

Therapeutic Trust

Abstract This paper develops and defends a new account of *therapeutic trust*, its nature and its constitutive norms. Central to the view advanced is a distinction between two kinds of therapeutic trust—*default therapeutic trust* and *overriding therapeutic trust*—each which derives from a distinct kind of trusting competence. The new view is shown to have advantages over extant accounts of therapeutic trust, and its relation to standard (non-therapeutic) trust, as defended by Pamela Hieronymi,¹ Karen Frost-Arnold,² and Karen Jones.³

1. Suppose you are leaving town for the weekend and need someone to watch your house, feed your pets and water your plants. Now imagine two choices you might make. You might trust a reliable friend who has an established track record of responsibility and loyalty. But, you might instead trust your 16-year-old nephew with no such track record to speak of. In the latter kind of case, suppose trust is undertaken with the intended aim of bringing about (or increasing) trustworthiness.⁴ Philosophers of trust often use the term ‘therapeutic trust’ to refer to this latter species of trust, in order to distinguish it from more standard cases of (non-therapeutic) interpersonal trust.

The matter of how exactly to characterise the relationship between non-therapeutic and therapeutic trust is contested.⁵ Here is the problem in a nutshell. Philosophical accounts of the nature of trust attempt to say what trusting someone with something essentially involves⁶, typically by focusing on how exactly to characterise the kind of trusting attitude one has towards her trustee.

1 “The Reasons of Trust,” *Australasian Journal of Philosophy* 86, no. 2 (2008): 213–36.

2 “The Cognitive Attitude of Rational Trust,” *Synthese* 191, no. 9 (2014): 1957–74.

3 “Trust and Terror,” in *Moral Psychology: Feminist Ethics and Social Theory*, ed. Peggy DesAutels and Margaret Urban Walker (Rowman & Littlefield, 2004), 3–18.

4 See, for example, Howard John Neate Horsburgh, “The Ethics of Trust,” *The Philosophical Quarterly* (1950-) 10, no. 41 (1960): 343–54 and Jones, “Trust and Terror.”

5 See, e.g., Horsburgh, “The Ethics of Trust.” and Victoria McGeer, “Trust, Hope and Empowerment,” *Australasian Journal of Philosophy* 86, no. 2 (2008): 237–54. For an overview, see Carolyn McLeod, “Trust,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2015 (<https://plato.stanford.edu/archives/fall2015/entries/trust/>; Metaphysics Research Lab, Stanford University, 2015), 1.

6 The kind of trust that is principally at issue in debates about therapeutic trust is *three-place* trust, e.g., with an infinitival component (schematically: A trusts B to X). As Baier “Trust and Antitrust,” *Ethics* 96, no. 2 (1986): 236. puts it, philosophers of trust—and not just those interested in

Once such accounts are made precise, it looks like therapeutic trust—given how the attitude we have in such cases about the trustee’s reliability is usually much *less* optimistic than in non-therapeutic cases—either (i) simply doesn’t get ‘ruled in’ as genuine trust on the account, or (ii) the account gets modified—perhaps stretched quite thin—in order to fit therapeutic trust in.

Here is the plan. In §§2-4, I discuss three notable ways philosophers of trust have attempted to deal with the tricky issue of therapeutic trust and its relationship with ordinary non-therapeutic trust: (§2) Pamela Hieronymi’s pure/impure approach⁷; (§3) Karen Frost-Arnold’s ‘unity’ approach⁸; and (§4) Karen Jones’s ‘normative difference’ approach⁹. Each is shown to be problematic. §§5-9 then develops a new way of thinking about therapeutic trust which avoids the problems facing the other three views while at the same time offering its own additional advantages. §10 concludes by canvassing some potential objections and replies.

2. According to Pamela Hieronymi,¹⁰ therapeutic trust is not ‘pure’ or ‘full-fledged’ trust. Trust is full-fledged (alternatively: pure) only if one actually *believes* that the person in question will do the thing in question.¹¹ One can risk

therapeutic trust—are concerned centrally with ‘one person trusting another with some valued thing’; likewise, as Katherine Hawley, “Trust, Distrust and Commitment,” *Noûs* 48, no. 1 (2014): 2 puts it: trust is ‘primarily a three-place relation, involving two people and a task’. On one popular way of thinking about the relationship between three-place trust and two-place trust, latter can be explained in terms of the former, which is the comparatively more fundamental notion. Explicit defences of ‘three-place fundamentalism’ can be found in, e.g., Richard Holton, “Deciding to Trust, Coming to Believe,” *Australasian Journal of Philosophy* 72, no. 1 (1994): 63–76 and Karen Jones, “Trust as an Affective Attitude,” *Ethics* 107, no. 1 (1996): 4–25. For some recent resistance to this idea, see Jacopo Domenicucci and Richard Holton, “Trust as a Two-Place Relation,” *The Philosophy of Trust*, 2017, 149–60.

7 Hieronymi, “The Reasons of Trust.”

8 Frost-Arnold, “The Cognitive Attitude of Rational Trust.”

9 Jones, “Trust and Terror.”

10 “The Reasons of Trust.”

11 For further support of the idea that non-therapeutic trust requires belief that the trustee will prove trustworthy, see Jonathan E Adler, “Testimony, Trust, Knowing,” *The Journal of Philosophy* 91, no. 5 (1994): 264–75, Arnon Keren, “Trust and Belief: A Preemptive Reasons Account,” *Synthese* 191, no. 12 (2014): 2593–2615 and Benjamin McMyler, *Testimony, Trust, and Authority* (Oxford University Press USA, 2011). For criticism, see Jones, “Trust as an Affective Attitude.”, Carolyn McLeod, *Self-Trust and Reproductive Autonomy* (MIT Press, 2002), McGeer, “Trust, Hope and Empowerment.”, Paul Faulkner, “On Telling and Trusting,” *Mind* 116, no. 464 (2007): 875–902; Paul Faulkner, *Knowledge on Trust* (Oxford University Press, 2011); Judith Baker, “Trust and Rationality,” *Pacific Philosophical Quarterly* 68, no. 1 (1987): 1–13.

betrayal by entrusting something to someone without believing they'll¹² actually do what they've been entrusted to do. But this is not 'pure' trust.

In support of this way of thinking about therapeutic trust, Hieronymi offers the following case-pair involving the betrayal of a secret.

SECRETS: Consider two cases. In one, I fully believe you are trustworthy; in the other, I have doubts about your trustworthiness, but, for other reasons (perhaps to build trust in our relationship, perhaps because I think friends should trust one another, or perhaps simply because I have no better alternative), I decide to tell you my secret. Suppose that, in both cases, you spill the beans, and that you do so in the same circumstances, for the same reasons.¹³

According to Hieronymi, once we thus hold fixed both (i) the 'importance of the good entrusted';¹⁴ and (ii) 'the wrongness of the violation,'¹⁵ then:

[...] it seems plausible that one's degree of vulnerability to betrayal tracks one's degree of trusting belief ... further, this seems to be because, in the second case, there was less trust to betray.¹⁶

There are, however, two problems with this diagnosis of SECRETS. The first is that it's not at all clear that one's degree of vulnerability to betrayal really tracks one's degree of trusting belief, *even when* the importance of the good entrusted and the wrongness of the violation are held fixed. To see why, just suppose we run a variation on Hieronymi's SECRETS case-pair where, in the first case, my belief that you are trustworthy is full (stipulate: credence 1) but, at the same time, completely *irrational*. By Hieronymi's reasoning, the betrayal is more greater by degree in the first case simply because of the irrationally ratcheted up belief. But it's not. I might, due to having this strong albeit irrational credence that you are trustworthy, be even more inclined than otherwise to *think* that the betrayal is serious. But it wouldn't in fact be a worse betrayal simply *on account of* the ratcheted up irrational credence.

