Philosophy Compass 8/1 (2013): 1–14, 10.1111/j.1747-9991.2012.00533.x

Is it Rational to Trust?

Jeremy Wanderer and Leo Townsend*
University of Cape Town

Abstract

It is common in popular depictions to portray the attitude of trusting and the norms associated with rationality as standing in some kind of tension. In this article, we suggest a way of capturing this tension, and explore some recent attempts at resolving it.

1. Setting Out the Tension

In popular discourse, trust and rationality are often portrayed as standing in some kind of tension. Consider the case of a paranoid parent who 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 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. Trust here seems to be at odds with a certain ideal of theoretical rationality, the requirement that we bring our cognitive attitudes into line with the best available evidence.

This tension between trust and rationality can be set out as resulting from the conjunction of the following three claims:

1. Trust is essentially a cognitive attitude

To say that trust is essentially a cognitive attitude is to say two things about trust. First, whatever else trust might involve, it must involve taking a stance on the truth of some content that can be captured in a that-clause (such as 'that the babysitter will care for my child'). Standardly, the cognitive attitude is that of belief or expectation, and the content of a 'trusting belief' (or expectation) is ordinarily thought to involve the trustworthiness of the person trusted. When we trust, we hold it to be true that the person we trust is trustworthy, or at least expect that we won't be let down in trusting them.[1]

The second thing which is meant by saying that trust is *essentially* a cognitive attitude is that this aspect of trust – that it is a cognitive attitude – is somehow central to an adequate understanding of trust. So, not only must cases of trust involve a cognitive element, but this element must not be seen as merely the consequence of some other, more basic, element in an analysis of trust.

2. The primary set of norms governing the rational propriety of cognitive attitudes is that associated with Evidentialism

Many philosophers contend that some form of Evidentialism is central to our conception of theoretical rationality.[2] According to Evidentialism, a core part of what it means to be theoretically rational is that one's cognitive attitudes stand in the right sort of relationship to evidence. Given that 'evidence' here means something like those considerations which

count in favour of the truth of a proposition, Evidentialism has a great deal of intuitive appeal. If certain considerations count for the truth of the propositional content of one's beliefs, then it seems that one's attitude of holding those propositions true had best respect those considerations. The chief disputes over Evidentialism are not, therefore, over whether Evidentialism in this broad sense is true, but over details (such as how best to characterise the 'right sort of relationship' that one's attitudes must bear towards the relevant considerations), and scope (such as how to proceed in cases where the relevant evidence is counter-balanced.[3])

Some epistemologists favour a simple synchronic version of Evidentialism which requires just that one's beliefs reflect the evidence one has at a given time.[4] Others build in diachronic requirements concerning the ways in which attitudes must be responsive to changes in that evidence over time.[5] It is worth noting that a great deal of the literature on trust and rationality goes beyond both of these, and attempts to reconcile attitudes of trust with what we might call the 'Evidentialist ideal'. According to this ideal, a rational thinker ought to respect evidence, by putting herself in the best possible position for discovering the evidence relevant to her beliefs, and engaging in on-going rational reflection on that evidential base.[6]

3. Trust stands in conflict with Evidential norms

No matter how the norms of Evidentialism are construed, trust invariably seems to stand in tension with them. Two such purported divergences are noted in the literature.[7] First, trust is typically formed and maintained in ways that extend beyond that which is supported by evidence. Second, the attitude of trust itself is one that resists being formed and maintained via a process of weighing evidence, whatever that evidence suggests. In to our opening example, our paranoid parent is a lousy trustor because a good trustor would not engage in the practice of collecting and weighing the evidence in favour of the on-going trustworthiness of the baby-sitter.

The point here is not that trust might on occasion fail to meet Evidentialist demands, for there is nothing controversial with concluding that sometimes we trust irrationally. Nor are these purported divergences from the Evidentialist ideal offered as mere empirical generalizations regarding what trustors typically do. Rather, it is claimed that the tension between trust and Evidentialism is constitutive: part of what it is to trust is that one's attitude of trust bears a relationship to the relevant evidential base that is explicitly prohibited by Evidentialism.

Joint commitment to each of the above three claims, seems to entail the following conclusion:

(C) Trust is not a rationally appropriate attitude.

This conclusion is unwelcome, not least because the negative verdict on trust, delivered from the perspective of theoretical rationality, runs counter to the positive appraisal of trust implicit in everyday practice and discourse, where trust is treated as a desirable and healthy attitude that forms the cornerstone of successful interpersonal relationships and societies. In what follows, we survey some of the attempts that philosophers have made to deny certain of the premises listed and so resist the troubling conclusion they appear mutually to support.

In order to provide this survey, we have had to simplify and/or take a stance on some aspects of both trust and rationality that are complex and contested. In so doing we have been largely guided by what we take to be the orthodoxy in the literature, but would like to disclose some of these assumptions at the outset. First, we focus solely on cases of

interpersonal trust, setting aside the question of its relation to other forms of trust, including institutional and self-trust. Second, the persons in question are assumed to be human adults, thereby setting aside the role of trust relations in developmental contexts.[8] Third, we take for granted that there is a basic difference between reliance and trust, where the hallmark of the latter is a propensity to feel resentment towards the one trusted when let down.[9] Fourth, we work within a broadly internalist epistemic framework, according to which, for an agent's belief to be justified, the reasons that justify that belief must be reflectively recognisable by that agent. Finally, our interest here is on doxastic and not simply propositional justification of belief, such that to be justified in believing something requires not just having good reasons for that belief but that one's belief be *based on* those reasons.

