful, and although it is competent, it is not successful *because* competent. See Sosa (2009, 22; 2010, 465–467 and 2017, 72–73) for discussion.

[33] For discussion of finks, mimics and masks as they interface with competences, see Sosa (2015, 96, fn. 3 and 145).

[34] Elsewhere, in the emotions literature, 'apt' may have other senses—viz., fittingness. All uses of 'apt' and 'aptness' refer to the technical sense described here, of successful because competent.

8.

Is apt trust then the best kind of trust we should aspire to? Recall here the axiological analogy with belief: belief that hits its aim (truth) because of ability—viz., knowledge, according to (robust) forms of virtue epistemology[35]—is an achievement, more valuable than true belief otherwise attained.

Apt trust, likewise, is an achievement—a success through ability. But even so, apt trust may be *fragile* in the following respect: (i) one trusts aptly when one's successful Trust-A'ing manifests one's complete Trust-C competence; (ii) *that one* manifests a complete Trust-C competence requires that one in fact be in proper shape and properly situated; (iii) one, properly constituted and situated, might nonetheless *very easily trust when not in these conditions*, where doing so would not reliably lead to successful trust.

Consider now the following case:

> *Mr. X*: Mr. X, having read *The Art of the Deal* along with several books by Tony Robbins, fancies himself a charismatic dealmaker, overestimating his influence. Mr. X entrusts Mrs. Y with information *I*, in a situation within normal bounds of risk, effort and skill, and Mrs. Y does not betray Mr. X. Mr. X's trust-A on this occasion may be apt—his successful trust manifests his competence to trust reliably enough in normal conditions. However, suppose that while Mr. X in trusting Mrs. Y has trusted aptly, he very easily would have trusted *inaptly* in those conditions. Although in entrusting Mrs. Y with information *I*, the risk to Mr. X is in fact not excessively high and gains of betrayal are within normal bounds, Mr. X. (with a distorted view of his charisma and influence, thanks to the Tony Robbins books) would easily have entrusted Mrs. Y with information *I* outside such bounds (e.g., had *I* been information that would have given Mrs. Y huge gains if divulged with little threat of her detection), in a situation where he would not have been a reliable enough truster.

In the above case, Mr. X's trusting is apt, though very easily would he have trusted *inaptly*, outside of his range of sufficient reliability, given the distorted view he has of his competence to trust, a distorted view which precludes him from accurately gauging the risks of trusting inaptly that are present. *That* Mr. X trusted aptly on this occasion as opposed to inaptly doesn't owe to any awareness of his of the threshold of his own competence to trust reliably (an awareness he lacks *ex hypothesi*), but rather just to good fortune.

9.

What, exactly, is 'missing' in the case of Mr. X? How would this best be articulated? Here it will be helpful to introduce what Sosa calls *full aptness*, as this applies to performances more generally:

[35] E.g., Greco (2003, 2010), Turri (2011) and Zagzebski (1996).

> The fully desirable status[36] for performances in general is full aptness: it is aptness on the first order guided by apt awareness on the second order that the first order performance would be apt (likely enough).[37]

In order to unpack this idea, compare Mr. X with a basketball player who shoots aptly within his threshold of sufficient reliability, but *unaware of where that threshold lies*, very easily would have shot from outside it. In such a case, the shot is apt—viz., it issues from the shooter's complete competence—despite the shooter lacking a second-order ability to competently gauge the risk of inaptness, something that requires (among other things) a competent view of the shooter's own competence.

The basketball player, then—like Mr. X.— performs aptly, but not fully aptly. There is, though, a more incisive way of putting Mr. X's situation, one that distinguishes between two *aims* one has in trusting.

Consider, as Sosa (2015) says of the basketball player, that he:

> … aims not just to succeed no matter how aptly [sic. but] to succeed aptly enough (through competence), while avoiding too much risk of failure. Their shots are assessed negatively when they take too much risk (2015, 85).

On the one hand, the truster (like the basketball shooter) aims not merely at *success*, a first-order aim, but also at *apt* success, a second-order aim.

Domain of endeavour	First-order aim	Second-order aim
Basketball	Make basket (i.e., *successful shot*)	That one's successful shot manifests one's complete shooting competence (i.e., *apt shot*)
Trust	That one's Trust-A in another is fulfilled, viz., when another has taken care of things as we have entrusted them (i.e., *successful trust*)	That one's successful Trust-A'ing manifests one's complete Trust-C competence (i.e., *apt trust*)

With respect to the first-order aim of trusting successfully, there is the matter of whether that is aptly attained (i.e., whether one's trusting successfully manifests a complete Trust-C competence). But it is a distinct question whether the further aim of *trusting aptly* is itself attained aptly. If so, this will not *just* be a matter of one's successful trusting manifesting a Trust-C competence—viz., a competence to trust-A successfully reliably enough when in proper shape and situation; aptly hitting the aim of *trusting aptly* requires also manifesting a *second-order* or reflective

[36] See, however, Kelp et al. (2017, Sect. 5) for a recent criticism of Sosa's view that performances attain fully desirable status if, and only if, they are fully apt.