But suppose, for the sake of argument, we grant that one's degree of vulnerability to betrayal tracks one's degree of trusting belief. It is worth noting that

12 The singular pronouns 'they' and 'them' are used throughout whenever gender is unknown or irrelevant.

13 "The Reasons of Trust," 230.

14 230.

15 230.

16 230–1.

even if this were true, it could be explained without recourse to the idea that there is—in the therapeutic case—“less trust” to betray. For example, it might be that vulnerability to betrayal is one among various features of trust, and it’s a feature that lines up with (e.g., by closely tracking) trust’s doxastic component, whereas other features of trust (for example: whatever features makes it resilient to certain kinds of monitoring¹⁷) might track some *non-doxastic* (e.g., affective) component of trust.

If something like this were right, then we couldn’t move simply from the idea that one’s degree of vulnerability to betrayal tracks one’s degree of belief to the conclusion that there is less trust in cases with less belief. After all, such cases might feature more prominently some other (e.g., affective) aspect of trust, and in virtue of the presence—perhaps, surfeit—of which there is not, on the whole, “less trust.”

But Hieronymi has a second argument for relegating therapeutic trust the ‘impure’ category. The second argument has to do not with vulnerability but with the legitimacy of certain kinds of complaints. This second line of reasoning goes as follows. People can *legitimately complain* about not being trusted fully when they are trusted in the absence of belief, which occurs only when other people lack confidence in them but trust them nonetheless.¹⁸ For example, imagine the 16-year-old from our opening case saying: “But you don’t *really* trust me”, upon finding out that the rationale for the trust was largely trust-building, in the absence of a belief that they’d prove trustworthy. The felicitousness of such a complaint is, for Hieronymi, meant to support the idea that therapeutic trust is not pure or full-fledged trust.

This reasoning is also problematic, though, in so far as it’s supposed to motivate relegating therapeutic to a ‘second tier’. Just as the teenager could complain in this scenario, they could also felicitously praise or thank you for trusting them *despite* lacking confidence. “Wow, you trusted me without believing—you must have *really* trusted me!”¹⁹ This is not to say that Hieronymi’s example complaint is *not* felicitous, nor that praise or gratitude for trusting despite lacking confidence is any more felicitous than is complaining that one has trusted in the absence of belief. Rather, the point is that it is not clear that complaining about

17 For discussion on this point, see, e.g., Baier, “Trust and Antitrust.” and Jeremy Wanderer and Leo Townsend, “Is It Rational to Trust?” *Philosophy Compass* 8, no. 1 (2013): 1–14.

18 “The Reasons of Trust,” 230.

19 For a defence of the idea that trust’s most pure form involves the absence of belief, see Guido Mollering, *Trust: Reason, Routine, Reflexivity* (Emerald Group Publishing, 2006).

trust in the absence of belief is in any way *more* felicitous than praising or thanking a trustor who trusts one in the absence of it.²⁰

In sum, Hieronymi's arguments from vulnerability and complaint legitimacy don't give us good reason to think that the difference between non-therapeutic trust and therapeutic trust is a difference in 'purity' of trust.

3. Let's look now at an attempt—due to Karen Frost-Arnold²¹—to 'broaden' an account of trust so that it is wide enough to rule both in both varieties of trust. On Frost-Arnold's proposal, *A* trusts *B* to φ iff the proposition that *B* will φ is part of *A*'s 'adjusted cognitive background,'²² where one's adjusted cognitive background includes all and only those propositions that one *accepts* for the purposes of practical reasoning—where acceptance does not entail positive belief²³ (e.g., positive belief of the sort that is generally lacked in therapeutic cases, even if often present in non-therapeutic cases). This kind of 'unity' view does not relegate therapeutic trust to a second-tier, as Hieronymi's proposal does, but rather 'brings it in to the first tier'—e.g., by subsuming it within a wider account of trust *simpliciter*.

There are two main problems with Frost-Arnold's 'unity'-style approach. The first is that the acceptance requirement needn't be satisfied in all cases of therapeutic trust. Suppose you trust your teenager to drive your car for the weekend and return it safely. Suppose further that, upon doing this, you purchase additional insurance, just in case. By purchasing this additional insurance, you are not accepting the proposition that the teenager will return the car safely—viz., to do what you'd trusted them to do—in the course of your practical reasoning. You act instead on the proposition that they might realistically enough not do so.²⁴

20 Moreover, the felicitousness of such a reply gains some support in the literature on the psychology of gratitude see, e.g., Robert A Emmons and Michael E McCullough, *The Psychology of Gratitude* (Oxford University Press, 2004); for instance, gratitude is a predictable response by one to another who has placed a kind of 'faith' in their good will or competence.

21 "The Cognitive Attitude of Rational Trust."

22 1963–4.

23 The idea that accepting a proposition, understood as taking it for granted in one's practical deliberations—alternatively: acting 'as if' the proposition is true—does not entail that one believes the proposition to be true has been defended variously in epistemology, the philosophy of science and elsewhere. For some representative discussions of how belief and acceptance come apart, see L. Jonathan Cohen, "Belief and Acceptance," *Mind* 98, no. 391 (1989): 367–89, Michael E. Bratman, "Practical Reasoning and Acceptance in a Context," *Mind* 101, no. 401 (1992): 1–15, and Andrei Buckareff, "Acceptance Does Not Entail Belief," *International Journal of Philosophical Studies* 18, no. 2 (2010): 255–61.

24 This is the case, to be clear, even though we needn't suppose that you positively believe that the teenager *won't* return the car as entrusted.

But the existence of this mitigating back-up plan doesn't preclude the case from having been a case of therapeutic trust in the first place. That is, you don't suddenly cease to be therapeutically trusting the teenager with whom you aspire to build trust once you buy the insurance. It's not as though the vulnerability to betrayal one subjects oneself to is eliminated by one taking steps to mitigate damages against the risk occurring.²⁵

Some philosophers of trust have pressed back on this point. For example, Arnon Keren²⁶ holds that trusting involves declining to take precautions against the trustee's failing to come through.²⁷ This idea seems *prima facie* plausible in the epistemic case, specifically, where what the truster trusts the trustee to do is to tell them the truth. What 'taking precautions against the trustee's failing to come through' would amount to in this case would be finding additional evidence that bears on the truth of the proposition. But *then*, having sought sought such evidence, it doesn't look as though you are trusting the person's word at all.

To the extent that trusting (therapeutic or otherwise) *does* involve declining to take precautions against the trustee's coming through, this might very well be idiosyncratic to the epistemic case where there is a constitutive tension between relying on one's word and acquiring the kind of evidence one would acquire by taking precautions. Crucially, we find no such tension though outside the epistemic case, at least when we hold fixed that the precautions are (as in the

25 This point, it is worth noting, is compatible with the widely accepted idea that *monitoring* is incompatible with trusting, either of a non-therapeutic or therapeutic variety. See, e.g., Baier, "Trust and Antitrust," 260.

26 "Trust, Preemption, and Knowledge," *Trust in Epistemology*, 2019.

27 Keren, 121 actually formulates his position as follows, with the qualifier 'every': 'If you rely on a person to ϕ but take *every* precaution against the possibility that she might not ϕ —by seeking evidence that might indicate that she might fail to ϕ and by acting in order to minimize the harm caused in case she fails to ϕ —then you do not trust her to ϕ . You might rely on her to ϕ , but you do not trust her to do so' (my italics). As formulated, this is not controversial, as this is tantamount to the statement that trusting is incompatible with certain kinds of monitoring. What is at issue in the example I am discussing above, involving an insurance policy, is rather whether mitigating at all against the risks of the damage that would be incurred by the trustee's betrayal would be compatible with nonetheless therapeutically trusting that person. My contention that it is is thus compatible with granting Keren's point that some kinds of monitoring—i.e., such as those that involve taking every precaution against the possibility the trustee won't come through—are incompatible with trust (therapeutic or otherwise). That said, Keren also makes claims about taking precautions, in the specific case of epistemic trust, which appear to go beyond the statement of his view noted above, and which appear to imply that it is essential to epistemically trusting someone that you decline entirely from taking precautions against their not coming through. Because this thesis, at least if applied generally and not just in the epistemic case, is in tension with my assessment of the insurance policy case, I focus in the main text on it as opposed to on Keren's less contentious formulation quoted above.

example of the insurance policy) precautions that are solely designed to mitigate damages *if* the trustee does not come through. Compare: this is importantly different from, and does not imply, taking precautions designed to lower the likelihood that the trustee will *fail* to come through — as one might do by hiring a team to accompany the teenage driver. Accordingly, the attempt to reply to the objection raised to Frost-Arnold by way of appealing to Keren’s insights about epistemic trust looks to come up short.

The upshot is that Frost-Arnold’s unified account of trust which frames trust in terms of acceptance is still too narrow to do what she wants it to do, which is to rule in all cases of trust, non-therapeutic and therapeutic alike.

Even more, the proposal faces a second problem. The second problem concerns the evaluative normativity of trusting. Generally speaking, evaluative norms – unlike prescriptive norms, which prescribe conduct – regulate what it takes for a token of a particular type of thing to be good or bad with regard to its type, where the ‘goodness’ or ‘badness’ here is *attributive* in Geach’s²⁸ sense – viz., the sense in which a sharp knife is a good knife, *qua* knife, regardless of whether it is good or bad *simpliciter*. (Likewise, in this sense, a known belief is a good belief, regardless of whether it would be good or bad *simpliciter* – viz., as it would be were the content of the knowledge instructions for igniting a terrible bomb.²⁹) The worry is that the view lacks the resources to account for why reasonable therapeutic trust isn’t just *bad* as an instance of trusting.

Continuing with the teenager car case: let’s suppose you have no trust-building objectives in mind, and simply want someone dependable to drive your car for the weekend and bring it back safely. Foolishly, you choose the teenager with a record you know is patchy at best. This looks like *bad* trust, even if it would *not* be so with therapeutic purposes in play.³⁰ But it’s hard to see how we’d explain this normative difference on Frost-Arnold’s unity-style account. One might try to begin to tell such a story by appealing to the ‘epistemic constraint’ that Frost-Arnold places on the kind of acceptance that matters for a proposition’s being ruled-in the adjusted cognitive background. But the epistemic constraint she places on acceptance is really a very minimal one. It precludes just one thing: positive belief that the person will *not* do the thing in question. This kind of

28 “Good and Evil,” *Analysis* 17, no. 2 (1956): 33–42.

29 For a helpful overview of the prescriptive/evaluative norm distinction, with reference to attributive as opposed to predicative goodness, see Conor McHugh, “The Truth Norm of Belief,” *Pacific Philosophical Quarterly* 93, no. 1 (2012): 22 and, as this distinction applies to belief specifically, Mona Simion, Christoph Kelp, and Harmen Ghijsen, “Norms of Belief,” *Philosophical Issues* 26, no. 1 (2016): 384–6.

30 One sense in which the trust here is bad is that it is not likely to be *successful*, in that trusting a teenager involves incurring a relatively higher risk of betrayal than normal. For discussion on this point, see [Redacted].

constraint won't help us in any way to adjudicate the normative question we're interested in—viz., how to distinguish at least some cases of good therapeutic trust from plain old bad trust.

There's another thread to this point. Just as picking out an unreliable person is 'bad' trusting (likely to lead to one's trust being betrayed³¹) even if trusting that unreliable person could have been reasonable were therapeutic purposes suitably in play, it doesn't follow that *simply stipulating* a therapeutic purpose suffices to make any instance of therapeutic trust *good* therapeutic trust. One can surely be better or worse at therapeutic trusting, just as one could be better or worse at trusting more generally—and indeed, very plausibly in light of different kinds of skill sets. None of this looks explicable (at least, in any straightforward way) if we embrace a unity-style view like Frost-Arnold's.

4. The foregoing discussion suggests that what's needed is an account of therapeutic trust which explains clearly how it features some kind of *normative difference* with respect to ordinary, non-therapeutic trust. This is exactly what Karen Jones's³² account of therapeutic trust offers. Unfortunately, she identifies the *wrong kind* of normative difference.

According to Jones, therapeutic trust involves the normative attitude that the trustee *ought* to do what one trusts them to do, rather than optimism that they will do it. With reference to our opening case pair: when you trust the reliable friend to watch your house for the weekend, you are optimistic that they will do this as you've entrusted them. While you're not optimistic that the teenager will do the same when you trust them with the task, you nonetheless think in trusting them that they *ought* to do what you've entrusted them to do.

There are three main problems with this proposal. First, the normative attitude that the trustee ought to do what one trusts them to do is not necessary for therapeutic trust. Consider a case where a CEO, with the aim of striking up a romantic relationship with a low-level employee, entrusts that employee with an inappropriately enormous responsibility—hoping that doing so will help generate a trusting relationship between them as a precursor to such a romance. If the CEO is not blind to their exploitative reasons underlying the trust they are placing in this inexperienced employee, then they will not have the view that the trustee *ought* to actually do what they are entrusted to do. Quite the contrary, the CEO might well know that that the employee's succeeding in doing what they've been entrusted with is beyond reasonable expectations. But this is therapeutic

31 See [Redacted] for a defence of this way of thinking about bad trust.

32 "Trust and Terror."

trust nonetheless, in that it is—albeit for morally dubious reasons³³—aiming to bring about and strengthen a trust relationship.³⁴

A second objection to Jones's proposal—in so far as it purports to distinguish therapeutic from non-therapeutic trust—is that some cases of non-therapeutic trust involve not only optimism that the trustee will do what they are entrusted to do, but *also* the normative attitude that the trustee ought to do what they are entrusted to do. Suppose someone is drowning. I am nearby with a life vest—my expensive life vest—but my arm hurts, and so I can't throw it off the boat to help. I trust my able-bodied and reliable friend to throw it. In this case, where I trust my friend to throw my vest, the trust isn't therapeutic in any interesting sense.³⁵ And yet, I have the (strong) normative attitude that the trustee *ought* all-things-considered to do what I've entrusted them to do. Thus, believing that the trustee ought to do what one entrusts them to do isn't distinctive of therapeutic trust but not ordinary non-therapeutic trust.