[37] Sosa (2015, 85).

competence—call this *meta-Trust-C*—a competence to trust not just successfully, but *aptly*, reliably enough. Unlike the first-order Trust-C competence, a meta-Trust-C competence will be a kind of second-order 'monitoring' competence,[38] one by which the truster reliably *judges risk of inaptness*, as opposed to *merely* trusting reliably enough when in proper shape and properly situated (viz., within normal bounds of risk, effort and skill).

With reference to the foregoing, we can now capture a richer level of trust—call it *convictive trust*—as follows[39]:

> *Convictive trust* (*i.e., fully apt trust*): A subject S's trust is convictive if and only if S's successful Trust-A'ing (i) manifests one's complete Trust-C competence (i.e., is apt); and (ii) is guided to aptness by the truster's (second-order) assessment of risk through meta-Trust-C.

A notion that needs unpacking, of course, is that of *guidance* as it features in the above account of convictive trust. Guidance is needed to close a certain gap that might otherwise be present between (i) one's apt trusting; and (ii) one's apt (second-order) assessment of risk of trusting inaptly. Consider the following:

> *Sherlock*: Sherlock trusts Mrs. Hudson to complete an important task, the trust is successful (she takes care of what he entrusted her to do as he entrusted her to do it), *and* the trust is apt: his trusting successfully manifested a complete Trust-C competence. Suppose further that his trusting was *aptly risk assessed*; Sherlock aptly appreciates that *not easily would he trust inaptly* in these conditions. But because life has gotten a bit too boring, Sherlock decides *whether* to actually trust by flipping a coin, and so his apt risk-assessment in this case is in fact disconnected from his apt trusting.[40]

What separates convictive trust, trust that is fully apt, from trust like Sherlock's, which is *merely* apt and aptly assessed for risk of inaptness—is that it is not merely apt 'in the light of' an apt risk assessment (through a competent view of the subject's own competence), but also guided by that assessment—viz., where the trusting itself is because of, and not merely compresent with, the second-order risk assessment.

10.

Fully apt trusting, despite its being apt on two levels, depends—like any kind of human performance—on a certain kind of background. The most capable poker player, for instance, seems as though she could play her hand fully aptly—apt on

[38] For an early presentation of the features (including coherence) of the kinds of information monitored by reflective competence more generally, see Sosa (1997).

[39] I am of course using terminology that is analogous to Sosa's animal/reflective knowledge distinction, which focuses on belief as a performance-type with an aim. For the initial presentation of the animal/reflective distinction, see Sosa (1991). See also Sosa (2009) and (2011).

[40] Compare with Sosa's case of Diana the huntress in Sosa (2015, 69).

both levels—even if the electricity grid responsible for the ambient lighting very easily could have gone out, ruining the hand, when it luckily happened to work normally, and when the poker player simply took for granted *that* it would work properly.

This raises an important question for our account of trust: what kinds of things can a truster *non-negligently ignore*—and simply assume to be in place—when exhibiting convictive (i.e., fully apt) trust?

This is a complicated issue, and I believe it is best approached via a performative analogy from Sosa's *Judgment and Agency*. Here's Sosa:

> The athlete needs to consider various shape and situation factors: how tired he is, for example, how far from the target, and so on, for the many shape and situation factors that can affect performance. But there are many factors that he need not heed. It is no concern of an athlete *as such* whether an earthquake might hit, or a flash tornado, or a hydrogen bomb set off by a maniac leader of a rogue state, and so on. As an athlete, he is not negligent for ignoring such factors (2017, 191, my italics).

There is, plausibly, a distinction *within* the class of factors that could cause a given performance to fail, between

(i) the kinds of things a fully apt performer must heed in order to safeguard against credit-reducing luck; and
(ii) the kinds of things he or she is free to non-negligently assume are already in place.

Sosa refers to the kinds of things in category (ii) *background conditions*.[41] Cobbling together from his various descriptions of them,[42] we can identify five key features of background conditions: logical, functional, modal, epistemological and normative.

> *Logical* Background conditions are *entailed* by the presence of pertinent seat, shape and situation conditions; they must hold if the relevant 'S' [seat/shape/situation] is in place at the time of the performance.[43]

[41] It may be helpful here to register some parallels between what Sosa calls background conditions with what Dancy (2000, 127–130) calls *enabling conditions*, as they feature in account of acting for a reason. Being born, for instance, is an an enabling condition for my (say), doing something ϕ-ing for a reason at a later time in the sense that it is a necessary precondition for my doing so, even though it is not a *reason* for my ϕ-ing. Likewise, he thinks, *believing that p* is something that's necessary for one's *ϕ-ing because p* despite not being one's reason for ϕ-ing. For helpful discussion of Dancy's and related accounts of acting for a reason, see Marcus (2012, Ch. 2).

[42] See, for example, Sosa (2017, 215–21; 2015, Ch. 3).

[43] Sosa (2017, 218). This relationship may also be captured as metaphysical. As an aside, the description 'logical' is used here because the condition captures a kind of entailment, and because the modal characterisation is likewise metaphysical.

Functional Background conditions are orthogonal to a performance *qua* the kind of performance it is.[44]
Modal Background conditions need not hold safely
Epistemological When a background condition holds, the performer need not know that it does.
Normative The quality of a performance is not reduced or in any way effected by reducing the safety a background condition.