A third objection to Jones's proposal is that some cases of therapeutic trust positively *do* involve optimism that the trustee will do what they are entrusted to do, even if this optimism persists along with some serious doubts. To see why, it will be instructive to first consider how optimism comes apart from belief in both directions. In the literature on the psychology of optimism,³⁶ an optimistic attitude, with respect to some situation X, is often characterised in terms of a

33 No assumption is being made, to be clear, that the norms of trust are moral norms. I have suggested (in responding to Frost-Arnold) that we should expect an account of therapeutic trust to be reconcilable with plausible claims about the evaluative normativity of trusting; but this commitment is a very general one—viz., to there being norms (however we best articulate them) that regulate what it takes for a token of trusting to be good or bad with regard to its type. This is at most a commitment to attributive (rather than predicative) goodness of trusting in certain cases.

34 The same kind of point can be made with reference to a more paradigmatic kind of 'teenager trust' case. Suppose a mafioso with a conscience but a weak will accepts a hit job and, rather than to do it himself, entrusts his unreliable teenager to carry out the hit—hoping that doing so will build trust. Assume the target of the hit is known by the mafioso to be completely innocent. It is entirely plausible here that the mafioso, bering this in mind, will appreciate that what he's entrusted the teenager to do is not something the teenager ought to do. Yet, this fact (as in the CEO case) does little to change the fact that the trust here is of a therapeutic variety. Granted, this—as well as the CEO case—relies on a weak assumption in moral psychology, which is that one can desire to bring about some state of affairs while acknowledging that its being brought about would violate one (or more) prescriptive norm. For discussion, see Michael Stocker, "Desiring the Bad: An Essay in Moral Psychology," *The Journal of Philosophy* 76, no. 12 (1979): 738–53.

35 As we'll see in §7, though, there is plausibly an uninteresting kind of 'default' therapeutic trust in play here that is implicated by most cases of non-therapeutic trust.

36 E.g., Charles S. Carver, Michael F. Scheier, and Suzanne C. Segerstrom, "Optimism," *Clinical Psychology Review* 30, no. 7 (2010): 879–89.

kind of attention profile directed at favourable features of that situation. For example, if my car breaks down and I'm stranded on a highway, then an optimistic attitude might lead me to focus my attention on how doing certain things under my control (e.g., walking to the nearest petrol station) could better my situation.

Coming back to therapeutic trust: one can distribute one's attention patterns in ways that line up with optimism (with respect to a trustee proving trustworthy) *without* having any positive belief that the trustee will prove trustworthy. (Compare: I can be optimistic when stranded without actually having the *belief* that I will be saved). For example, being optimistic that the teenager will look after the house properly or return the car might involve focusing on the teenager's good traits, feeling pride in remembering past times they've exceeded expectations, etc. This is all compatible with a lack of belief that they will in fact bring the car back. Note, furthermore, that belief and optimism come apart in the other direction as well. You could believe someone will bring a car back without being optimistic simply because your attentional profile does not line up with what you believe. You might be irrationally paranoid, given a pessimistic perspective that does not line up with your belief that the trustee will prove trustworthy. These considerations in favour of the idea that optimism can float freely of one's doxastic attitudes supports that (*contra* Jones) the kind of doubts one has in the case of therapeutic trust aren't doubts that, as such, would preclude optimism that the trustee will prove trustworthy.

In sum, Jones is mistaken that therapeutic trust involves the normative attitude that the trustee ought to do what one trusts them to do, rather than optimism that they will do it. This is because it neither requires the normative attitude that the trustee ought to do what one trusts them to do—as per the CEO case—*nor* does it preclude optimism that they will do what they are entrusted to do, at least in so far as optimistic attitudes are plausibly demarcated by their attentional profiles.

5. So far, we've seen that prominent extant accounts of therapeutic trust run into various kinds of problems. A presupposition common to each of the three views considered is that therapeutic trust is a univocal kind, and this is a presupposition we'd be better off rejecting.

There are two importantly different species of therapeutic trust—*default therapeutic trust* and *overriding therapeutic trust*. Each species of therapeutic trust interacts with ordinary (non-therapeutic) trust differently. And, each is *normatively constrained* differently from each other. Appreciating how this is so, we can—in addition to avoiding the kinds of problems considered—make sense of something other views can't, which is what makes therapeutic trust (of a philosophically interesting sort) *good* when it is.

6. Any kind of performance with an aim internal to that performance-type can be evaluated along three dimensions: it can be evaluated for *accuracy* (i.e., did the performance succeed in attaining its aim), *adroitness* (i.e., was the performance skilful), and *aptness* (i.e., was the success *because* of the skill?).³⁷ Take, for example, archery. An archer’s shot is accurate if and only if it hits the target.³⁸ Regardless of whether it hits the target, it might be adroit—viz., it might be fired in a manner that would usually (enough³⁹) result in a successful shot, in normal conditions, and regardless of whether it in fact did. If the shot is both accurate and adroit—and further, if the accuracy is *because* of the adroitness—then the shot is *apt*.⁴⁰

A key insight of recent virtue epistemology is that belief is a kind of performance with an aim, the aim of truth.⁴¹ With reference to the ‘AAA’ model, we can ask three distinct questions for any given belief: is it true (accurate), is it

37 This three-fold model for evaluating performances is due to Ernest Sosa e.g., Ernest Sosa, *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1* (Oxford University Press, 2007); Ernest Sosa, “How Competence Matters in Epistemology,” *Philosophical Perspectives* 24, no. 1 (2010): 465–75; Ernest Sosa, *Judgment and Agency* (Oxford: Oxford University Press, 2015). For other applications of this ‘AAA’ model of performance evaluation, see for example, John Turri, “Knowledge as Achievement, More or Less,” *The Present and Future of Virtue Epistemology*, 2016, [Recacted], Jesper Kallestrup, “Group Virtue Epistemology,” *Synthese*, 2016, 1–19, Christoph Kelp et al., “Hoops and Barns: A New Dilemma for Sosa,” *Synthese*, 2017, 1–16, and (ed.) Miguel Ángel Fernández Vargas, *Performance Epistemology: Foundations and Applications* (Oxford University Press, 2016). For criticism, see Matthew Chrisman, “The Normative Evaluation of Belief and the Aspectual Classification of Belief and Knowledge Attributions,” *Journal of Philosophy* 109, no. 10 (2012): 588–612.

38 For an initial use of this analogy in order to illustrate performance normativity, see Sosa, *A Virtue Epistemology*.

39 The threshold for reliability that is required for adroitness differs across performance domains. For example, to be adroit at shooting free-throws, one might require at least 70% reliability. The very best baseball-hitting averages, however, are barely over 30%. For discussion, see Ernest Sosa, *Epistemology* (Princeton University Press, 2017), 75 and John Greco, *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity* (Cambridge University Press, 2010), 77–78.

40 Note that the conjunction of a shot’s being accurate and adroit is compatible with it’s failing to be apt. In the archery case, this might occur if a shot is fired adroitly while a chance gust of wind (i) first blows the arrow off target; and then (ii) a second gust of wind blows the arrow back on target, guiding it to the bullseye. In such a case, the accuracy + adroitness conjunction falls short of aptness because the accuracy is not *because* of the adroitness, but because of the second gust of wind. For discussion of cases with this kind of structure, see, along with Sosa, *A Virtue Epistemology*, also John Greco, *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity* (Cambridge University Press, 2010) and Duncan Pritchard, “Anti-Luck Virtue Epistemology,” *Journal of Philosophy* 109, no. 3 (2012): 247–79.

41 The canonical presentation of this idea is due to Sosa, *A Virtue Epistemology*. See also the essays in Fernández Vargas, *Performance Epistemology*. For a revisionist ‘knowledge-first’ approach to this model, which replaces truth with knowledge as the aim of belief, see, e.g., Christoph Kelp, “Knowledge First Virtue Epistemology,” in *Knowledge First: Approaches in Epistemology and Mind*,

formed in a manner that would reliably enough attain the truth aim (e.g., is it adroit), and thirdly, is the belief *apt*—viz., is it accurate (i.e., true) *because* of its adroitness.

Conveniently, non-therapeutic trust—as a kind of performance—can be subsumed within this kind of ‘AAA’ template just as belief can.⁴² By placing it within the template, it will become clearer how we should think about therapeutic trust, or so I’ll argue. In order to get things off the ground, though, we need clarify the sense in which there is an aim internal to trusting in the standard non-therapeutic case. Such an aim, I submit, falls out of a more or less generic characterisation of (non-therapeutic) trust as an attitude.

The idea is as follows: just as an archer’s shot attains its aim if and only if it hits the target, and a belief succeeds in attaining its aim if and only if true, trust succeeds in attaining its aim if and only if *the trustee takes care of things as entrusted*.⁴³ If the trustee does not take care of things as entrusted, then that trust is not successful *even if* the trust is adroit—viz., even if one trusts in ways (e.g., by seeking out a reliable trustee, assessing risks of betrayal, etc.) that *ordinarily* would lead to successful trust, but was just unlucky on this occasion.⁴⁴

ed. J. Adam Carter, Emma C. Gordon, and Benjamin Jarvis (Oxford University Press, 2017), 223–45 and Lisa Miracchi, “Competence to Know,” *Philosophical Studies* 172, no. 1 (2015): 29–56.

42 For a development of this idea, see [Redacted].

43 This phrasing is intentionally compatible with different kinds of substantive glosses. For example, according to Baier, “Trust and Antitrust,” 234, taking care of things ‘as entrusted’ will include at least that the trustee do so out of goodwill toward the trustor. For criticism of this goodwill caveat, see Holton, “Deciding to Trust, Coming to Believe,” 65. Alternatively, for a very different way to gloss the idea of taking care of things as entrusted, consider Katherine Hawley, “Trust, Distrust and Commitment,” 10 view that the trustee must be *believed to have* a commitment such that she is relied on to meet that commitment. For a more recent development of this idea, see *How to Be Trustworthy* (Oxford University Press, USA, 2019).

44 One might suggest that additional clarity is needed about what “taking care of things as entrusted” involves. As Baier, “Trust and Antitrust.” points out in “Trust and Anti-Trust”, it is often ambiguous just what one person trusts another for (do we trust the babysitter, however well-intentioned, not to “tidy” our kitchen so we can’t find anything anymore?). Rather than to attempt to clarify “taking care of things as entrusted” in a way to avoid any ambiguity, we should take Baier’s point at face value; sometimes there is some ambiguity, perhaps vagueness, in what the precise object of trust is, especially when we ask fine-grained questions. This is not any different than what we find when thinking about how to delineate ‘success’ for other kinds of performances. Even in the archery case – ‘hit the target’ might be ambiguous between hitting anywhere in the bullseye, just the very centre of the bullseye, etc. For my own purposes, we should follow Baier’s lead and accept that we might not always know the sharp boundaries of the object of trust, and thus, what is required to take care of things as entrusted in a particular case. However, just because we needn’t know these sharp boundaries does *not* mean whether the trustee has taken care of things as entrusted isn’t obvious enough most of the time. As a final point of clarification: the locution *as entrusted* is neutral with respect to whether ‘taking care of things’ in a paradigmatic trust exchange will involve the trustee’s doing so out of goodwill e.g.,

The ‘AAA’ model of performance assessment—extended to ordinary (non-therapeutic) trust—takes on the following template shape:

	Accuracy/success	Adroit/skilful	Apt
<i>archery shot</i>	hit target	shot issued from archery competence	shot successful (hit target) because of (archery) competence
<i>belief</i>	true	belief issued from epistemic competence	belief successful (true) because of epistemic competence
<i>trust</i>	trustee takes care of things as entrusted	trust issued from trust competence	trust successful because of trust competence

Throughout this analogy, adroitness is understood in terms of *competence*, which is an important concept in performance-theoretic evaluations. Before zeroing in on therapeutic trust specifically, it is worth clarifying a point about the structure of competences generally.

An archery competence is a disposition to reliably enough hit the archery-relevant aim (i.e., the target) when one tries. But there is an important qualification here, which is that—just as dispositions like flammability are indexed to manifestation conditions⁴⁵—competences are indexed to performance conditions (i.e., shape and situation) which are the pertinent ones to that performance

Jones, “Trust as an Affective Attitude.” I’m remaining neutral for my own purposes on whether the trustee’s taking care of things as entrusted requires this, and so the model I’m advancing here could be accepted by philosophers of trust who both embrace (and reject) such an attitudinal requirement on the part of the trustee in order to do what they are entrusted to do as entrusted.

⁴⁵ For instance, flammability is a disposition a match has if and only if the match would struck if lit *under certain normal conditions*, which include, e.g., being suitably dry, being struck in the presence of ambient oxygen, etc. For discussion, see Sosa, “How Competence Matters in Epistemology,” 466. Note that the indexing of dispositions (and by extension, competences) to manifestation conditions fits snugly conditional approaches to analysing dispositions, but doesn’t rely on any such commitment. For classical approaches to indexing dispositions to manifestation conditions, see, e.g., Gilbert Ryle, “The Concept of Mind,” *London: Hutchinson*, 1949, Nelson Goodman, *Fact, Fiction, and Forecast* (Harvard University Press, 1954), Willard Van Orman Quine, *Word and Object* (MIT press, 1960). For an overview, see Jennifer McKittrick, *Dispositional Pluralism* (Oxford University Press, 2018) and Sungho Choi and Michael Fara, “Dispositions,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2018 (<https://plato.stanford.edu/archives/fall2018/entries/dispositions/>; Metaphysics Research Lab, Stanford University, 2018).

type,⁴⁶ viz., the ones in which reliability is valued. It doesn't count against your archery competence, for instance, if you would reliably *miss* the target when releasing the arrow while drugged or placed in unusually high winds. Likewise, it doesn't count against a competence to trust successfully (reliably enough) if, too often, your trust would be betrayed when in improper shape (e.g., mentally incapacitated prior to selecting whom to trust with something) or improperly situated (e.g., unbeknownst to you, placed in an abnormal social-epistemic environment, viz., a city of liars, or a deception-driven flash mob).⁴⁷

7. The previous section has shown how we can situate paradigmatic non-therapeutic trust within the performance-theoretic framework that is familiar in other areas of philosophy. With non-therapeutic trust repositioned in this way, we now have a new vantage point to theorise about *therapeutic* trust—and in particular, about what I'll call *overriding* therapeutic trust—which is the most philosophically interesting variety of therapeutic trust.

But first it is worth making the following explicit: there is a kind of *default* therapeutic trust—viz., therapeutic insofar as it (trivially) aims at trust-building—that is *implicit* in paradigmatic cases where one trusts with the aim that the trustee take care of things as entrusted. We take for granted in trusting that trust will—apart from whatever else it does—play its normal social functions, functions that plausibly include the social function of strengthening trust relations.⁴⁸ This is so even when we trust—as we do when we seek out someone reliable and trustworthy—with the basic aim that the trustee take care of things as entrusted.