The epistemological and normative features of background conditions are of special importance, as they capture the threshold of permissible ignorance in fully apt trusting—one may take for granted whatever, not related to the performance-type *qua* performance-type—is entailed by the relevant 'S' [seat/shape/situation] being place at the time of the performance.

We've seen what fully apt (i.e., convictive) trust demands of a truster. With reference to the above account of background conditions, let's now investigate what such trust *permits*.

11.

Let's use the following case as a reference point:

Loan Payment: A trusts B to pay back a (modern, online) financial debt, which B repays as entrusted to do. Let's assume further that all conditions for fully apt (convictive) trust are met. So A's trusting B on this occasion is apt—and even more—guided to aptness by A's assessment of risk through meta-Trust-C.

What, exactly, are the *background conditions* in *Loan Payment*, conditions which we may suppose that A could freely and non-negligently assume to be in place, and which—even if they didn't hold safely—this wouldn't count against the quality of A's trusting?

As noted in the previous section, background conditions are entailed by the presence of pertinent seat, shape and situation conditions in the sense that they must hold if the relevant 'S' [seat/shape/situation] is in place at the time of the performance. Let's focus on what is entailed by the presence of the relevant (a) shape and (b) situation in the case of Trust-C.

(a) Shape: Background conditions

One is in proper shape, *vis-a-vis* trusting, only when at least (i) mentally fit (i.e., not physically or mentally incapacitated in a way that would have a deleterious effect on the reliability one's trusting) and (ii) when not coerced or manipulated *ex ante* intro trusting (see §6). Implied by the relevant sort of mental fitness is the presence of healthily functioning neurotransmitters; suppose that, had A been slipped (for

[44] Consider, in particular, Sosa's remark that it is of no concern of an athlete *as such* whether, e.g., a tornado might easily have it. See Sosa (2017, 218). For related discussion, see Carter (2017).

instance) Alpha-PVP (i.e., flakka), A's neurotransmitters would be functioning abnormally and would have impeded A's ability to understand details relevant to A's trust exchange with B. Even if A were such that very easily A may have been (undetectably) slipped flakka prior to the trust exchange which would have ruined A's shape, but by luck was not, A may (while trusting fully aptly) non-negligently take for granted that this was not so.

A precondition of S's *not* being manipulated or otherwise coerced into trusting *ex ante* is that A is capable of exercising base-level autonomy with respect to whom to trust; if (for instance) A were suddenly hypnotized, A would not be in such an autonomous position—A's capacity to avoid manipulation and coercion would be undermined. Even if A very easily could have been, prior to trusting, subjected to undetectable shape-ruining hypnosis, A may, while trusting fully aptly, non-negligently take for granted that this was not so.

(b) Situation: Background conditions

More complex and perhaps also interesting, ethically and epistemologically, are the background conditions that underlie the *situational* component of the Trust-C competence: normal bounds of skill/risk/effort (see §6). Let's take each in turn.

Skill. Thanks to working online banking services, replaying a complex modern loan as entrusted doesn't require the trustee have specialised knowledge, e.g., of partial differential equations, stochastic calculus, etc., that are beyond *normal* intellectual capacities[45]; doing so requires performing a sequence of simple online steps within what is usually an intuitive choice architecture design; were online banking services to suddenly fail (and experts to charge exorbitant fees), the trustee could successfully (and with suitable accuracy) pay back and properly document the complex loan only through possessing abnormally high financial and mathematical skill. Even if world online banking systems easily could have crashed beyond A's ken but did not, A may, while trusting fully aptly, non-negligently take for granted that this was not so.

Risk Thanks to a well-functioning stock market and currency system, A would not be abnormally exposed if the debt were not paid back; if the market suddenly crashed, A would be so exposed (and likewise, suppose, B would stand to gain significantly more through betrayal than B does with the stock market failing to surprisingly crash). Even if, beyond A's ken, a cabal of oligarchs easily could have, but just so happened to not, sabotage the stock market and world currency values in a way that would have left B standing to gain much more by betrayal than otherwise, A may, while trusting fully aptly, non-negligently take for granted that this was not so.

[45] I am not suggesting here that such mathematics falls outside of what normal, healthy cognitive functioning is *fit* for. Such mathematical skill, even if within the reach of normal human cognition when suitably applied, is *specialised*; it reflects expertise and is the product of training the lack of which is both normal in the sense of widespread in typical communities, and which would (in the circumstance described) render repaying the debt considerably more difficult and thus unlikely.

Effort On the supposition that A is easily locatable by B through normal means, it is does not take much effort for the trustee to repay the debt to the truster; if A's identity had been stolen, even a mathematically skilled and well-intended trustee would have to expend considerable effort to repay A *rather* than A's identity thief. Even if, beyond A's ken, a group of hackers might easily have, but did not, steal A's identity in a way that would have required enormous effort by B to locate A to repay the loan, A may, while trusting fully aptly, non-negligently take for granted that this was not so.[46]

In sum, the fully apt truster can non-negligently take for granted background conditions that are implied by the obtaining of the relevant seat, shape and situation constituents of Trust-C. Even if the background conditions themselves hold unsafely beyond the subject's ken, this is compatible with the subject trusting fully aptly. *Incompatible* with trusting fully aptly is trusting even when shape and situation conditions could easily not hold (but for reasons that are not down to the unsafety of background conditions)—viz., as in the case of Mr. X, but *not* in the case of *Loan Payment*.