In this respect, there is a minimal and trivial kind of therapeutic trust that is going to be implicit in garden variety (non-therapeutic) trusting, and this is so even in the absence of any explicit *intention*—the kind of intention that is explicit in teenager-style cases—to satisfy the aim of building trustworthiness.⁴⁹

46 See Sosa, “How Competence Matters in Epistemology,” 466–7 and *Epistemology*, 195.

47 For a detailed discussion of the structure of competences, see, along with Sosa, “How Competence Matters in Epistemology,” and also Sosa, *Epistemology*, 191–2.

48 For some representative defences of the role of trusting in trust-building, see Faulkner, *Knowledge on Trust* Ch. 1, Mark Alfano, “The Topology of Communities of Trust,” *Russian Sociological Review* 15, no. 4 (2016): 30–56, Mark A. Hall, “The Importance of Trust for Ethics, Law, and Public Policy,” *Cambridge Quarterly of Healthcare Ethics* 14, no. 2 (2005): 156–67, and Robert C. Solomon and Fernando Flores, *Building Trust: In Business, Politics, Relationships, and Life* (Oxford University Press USA, 2003).

49 Consider, by way of analogy, one of the plausible social-epistemic functions of *assertion*, which is to generate knowledge in the hearer e.g., Christoph Kelp, “Assertion: A Function First Account,” *Noûs* 52, no. 2 (2018): 411–42; Mona Simion, “Knowledge-First Functionalism,” *Philosophical Issues* 29, no. 1 (2019): 254–67; cf., Timothy Williamson, *Knowledge and Its Limits* (Oxford: Ox-

Moreover, the ‘implicit’ kind of therapeutic trust that accompanies ordinary trust as a default does not have its own constitutive aim. This is because default therapeutic trust is just *implicated by* normal, non-therapeutic trusting, which itself constitutively aims at the trustee’s taking care of things as entrusted. In this respect, the competences that are relevant—trivially—to default therapeutic trust are just those that matter for non-therapeutic trust that implicates it.

8. The most interesting kind of therapeutic trust is not default therapeutic trust, but *overriding therapeutic trust*. This occurs when, as in our paradigmatic teenager cases, the aim of successful trust—given perceived vulnerabilities—isn’t itself enough to *motivate* one to risk trusting. Simply wanting your house to be watched over responsibly wouldn’t, from the perspective of a trustor, favour entrusting such a task to the teenager, as opposed to someone regarded to be more reliable; rather, the opposite would be the case. Necessary for bringing about overriding therapeutic trust is thus an ‘overriding’ and intentional aim—the aim of building or strengthening trust—that is distinct from the constitutive aim of ordinary non-therapeutic trust.

Unlike default therapeutic trust that is implicit in most normal trusting, *overriding* therapeutic trust is a distinct kind of performance from normal (non-therapeutic) trust, with its own constituent normativity. The constitutive aim of overriding therapeutic trust is not *merely* to trust successfully (*viz.*, that the trustee take care of things as entrusted). But nor, it should be emphasised, is it *merely* to build trust. It is to *build trust through trusting successfully*.

Consider that just as ordinary (non-therapeutic) trust is defective when it misses its internal aim (that the trustee take care of things as entrusted), your choosing to trust your teenager to watch over the house has missed *its* mark if either (i) trust is not built (e.g., if a result of this trusting is not a strengthened trust relationship) *or* if (ii) trust is not successful (e.g., if the teenager throws a party, during which items from the house are stolen). *Even more*, though, your trust will have missed its mark even if the trust serves to build trust *and* the trust is successful, but (iii) if the trust built is not built *through* the successful trust, but for some reason disconnected with the therapeutic trust placed in them. This might be the case, for example, if the teenager watches over the house

ford University Press, 2000). On the assumption that assertion has such an aim, constitutively, it’s easy to see how asserters *implicitly*, in asserting, aim at other things (even if not intentionally), namely, whatever generating knowledge in a hearer generally involves, including playing roles that knowledge normally plays for the hearer. For example, one role that knowledge plausibly plays for a hearer who acquires it is that of being a possible premise in the hearer’s practical reasoning e.g., John Hawthorne and Jason Stanley, “Knowledge and Action,” *Journal of Philosophy* 105, no. 10 (2008): 571–90.

successfully, though—unaffected entirely by the trust you’ve placed in them—comes to trust you more nonetheless due to having, while watching over the house, spent some time reading false accounts of sacrifices you’ve made for them in the past, and only on this basis, develops toward you a stronger bond of trust.

Question: if overriding therapeutic trust constitutively aims not at mere successful trust, nor at the mere building of trust, but at building trust *through* successful trust, then what do (i) adroit and (ii) apt overriding therapeutic trust consist in? Put another way: when we add overriding therapeutic trust and its constitutive aim to our ‘AAA’ template, how should we fill the rest in?

	Accuracy/success	Adroit/skilful	Apt
<i>archery shot</i>	hit target	shot issued from archery competence	shot successful (hit target) because of (archery) competence
<i>belief</i>	true	belief issued from epistemic competence	belief successful (true) because of epistemic competence
<i>trust</i>	trustee takes care of things as entrusted	trust issued from trust competence	trust successful because of trust competence
<i>trust</i> _{O.T.}	trust built through successful trust	?	?

Adroit overriding therapeutic trust, on the ‘AAA’ model, will derive from a competence to attain the constitutive aim of *overriding therapeutic trust* reliably, which is the aim of building trust *through* successful trust. Given that competences are indexed to performance conditions, a clear view of the kind of competence that matters for overriding therapeutic trust requires an understanding of the conditions under which reliable performance matters for this *particular* kind of trusting. These conditions include (at least) the satisfaction of what I’ll call an *openness condition* and a *reciprocity condition*.

To appreciate the former condition, consider the following case:

DIANE: You need someone to babysit on short notice. There are a number of people you could ask, however, you choose a local teenager, Diane, whose parents you know. You have heard that Diane is troubled, and you have had a standing desire to take Diane under your wing in hopes of having a positive influence

on her. A first step toward having such a positive influence, you think, will be to establish a bond of trust, a bond you hope to develop by entrusting her with the babysitting task despite her reputation. Unfortunately, and unbeknownst to you, Diane recently experienced a highly traumatic event, to which she has responded by closing off the possibility of developing a trusting relationship with *anyone*, at least until she has worked through this trauma. She succeeds in the task of babysitting, though at no point was she in a position where her being entrusted with this would have changed her distrusting stance of others.

In DIANE, the conditions for successful overriding therapeutic trust are simply not in place *ex ante*—and this is so even though the conditions in DIANE do not preclude her in any way from taking care of things as entrusted. With respect to the aim you have of building trust through successful trust, Diane is ‘closed’. She is not in a position where your trusting her with the task that you do could—even when that trust is fulfilled by her—contribute to building trust on account of that fulfilment.