12.

If the foregoing view is right, then a bi-level approach is needed to distinguish between two species of trust, each of which is of human interest, and each of which is structured differently with respect to competence and risk.

In this section, I want to outline some additional benefits that such a view offers over more traditional univocal accounts of trust in ethics and epistemology.

(a) Puzzles about reflection

One of the perennial problems about the nature of trust concerns the interplay between trust, risk and reflection. As Annette Baier[47] vividly expresses the idea:

> Trust is a fragile plant […] which may not endure inspection of its roots, even when they were, before inspection, quite healthy.

To make this idea—viz., that reflection on (and/or monitoring of) the trust relationship imperils trust—more concrete, consider the following case, due to Wanderer and Townsend (2013):

> *Paranoid parent*. A paranoid parent […] organises a babysitter for their child, and then proceeds to spend the evening out monitoring their babysitter's antics remotely, via a 'nanny-cam'. The paranoid parent is not only a lousy date, but

[46] A prerequisite for the obtaining of *all* situational features is that that the world exists. Had a maniac detonated a bomb, the truster, trustee and the banks and financial systems would be obliterated and the trustee could not repay the debt. We may suppose accordingly that even if such a maniac easily could have done so but did not, A may trust fully aptly while taking for granted that whatever preconditions to trusting are furnished by the existence of the world are intact. For related discussion, see Sosa (2017, 191).

[47] Baier (1986, 260).

> also a lousy trustor; in performing the seemingly rational act of broadening the evidential base relevant to her judgments of trustworthiness, she is, precisely, *failing to trust the babysitter* (2013, 1).

The idea that actively reflecting on the trust relationship so as to minimise risk is *itself* at tension with genuine trust has been raised in various ways in the literature on the rationality of trust.[48] As McLeod (2015, §2) states the puzzle succinctly:

> Since trust inherently involves risk, any attempt to eliminate the risk through rational reflection could eliminate the trust at the same time.

A bit more permissively, Faulkner (2018, §5) remarks that:

> […]*[T]oo much* reflection can undermine trust.[49]

The principal challenge for univocal accounts of trust is to explain what trust is, i.e., its nature, while—at the same time—making sense of how the kind of reflection or monitoring that would seem *prima facie* needed to improve the quality of trusting would not at the same time have the consequence of destroying it.[50]

A pleasing feature of the bi-level account proposed is that—at least in one important respect—the account offers a way to bypass the puzzle.[51] Second-order reflection improves the quality of trusting, and at the same time, it is compatible with *genuinely* trusting. Genuine trust seems, as the puzzle goes, threatened by monitoring of the trustee. But the principal role of reflection in the proposed account is not reflection on, or monitoring of, whether the trustee will betray one's trust; its role in the account of convictive trust is one of *self*-regulation, where the object of the reflection, monitoring and awareness at the second order is in the main one's *own* competence and the conditions of its exercise.

(b) Internalism and the quality of trust

However one addresses the previous puzzle, it is widely thought—in particular among philosophers of trust sympathetic with internalist epistemology[52]—that one

[48] As Dasgupta (2000, 51) puts it, trust must be *prior* to any monitoring of the trustee: 'If I can monitor what others have done before I choose my own action, the word'trust' loses its potency'. Likewise, accordingly to Baier (1986, 260): 'Trust is a fragile plant […] which may not endure inspection of its roots, even when they were, before inspection, quite healthy'. For a helpful overview, see Wanderer and Townsend (2013). Cf., Möllering (2006, Ch. 5) for a more radical expression of the idea that trust cannot withstand rational scrutiny.

[49] Note that, beyond philosophical considerations in favour of this constraint on trust, there are also empirical reasons to embrace it, or something like like it. See for instance Fuchs (2010) for a discussion of how reflection and monitoring have a deleterious effect on trust in psychopathology.

[50] In the recent literature, doxastic accounts of trust, according to which trusting either is or involves the belief that the trustee will be trustworthy, have in particular attempted to resolve this problem. See, for instance, Hieronymi (2008, 213–36) and McMyler (2011). For discussion, see Faulkner (2018, §5).

[51] To be clear, the puzzle is bypassed only in the sense that the view proposed is able (better so than other proposals) to preserve the compatibility of trust and (trust-relevant) reflection. Whether the puzzle can be *solved*—viz., presumably by showing either that one of the claims that generates it is false—is beyond what I'm aiming to show here.

[52] See, for instance, Fricker (1995), Lipton (1998), and Lehrer (2006).

ought to have good *reasons* for trusting others, in particular, in cases where there are significant stakes. A correlative of this idea is that successful trust, backed by good reasons for trusting, is better (from the point of view where good trusting is valued) than trust otherwise secured.

For example, Russell Hardin (2002, 12) maintains that the quality of one's trusting is a function of the (continually) updated reasons one has that the trustee will be trustworthy, which might come from such things as inductive generalisations and past experiences, etc.[53] Having such an internal perspective that backs one's trusting is surely valuable from the perspective where trusting well is valued. Though it cannot exhaust such value.

Here a comparison in epistemology is illustrative: why is it epistemically valuable (i.e., from the point of view where we care about the truth) for our beliefs be reasonable in light of one another?