Now consider a twist on this case:

DIANE*: You need someone to babysit on short notice. There are a number of people you could ask, however, you choose a local teenager, Diane*, whose parents you know. You have heard that Diane* is troubled, and you have had a standing desire to take Diane* under your wing in hopes of having a positive influence on her. A first step toward having such a positive influence, you think, will be to establish a bond of trust, a bond you hope to develop by entrusting her with the babysitting task despite her reputation. Diane* is open in principle to building trust with someone who would entrust her with this kind of task. Unfortunately, and unbeknownst to you, Diane* bears a deep-seated grudge against you. Though she babysits the kids successfully (suppose, she needs the money)—and though her doing so successfully in fact contributes to making her more trustworthy *generally* speaking—it plays no role in establishing or strengthening any trust between you and her.

Diane* is *not* closed to building trust through successful trust, as Diane is, generally. However, the conditions in DIANE* are such that they prevent building her trust *with you* through successful trust. This is not to say that Diane*'s grudge

would never subside so as to open up such a possibility later. The point is that the situation in which you encounter Diane* is not one in which, were she to come to establish and build trust with you, this could be achieved in the way you're attempting to do so here—viz., through facilitating successful trust via the babysitting task.

There are two interrelated points to draw from the DIANE and DIANE* cases. The first is that it doesn't count against one's overriding therapeutic trust competence, viz., one's disposition to attain the aim of overriding therapeutic trust reliably enough, were one to be *unreliable* at attaining this aim in cases like DIANE or DIANE*, where the conditions are, for different reasons, not suitably conducive to building trust through successful trust. Secondly, and relatedly: the kind of competence that *matters* for overriding therapeutic trust is, accordingly, a disposition to build trust through successful trust reliably enough when one is in conditions that *are* appropriate to doing so, conditions that include at least that openness and reciprocity are satisfied, as they are not in DIANE and DIANE*, respectively.

A further point is that *when* these conditions are met, some are disposed to achieve the aim of overriding therapeutic trust *more* reliably than others. And that is just to say that, when it comes to overriding therapeutic trust, some are *more competent* than others, some of whom simply lack this competence by not being suitably reliable in conditions that are favourable to this kind of trust.

What makes the difference? One factor that's worth noting explicitly is that we vary in the capacities we have to reliably assess trust-building payoffs. For example, recall our case of the CEO (§4) who entrusted the low-level employee with a *disproportionately* large task, one which not easily the employee would have managed. The overriding therapeutic trust is unlikely to payoff here simply given that the difficulty of the task choice will make unlikely the building trust through *successful* trust. Conversely, entrusting *too small* a task, with therapeutic aims, is likewise unlikely to payoff, though, for a different reason. (Compare: suppose you were to, with trust-building aims, entrust a teenager not with looking over the house or the kids, but with looking after a small cactus for the weekend). The task is not certainly too difficult to undermine the likelihood that the trustee will take care of things *as entrusted*, but it is *so easy* that it undermines the likelihood that, through being undertaken successfully, it will play a (non-negligible) role in increasing any kind of trust bond with the trustee. In short: (i) a propensity to miss the mark too often in *either* direction will undermine one's reliability at attaining the aim of building trust through successful trust, and so (ii) a competence to hit this aim reliably (when appropriately situated

to do so⁵⁰) requires a capacity for the kind of risk assessment that's needed to prevent one from too often 'over' or 'under' trusting (as in the CEO and cactus cases, respectively).

Let's return now to our 'AAA' table. *Accurate* or successful overriding therapeutic trust (i.e., attaining the first 'A') occurs when overriding therapeutic trust hits its constitutive aim, which is the aim of building trust through successful trust. *Adroit* overriding therapeutic trust issues from a *competence* to hit this aim reliably enough when one trusts with a therapeutic aim whilst appropriately situated—where being appropriately situated for this kind of trust requires at least the satisfaction of the openness and reciprocity conditions. *Apt* overriding therapeutic trust can now be defined in terms of accurate and adroit therapeutic trust—viz., apt overriding therapeutic trust is overriding therapeutic trust that is *accurate because adroit*, viz., when one's building trust through trusting successfully manifests one's competence to therapeutically trust successfully reliably enough in appropriate conditions.

Apt overriding therapeutic trust is a kind of *achievement*, just like any kind of aim attained through skill rather than by other means.⁵¹ In this respect, apt overriding therapeutic trust stands to *mere* successful overriding trust as knowledge to lucky true belief, and to an archer's successful shot attained through skill to the same success attained any old way. However, as we've seen, the achievement of apt overriding therapeutic trust is a *different* achievement than the achievement of apt (non-therapeutic) trust, one that involves the attaining of a different aim through the manifestation of a different sort of competence.

9. We began with a puzzle about therapeutic trust and its relationship to ordinary non-therapeutic trust. Three prominent attempts to address this puzzle were considered, and each was shown to be problematic for different reasons.

50 That is: when the trust environment is such that the openness and reciprocity conditions described in this section are met.

51 For some representative discussions of the value of achievements understood as having a success-through-ability structure, see, e.g., Gwen Bradford, "The Value of Achievements," *Pacific Philosophical Quarterly* 94, no. 2 (2013): 204–24; *Achievement* (Oxford University Press, USA, 2015); Ernest Sosa, "Value Matters in Epistemology," *The Journal of Philosophy* 107, no. 4 (2010): 167–90; Duncan Pritchard, John Turri, and J. Adam Carter, "The Value of Knowledge," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2018 (<https://plato.stanford.edu/archives/spr2018/entries/knowledge-value/>, 2018); John Greco, "Episteme: Knowledge and Understanding," *Virtues and Their Vices*, 2014, 285–302; [Redacted]; Duncan Pritchard, "Knowledge, Understanding and Epistemic Value," *Royal Institute of Philosophy Supplement* 64 (2009): 19–43 and Linda Trinkaus Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge University Press, 1996).

One notable problem common to each of the three accounts was that none was well-suited to explain—given what each maintains, respectively, about therapeutic trust and how it differs from non-therapeutic trust—in virtue of *what* good therapeutic trust differs from plain old bad trust, including incompetent trust that just so happens to result in the building of trust, as well as successful *and* competent trust that builds trust for reasons that have nothing to do with the trust placed.

The account I've proposed has a number of advantages over these accounts. First, it avoids the traps that these other accounts were shown to fall into given their specific commitments. The key move proposed which helps to get things right involves the recognition of two kinds of therapeutic trust. There is a philosophically uninteresting species of therapeutic trust that is implicit in ordinary trusting—what I called *default therapeutic trust*. While default therapeutic trust (trivially) aims at building trust, it does so only because building trust is among the normal social functions of ordinary non-therapeutic trust, which aims constitutively at the trustee taking care of things as entrusted. *Overriding therapeutic trust*, by contrast, has its own constituent normativity—with reference to which we can normatively assess this kind of trust differently from how we normatively assess standard trust. In doing so, we can say *why* each kind of trust is good when it is, without reducing the goodness of either kind of trust to the goodness of the other. Moreover, the view can help us to make sense of how the skills needed for reliable therapeutic trust come apart from the skills needed to be good at trusting well more generally; the 'AAA' profiles of adroit trust and adroit therapeutic trust differ in clear ways. Finally, by distinguishing between ordinary apt trust and apt (overriding) therapeutic trust on the model proposed, we have a perspective from which to appreciate two different *achievements* in trusting and why neither of these achievements reduces to the other.

10.

(Objection 1). On the view proposed, the constitutive aim of overriding therapeutic trust is meant to be distinct from the constitutive aim of standard (i.e., non-therapeutic) trust in that: (i) the aim of the former is that the trustee take care of things as entrusted; whereas, (ii) the aim of the latter is to build (or strengthen) trust *through* successful trust, viz., through the trustee's taking care of things as entrusted.