Here's Sosa (1997) in an early paper on reflective knowledge:

> Coherence-seeking inferential reason, like retentive memory, is of epistemic value when combined with externally apt faculties of perception, because when so combined it, like retentive memory, gives us a more comprehensive grasp of the truth than we would have in its absence (1997, 421).

If externally competent faculties are working as they should, then having broad coherence in our beliefs is desirable in part because it is truth-conducive, and this is so 'even if in a demon world […] coherence fails this test' and would not promote true belief' (ibid., 422).

Analogously with trust: trusting backed by a rational and internally coherent perspective is valuable, *vis-à-vis* trusting well, *when* externally competent faculties (viz., Trust-C) are working as they should; in such circumstances, such coherence is desirable because it is conducive of successful trust; and this is so even if in a world where trust were categorically betrayed,[54] such internal coherence would fail this test.

If the foregoing is right, the value of trusting well cannot be explained as internalists such as Hardin suggests, even if internalist considerations can (if combined with externally competent faculties) promote successful trust. By contrast, the bi-level account fits snugly with the above points.

In short: internal coherence is (a) an important feature of convictive trust and is essential to one's (second-order) ability to place one's first-level trust in perspective; and (b) is desirable (*vis-à-vis* our aims in trusting) because in our actual world (where our first-order trust competence really is reliable) it is conducive to successful trust. This is not meant to suggest that such an internal perspective is

[53] For helpful discussion, as well as a survey of some criticisms, see McLeod (2015, §5).

[54] The viability of such a world has been contested. See, in particular Coady (1992, 85), which presents the famous Martian Argument which casts doubt on systematically false reporting. A more general about the untenability of systematic lying owes to Kant.

needed for *apt* trust—it is not (much, anyway)[55]—but rather to affirm that beyond apt trust there is a better species of trust.[56]

(c) Enriching trust (and trustworthiness)

A separate point worth addressing about the quality of trust is due to Jones (2012, 72), and in particular, to her remarks about *rich trustworthiness*. Firstly, note that in paradigmatic cases of interpersonal trust, we trust someone with *something* (i.e., to take care of some particular things, as entrusted). This is sometimes called 'three-place-trust'[57] so as to distinguish such trusting from *two-place* trust, as when one trusts another *simpliciter*.

While apt and convictive trust, as presented, are models of three-place-trust,[58] two-place trust is also valuable. Consider Jones' remarks on two- and three-place trustworthiness, and why two-place trustworthiness may be of special interest.

> When we talk about cultivating trustworthiness, we sometimes have three-place trustworthiness in mind, as when, for example, we talk about ways of fostering the trustworthiness of doctors with respect to their patients. However, often, when we talk about cultivating trustworthiness, our target is *rich trustworthiness*: we want both to increase the range of domains over which people will be competent and responsive to dependency and to improve those capacities required to have a reliable grasp of these zones of competence and to be able to signal them to others (2012, 72).

By parity of reasoning, in the case of cultivating the capacity to trust well, our target is an analogous one—viz., we value increasing (i) the range of domains over which people will *trust competently*; but in addition to this, we *also* aim 'to improve those capacities required to have a *reliable grasp of these zones of competence* [sic to trust well] and to be able to signal them to others' (ibid., 72, my italics).

It should be emphasised that if the capacity to trust well is likewise *enriched* by way of (i) *and* (ii), as it plausibly would, then this will require not *merely* the cultivation of Trust-C, but also the cultivation of meta-Trust-C through which one has a reliable grasp of one's zones of (Trust-C) competence.

The bi-level account neatly explains why trust would be enriched by cultivating *both* Trust-C and meta-Trust-C. The cultivation of the former is conducive to apt

[55] One might initially balk at this thought; doesn't the possession of Trust-C require internal coherence? Not necessarily. Consider that Trust-C is a function of one's being reliable at securing an end (successful trust) under a range of circumstances under which securing that end is of human interest—viz., when one is in proper shape and properly situated. The factors that make one reliable at securing successful trust when in proper shape and properly situated may in a truster lie below the surface of conscious endorsement.

[56] These points are presented, intentionally, in a way that is analogous to the way Sosa summarises the role of coherence in his bi-level virtue epistemology in Sosa (1997, 422).

[57] See, for example, Horsburgh (1961) and Hardin (1992).

[58] I am sympathetic to Hawley's suggestion that trust is '*primarily* a three-place relation, involving two people and a task'; this is compatible with recognising other valuable ways of trusting. See Hawley (2014, 2, my italics).

trust, which is good. The cultivation of the latter, when apt trust is secured, is conducive to convictive trust, which is even better.

(d) Varieties of risk assessment

We've already noted that trust and risk are conceptually connected[59]; trusting inherently involves vulnerability to the risk that the trustee will *not* take care of things as entrusted to.

However, the varieties of risk assessment that befit a good truster are not limited to assessing this risk—that the trustee will not take care of things as entrusted—even under a wide construal of what kinds of factors are relevant to this particular risk. There are *other* risks inherent in trusting.

Following Duncan Pritchard (2015, 437) it will be helpful to distinguish between (a) a *risk event*[60]; and (b) *riskiness*. A risk event (alternatively called a 'risk') is a possible but unwanted or harmful outcome, whereas an event is *risky* (or 'high risk') when the chance of the risk event materialising is higher than is normal for an activity of the relevant kind.

Harms internal to a domain are fixed by the aims of that domain. Trusting, on the account proposed (i.e., §9), has two aims—the first-order aim of trusting successfully, and the second-order aim of trusting aptly. The non-obtaining of these aims are distinct risk events. There are thus two *kinds* of trust-relevant riskiness: that which features when the chance[61] of the non-obtaining of the first-order aim (successful trust) is higher than normal; and that which features when the chance of the non-obtaining of the second-order aim (apt trust) is higher than normal.

Risk management techniques appropriate to each aim may come apart. One may manage first-order risk (by limiting first-order riskiness) while lacking any good way to manage second-order risk. This is the situation we find in cases like that of Mr. X., and more generally, in cases where a truster possesses, and exercises, Trust-C in ways that are not *fully* apt. Conversely, in cases like *Sherlock*, one's second-order risk management techniques may not bear on one's management of first-order risk.

An advantage that the bi-level account enjoys over univocal accounts of trust is that it can explain in a straightforward way why both varieties of trust-relevant risk management, neither of which entails the other, should be valued and cultivated.

(e) The distinction between trust and reliance

[59] See Nickel and Vaesen (2012) for further discussion.

[60] As is noted by Hansson (2004), in some academic disciplines, the term 'risk' is used alternatively to refer to (i) the probability of such a harm, or (ii) the expected disutility of such a harm. Hansson (2014).

[61] Such chance is generally modelled in terms of probabilities. However, there may be advantages to modelling risk modally rather than probabilistically. For some recent defences of this idea, see Pritchard (2015, 2016).

What is the difference between trust and mere reliance? One familiar line adverts to trust's involving a dependence not only on another to take care of something, but on another to do so compresently with certain attitudes—viz., with goodwill.

As Baier (1986) remarks, trust:

> […] seems to be reliance on their good will toward one, as distinct from their dependable habits, or only on their dependably exhibited fear, anger, or other motives compatible with ill will toward one, or on motives not directed on one at all (1986, 234).

In a similar spirit, Karen Jones (1996) writes:

> One can only trust things that have wills, since only things with wills can have goodwills-although having a will is to be given a generous interpretation so as to include, for example, firms and government bodies. Machinery can be relied on, but only agents, natural or artificial, can be trusted (1996, 14).

Furthermore, concerning the appropriateness of our attitudes toward misplaced trust compared to misplaced reliance, Hawley (2014) writes:

> Suppose I trust you to look after a precious glass vase, yet you carelessly break it. I may feel betrayed and angry; recriminations will be in order; I may demand an apology. Suppose instead that I rely on a shelf to support the vase, yet the shelf collapses, breaking the vase. I will be disappointed, perhaps upset, but it would be inappropriate to feel betrayed by the shelf, or to demand an apology from it (2014, 2–3).

For the present purposes, I'm disputing none of these points.[62] Rather, I want to suggest how the distinction between trust and mere reliance may in fact run deeper, and why the bi-level account is nicely situated to account for why this is so.

Consider that the satisfaction conditions for reliance are specifiable *extra-agentially*, in terms of whether someone, or something, does (or is) as one depends on it (him, or her) to do or to be. In the case of interpersonal reliance, we might rely on the reliably self-centred person to behave as expected; whether or not this reliance is *disappointed* (as opposed to betrayed) is not a matter of what the relier *herself* does, but of what the relied upon does.

The satisfaction conditions for trust, however, are importantly *not* (exclusively) extra-agential, but only partly so. Just like the basketball player who chucks it from anywhere aims, each time he chucks the ball, *to make the basket*, the better player aims to shoot aptly, an aim the satisfaction conditions for which include not just making it, but also making *a well-selected shot*, conditions satisfied in part by the agent's own contribution to the performance. Correspondingly—and this connects with the discussion in §9—in trusting of the sort that mature humans aspire to, we

[62] There is some scope to dispute them. One way to do so is to distinguish between trust that involves mere reliance and trust that does not; Faulkner (2007, 880) makes such a move in distinguishing between what he calls affective and predictive trust. I'm inclined to agree, though, with Hawley (2014, 4) on the point that a 'distinction is important because trust, not mere reliance, is a significant category for normative assessment'.

aim not *just* at avoiding being betrayed—at successful trust *however* we may get it[63]—but we also aim at apt trust, and the satisfaction conditions for the latter include not only the trustee's taking care of things as entrusted, but the also the truster's *selecting when to trust* well, conditions satisfied in part by what the trusting subject brings to the trusting.

Trust and reliance accordingly differ not only with respect to (i) the former but not the latter depending on the goodwill of the trustee; and (ii) the appropriateness of our attitudes toward misplaced trust in comparison with misplaced reliance; but also with respect to (iii) the role of the trusting subject in a specification of their respective satisfaction conditions.[64]

In sum, a bi-level account of trust has much to recommend it. The view has advantages over univocal accounts of trust in matters to do with reflection, trust quality, trust-enrichment, risk-assessment and the distinction between trust and reliance.