However, the suggestion that these aims are distinct is not so clear given that the view *also* holds that building trust is among the normal social functions that is played by (successful) ordinary trust. But if *that's* right, then isn't it the

case that standard trust constitutively aims not *merely* at the trustee's taking care of things as entrusted, but also, at this fact playing the social function of strengthening trust? If so, then it looks like the claimed difference between the constitutive aims of standard trust and overriding therapeutic trust collapses.

Reply: The fact that the constitutive aim of ordinary trust—viz., that the trustee take care of things as entrusted—is such that when this aim is met, it's doing so has a characteristic social function, X, does not imply that its actually playing that function, X, is thereby included *as part of the constitutive aim*. The aim would still be met even if that social function characteristic of attaining that aim were *not* played.⁵² (Compare: the aim of archery—hitting the target—is attained even if your hitting the target does not play any of the roles that attaining this aim would characteristically play, e.g., to build confidence, solidify social standing with peers, signal competence, etc.). Likewise, if you trust a reliable colleague to deliver an envelope to your boss without reading the message inside, and the colleague successfully does so without taking a peek, there is a clear sense in which your trust placed in your colleague on this occasion has attained *its* aim—no matter what further social functions your trust plays or does not play, including social functions you might reasonably *expect* it to play.

(Objection 2). The competences involved in ordinary (non-therapeutic) trust and overriding therapeutic trust are claimed to be *different* competences. But is this really so? Here is a reason to think the answer is 'no'. Adroit overriding therapeutic trust issues from a competence to reliably enough build trust through successful trust when one attempts to do so while appropriately situated. But then—being reliable at *this* was said to require a capacity for the kind of risk assessment that's needed to prevent one from too often 'over' or 'under' trusting (as in the CEO and cactus cases, respectively). But even if that's right—and here's the worry—doesn't being reliable at attaining the aim of *ordinary* trust *also* re-

52 There is a precedent for this kind of thinking about aims and defective functioning found in Tyler Burge, "Perceptual Entitlement," *Philosophy and Phenomenological Research* 67, no. 3 (2003): 509. According to Burge, evidence that something is or is not operating defectively offers us insight into what its aim (or, for Burge, function) is (or is not). For example, if we did not regard the heart as defective if it failed to pump blood, then this would cast doubt on the idea that the heart is normatively constrained by the aim of pumping blood. By parity of reasoning: my suggestion is that—in both archery and in ordinary trust—we would not regard a shot as defective if it hit the target but did not inspire confidence (a normal social function which, suppose, hitting a target plays) nor (ordinary) trust as defective if the trustee took care of things as entrusted despite this fact not going on to build further trust. This—with reference to the kind of reasoning we find in Burge—counts against the aim of archery being 'hitting the target *and inspiring confidence*', which is surely the right result, and likewise against the aim of ordinary trust as being 'that the trustee take care of things as entrusted *in a manner than builds trust*'.

quire a capacity for risk assessment? That is: a competence to attain the aim of ordinary trust reliably enough (when appropriately situated) surely requires a capacity to evaluate risks of *betrayal*, including risks of betrayal generated by, e.g., incentives the trustee has to betray, the difficulty of the task relative to the trustee's perceived abilities, etc. But once these points are granted, the distinction between the substance of the competences relevant to (i) ordinary trust versus (ii) overriding therapeutic becomes blurred.

Reply: In short, the answer is the kind of competence that matters for overriding therapeutic trust *asymmetrically entails* the kind of competence that matters for ordinary (non-therapeutic trust). While risk assessment is undeniably important to both kinds of competences, and thus to both adroit overriding therapeutic trust as well as adroit ordinary trust, the kind of risk assessment that competent overriding therapeutic demands is more sophisticated, and accordingly more demanding, than the kind of risk assessment that competent ordinary trust demands. Given that the constitutive aim of ordinary therapeutic trust (that the trustee take care of things as entrusted) is a *component* of the constitutive aim of overriding therapeutic trust (that trust is built through successful trust—viz., through the trustee taking care of things as entrusted), reliably attaining the latter will require the very same kind of risk assessment needed to reliably secure the former, *plus* the capacity to assess *additional* risks—risks specifically to the non-obtaining of *trust built through successful trust*. Ordinary trust competence doesn't demand one have the capacity to assess these further risks.

The above illuminates an interesting wider point about the difference between ordinary and overriding therapeutic trust, which is that the latter is—in short—*more difficult* to do well. Being competent at overriding therapeutic trust requires all the skills required to be competent at ordinary trust, plus others which the latter doesn't require. A corollary is that the achievement of *apt* overriding therapeutic trust is more substantial, arguably more valuable⁵³, than the achievement of *apt* ordinary trust, in that the former issues from a comparatively more sophisticated and demanding kind of competence to acquire and exercise.

(Objection 3). Certain kinds of 'forced-choice' cases seem like they would work as counterexamples to the proposed account of overriding therapeutic trust. Consider the following:

FORCED CHOICE: You've just moved to an apartment building in a new city, where the only person you know is teenager who lives

53 For some notable arguments that difficulty adds value to achievement, see, e.g., Bradford, "The Value of Achievements.", *Achievement*, and Duncan Pritchard, "Knowledge and Understanding," in *Virtue Epistemology Naturalized*, 2014, 315–27.

in the flat below you. You need to leave town for the weekend—suppose your job depends on it—and need someone to feed, water and walk your dog (you’ve tried kennels, etc., and all are fully booked). Your hand forced, you trust the teenager who lives in the flat below you with this task—someone whom, had you had a better range of options—you wouldn’t have chosen, as they’ve not established any track-record yet of responsibility with you, and your dog’s welfare is important to you.

Two things seem, *prima facie*, to be true in FORCED CHOICE. First, (i) it looks like a case of therapeutic trust of a philosophically interesting sort (you are, after all, placing trust in a teenager to whom you wouldn’t ordinarily trust a task like this); but, second, (ii) it doesn’t get ruled in on the account proposed. This is because in FORCED CHOICE, it’s *not* the case that the aim of ordinary trust is *not* sufficient to lead the truster to risk trusting. That aim *is* sufficient, *ex hypothesi*.

Reply: My response to FORCED CHOICE is to accept (ii) and press back against (i). It is a mistake to think that all cases in which one trusts a non-ideally suited trustee (e.g., by selecting someone regarded as being less reliable than would be preferred) are, in virtue of this, ‘therapeutic’ in some interesting sense. On the view I’ve proposed, therapeutic trust of a philosophically interesting sort misses its mark—viz., is defective—even if the trust is successful, provided the trust fails to build or strengthen through this successful trust. FORCED CHOICE, however, is a case where the trust placed in the teenager *succeeds perfectly well* so long as the the teenager takes care of the dog as entrusted. This is so in a way that is not interestingly different than were the teenager perceived to have been much more reliable than they are actually perceived to be. The situation is, however, very different if we suppose that the basic aim of successful trust *weren’t* enough (as it is in FORCED CHOICE) to motivate you to risk trusting the teenager—viz., as would be the case when you trust the teenager, e.g., rather than someone you think has a better track record, with the aim of using successful trust to building trust. The above diagnosis not only explains why we would be wrong to, in short, lump all high-risk cases together—but it also helps to highlight the important sense in which therapeutic trust of the philosophically interesting kind *uses* trust in a way that ordinary trust does not.