A transition from a single level to a bi-level account of trust also has more general advantages. It allows us to see more clearly what trusting well involves, how this connects with the goals mature humans have in trusting, why these goals are worthy ones, and why it matters how we attain them.

Acknowledgements Thanks to audiences at the University of Glasgow, the University of Edinburgh, and University College Dublin for helpful feedback. I'm also grateful to Emma C. Gordon, Christoph Kelp, Mona Simion and Ernest Sosa for helpful discussion, as well as to an anonymous referee at *Philosophical Studies.*

[63] Consider an analogy with belief: If we aimed to get the truth *any old way* (affording no weight to the disvalue of error) a wild guess (as opposed to withholding) would be advised even in the absence of evidence. Unless under practical duress, a good thinker aims affirms with the aim of getting it right not just any old way (but not by stab in the dark) but by a more reliable means. A similar point holds for trusting; a good truster doesn't trust in a way that disproportionately weights the competing aims of (i) attaining successful trust; and (ii) avoiding distrust, so as to give all the weight to the former and none to the latter; doing so would be trusting as the guesser guesses. For related discussion, see Carter (2017).

[64] This third difference between trust and reliance may be easy to overlook given the attention to the attitudinal features of trust that feature in debates between doxastic and non-doxastic accounts of (univocal) trust. Within this debate, proponents of doxastic accounts as well as non-doxastic accounts differ about *whether* one's trusting another to do something involves the belief that the trustee will do that thing. An affirmative answer is given by proponents of doxastic accounts such as Hieronymi (2008), and McMyler (2011). A negative answer is given by proponents of non-doxastic accounts such as Faulkner (2011) and Jones (1996). *Both sides*, though, seem hard pressed to account for (iii). In the former case, this is because the satisfaction conditions for reliance on another in conjunction with a belief that the other will do that thing, are (equally) extra-agential. Non-doxastic accounts, such as Faulkner's and Jones', according to which what's required in addition to reliance is an expectation that the trustee be moved by the fact that S relies on the trustee to do the thing in question, likewise have extra-agential satisfaction conditions. Whether the aim in trusting is satisfied, in either case, is just a matter of how things go outside the scope of the agent's own trusting.

References

Alfano, M. (2016). The topology of communities of trust. *Russian Sociological Review, 15*(4), 30–56.
Baier, A. (1986). Trust and antitrust. *Ethics, 96*(2), 231–260.
Carter, J. A. (2016). Sosa on knowledge, judgment and guessing. *Synthese*. https://doi.org/10.1007/s11229-016-1181-2.
Carter, J. A. (2017). Review of epistemology by Ernest Sosa, *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/news/epistemology/.
Carter, J. A. (2019). Exercising abilities. *Synthese*. https://doi.org/10.1007/s11229-019-02227-4.
Carter, J. A., Jarvis, B., & Rubin, K. (2015). Varieties of cognitive achievement. *Philosophical Studies, 172*(6), 1603–1623.
Carter, J. A., & McKenna, R. (2018). Kornblith versus Sosa on grades of knowledge, *Synthese*. https://doi.org/10.1007/s11229-018-1689-8.
Chan, T. (2013). *The aim of belief*. Oxford: Oxford University Press.
Coady, C. A. J. (1992). *Testimony: A philosophical study*. Oxford: Oxford University Press.
Coleman, J. (1990). *Foundations of social theory*. Belknap: Cambridge, MA.
Dancy, J. (2000). *Practical reality*. Oxford: Clarendon Press.
Dasgupta, P. (2000). Trust as a commodity. *Trust: Making and breaking cooperative relations, 51,* 49–72.
Engel, M. (1992). Is epistemic luck compatible with knowledge? *The Southern Journal of Philosophy, 30*(2), 59–75.
Faulkner, P. (2007). On telling and trusting. *Mind, 116*(464), 875–902.
Faulkner, P. (2011). *Knowledge on trust*. Oxford: Oxford University Press.
Faulkner, P. (2018). Testimony and trust. In S. Judith (Ed.), *The Routledge handbook on trust*. London: Routledge.
Fricker, E. (1995). Critical notice: Telling and trusting—Reductionism and anti-reductionism in the epistemology of testimony. *Mind, 104*(414), 393–411.
Fuchs, T. (2010). The psychopathology of hyperreflexivity. *The Journal of Speculative Philosophy, 24*(3), 239–255.
Gambetta, D. (1988). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–237). Oxford: Blackwell.
Greco, J. (2003). Knowledge as credit for true belief. In M. DePaul & L. Zagzebski (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology*. Oxford: Oxford University Press.
Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge: Cambridge University Press.
Haddock, A., Millar, A., & Pritchard, D. (Eds.). (2010). *The nature and value of knowledge: Three investigations*. Oxford: Oxford University Press.
Hansson, S. O. (2004). Philosophical perspectives on risk. *Techné: Research in Philosophy and Technology, 8*(1), 10–35.
Hansson, S. O. (2014). Risk. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Palo Alto, CA: Stanford University.
Hardin, R. (1992). The street-level epistemology of trust. *Analyse & Kritik, 14*(2), 152–176.
Hardin, R. (1996). Trustworthiness. *Ethics, 107*(1), 26–42.
Hardin, R. (2002). *Trust and trustworthiness*. New York City: Russell Sage Foundation.
Hawley, K. (2014). Trust, distrust and commitment. *Noûs, 48*(1), 1–20.
Hieronymi, P. (2008). The reasons of trust. *Australasian Journal of Philosophy, 86*(2), 213–236.
Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy, 72*(1), 63–76.
Honoré, A. (1964). Can and can't. *Mind, 73*(292), 463–479.
Horsburgh, N. J. H. (1961). Trust and social objectives. *Ethics, 72*(1), 28–40.
Jones, K. (1996). Trust as an affective attitude. *Ethics, 107*(1), 4–25.
Jones, K. (2012). Trustworthiness. *Ethics, 123*(1), 61–85.
Kelp, C., et al. (2017). Hoops and Barns: A new dilemma for Sosa. *Synthese*. https://doi.org/10.1007/s11229-017-1461-5.
Kenny, A. (1976). *Will, freedom, and power*. London: Blackwell.
Kornblith, H. (2009). Sosa in perspective. *Philosophical Studies, 144*(1), 127–136.
Kornblith, H. (2010). What reflective endorsement cannot do. *Philosophy and Phenomenological Research, 80*(1), 1–19.

Kornblith, H. (2012). *On reflection*. Oxford: Oxford University Press.
Kvanvig, J. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge: Cambridge University Press.
Kvanvig, J. (2008). Pointless truth. *Midwest Studies in Philosophy, 32*(1), 199–212.
Lackey, J. (2007). Why we don't deserve credit for everything we know. *Synthese, 158*(3), 345–361.
Lehrer, K. (2006). Testimony and trustworthiness. In J. Lackey & E. Sosa (Eds.), *The epistemology of testimony* (pp. 145–159). Oxford: Oxford University Press.
Lipton, P. (1998). The epistemology of testimony. *Studies in History and Philosophy of Science-Part A, 29*(1), 1–32.
Maier, J. (2018). Abilities. In E. N. Zalta (Ed.), The stanford encyclopedia of philosophy (Spring 2018 Edn.). https://plato.stanford.edu/archives/spr2018/entries/abilities/.
Marcus, E. (2012). *Rational causation*. Cambridge, MA: Harvard University Press.
McLeod, C. (2015). Trust. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Palo Alto, CA: Stanford University.
McMyler, B. (2011). *Testimony, trust, and authority*. New York: Oxford University Press.
Mele, A. R. (2003). Agents' abilities. *Noûs, 37*(3), 447–470.
Möllering, G. (2006). *Trust: Reason, routine* (p. 2006). Reflexivity: Emerald Group Publishing.
Nickel, P. J., & Vaesen, K. (2012). Risk and trust. In S. Roeser (Ed.), *Handbook of risk theory* (pp. 857–876). Berlin: Springer.
Perrine, T. (2014). Against kornblith against reflective knowledge. *Logos & Episteme, 5*(3), 351–360.
Potter, N. P. (2002). *How can i be trusted? A virtue theory of trustworthiness*. London: Rowman & Littlefield.
Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford University Press.
Pritchard, D. (2009). *Knowledge*. London: Palgrave Macmillan.
Pritchard, D. (2011). What Is the swamping problem. In A. Reisner & A. Steglich-Petersen (Eds.), *Reasons for belief* (pp. 244–259). Cambridge: Cambridge University Press.
Pritchard, D. (2012). Anti-luck virtue epistemology. *Journal of Philosophy, 109*(3), 247–279.
Pritchard, D. (2015). Risk. *Metaphilosophy, 46*(3), 328–349.
Pritchard, D. (2016). Epistemic risk. *Journal of Philosophy, 113*(11), 550–571.
Pritchard, D., Turri, J., & Carter, J. A. (2018). The value of knowledge. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (spring 2018 edition). https://plato.stanford.edu/archives/spr2018/entries/knowledge-value/.
Pritchard, D., & Whittington, L. J. (Eds.). (2015). *The philosophy of luck*. London: Wiley-Blackwell.
Shah, N. (2003). How truth governs belief. *Philosophical Review, 112*(4), 447–482.
Shah, N., & Velleman, J. D. (2005). Doxastic deliberation. *Philosophical Review, 114*(4), 497–534.
Simon, J. (2013). Trust. In D. Pritchard (Ed.), *Oxford bibliographies in philosophy*. Oxford: Oxford University Press.
Sosa, E. (1991). *Knowledge in perspective: Selected essays in epistemology*. Cambridge: Cambridge University Press.
Sosa, E. (1997). Reflective knowledge in the best circles. *The Journal of Philosophy, 94*(8), 410–430.
Sosa, E. (2009). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. I). Oxford: Oxford University Press.
Sosa, E. (2010). How competence matters in epistemology. *Philosophical Perspectives, 24*(1), 465–475.
Sosa, E. (2011). *Knowing full well*. Princeton: Princeton University Press.
Sosa, E. (2015). *Judgment and agency*. Oxford: Oxford University Press.
Sosa, E. (2017). *Epistemology*. Princeton: University Press.
Turri, J. (2011). Manifest failure: The gettier problem solved. *Philosopher's Imprint, 11*(8), 1–11.
Wanderer, J., & Townsend, L. (2013). Is it rational to trust? *Philosophy Compass, 8*(1), 1–14.
Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge: Cambridge University Press.