2. Eliminating the Tension

2.1. Rejecting (1): non-cognitivist accounts of trust[10]

The argument formulated above expresses the concern that trust could not be rational because it is fundamentally at odds with Evidentialism. But since Evidentialism is supposed to govern only our cognitive attitudes, the argument will have no traction if (1), the premise that trust is essentially a cognitive attitude, can be denied. Here we look briefly at two influential responses that could be construed as proceeding along these lines: Karen Jones's affective account of trust,[11] and Richard Holton's conception of trust as a practical stance.[12]

Jones's account deflects the Evidentialist concern by construing trust primarily in terms of an affectively-loaded 'way of seeing' another person, which she describes as a form of optimism. Trust is thus not a matter of judging *that* the person trusted is trustworthy, but rather a matter of viewing them in a certain favourable light, by pursuing distinctive lines of inquiry, patterns of salience and tendencies of interpretation. Trust does involve a cognitive component – specifically, an expectation that the one trusted will be favourably moved by the display of optimism – but this is itself grounded in the attitude of optimism, and hence is only justified to the extent that the affective attitude of optimism is itself justified. Jones acknowledges that the optimism constitutive of trust will play a significant role in the formation and maintenance of further cognitive attitudes – for instance, beliefs about the trustworthiness of those trusted – but she is adamant that it should not be mistaken for them. What this means is that, even if (3) is correct and the cognitive attitudes associated with trust – including the expectation that the trustee will be moved by one's trust, and other beliefs which result from trust – may violate Evidentialist norms, this does not entail that trust itself is irrational. For trust on Jones's account is not primarily a cognitive attitude, and so is not apt to be assessed in terms of the norms of theoretical rationality.

For Jones, the cognitive component of trust plays a secondary role to that of the affective component. Holton goes further, allowing for the possibility that trust need not involve a cognitive component at all. What we do when we trust, according to Holton, is we *rely from within the participant stance*. We *rely* in that we work the supposition that we will not be let down into our plans, where this need not involve anything like a belief or confident expectation that one will not be disappointed. And we do this *from within the participant stance*, i.e. the perspective from which one treats other people as autonomous agents; from which one ascribes praise or blame to such agents; and from which one becomes susceptible to 'reactive attitudes' of gratitude, resentment and betrayal in response. On neither of these two aspects of Holton's account is Evidentialism afforded

any grip. Just as an affective attitude is, for Jones, not a matter of taking some propositional content to be true, so the participant stance is, for Holton, not a metaphysical theory about the capacities of human creatures, and thus acting and reacting from within the participant stance is not something that could stand in need of external (evidential) justification. In short, Holton insists that to trust is to adopt a practical stance rather than to believe, which would place it outside the Evidentialist reach.[13]

A nifty feature of non-cognitive accounts like these is that in defending trust against the Evidentialist charge of theoretical irrationality, they simultaneously point towards a positive vindication of the rationality of trust in terms of practical reason. As long as belief is primary, then considerations of trust's value or utility are normally taken to be unavailable as justifiers, since these considerations do not bear on the truth of the belief constitutive of trust. Moreover, instrumental justification of trust – seeing trust as the means to some desired end – would depend on the possibility of trust being willed, and it is commonly assumed that belief is not volitional in this way. But once the primacy of belief is replaced by an affective attitude or a practical stance, then it is less controversial to say that trust can be willed (or at least, for Jones, 'cultivated'), and we are free to see trust as rational in virtue of its usefulness in attaining some goal. Of course, if the attitude of trust is resistant to the very process of weighing evidence in favour of the trustworthiness of the trustee (as suggested above), this would seem to extend to the weighing of practical reasons as well. Nevertheless, by permitting a wider range of reasons in support of the attitude of trusting, non-cognitive approaches offer more room for manoeuvre in resolving the perceived tension between trust and rationality.

In the course of their discussions, both Jones and Holton consider scenarios in which somebody trusts for instrumental reasons while lacking the belief that the trustee is trustworthy. Jones imagines a rape-survivor who decides to cultivate trust in order to realise a certain conception of herself,[14] and Holton considers a drama student who is 'most uncertain' whether or not his fellow students will catch him in a 'trust-circle' exercise, but decides to fall nevertheless, in the hope that by such behaviour an atmosphere of trust might be created in the class.[15] This tactic of reflecting on cases which seem to involve trust in the absence of trusting belief invites the rejoinder that such cases are not cases of trust at all, but of something else, such as *entrusting*.[16] In other words, an opponent can insist that the entrusting described in Jones's and Holton's cases is not an instance of fully trustful trusting, precisely because it lacks the cognitive features seen as constitutive of the attitude of trust.

To elaborate on this reason for rejecting non-cognitivism about trust, suppose, in a variation on our opening case, that our parent decides to have a friend babysit and does not set up the nanny-cam after all, but that she does continue to harbour genuine (i.e. not merely paranoiac) doubts as to the friend's trustworthiness. It may be granted that the not-so-paranoid parent 'trusted' the babysitter in a sense that includes an act of reliance from within the participant stance; an attitude of optimism towards the babysitter; and a vulnerability to betrayal. But it also seems that the very lack of confidence in the babysitter reveals a lack of trust in him in another, stronger sense, a sense in which the attitude of trusting requires belief in the trustworthiness of the person trusted. For some theorists of trust, it is only this stronger sense of trust that is 'fully-fledged', and anything short of involving a belief-like attitude can at most be *acting as if* we trust (or a lower grade of trust) rather than actual (or top grade) trust.[17]

Beyond the question of whether the cognitive component of trust can be eschewed or relegated to secondary status, there is further doubt about the adequacy of these attempts made by Jones and Holton to repel the Evidentialist objection. For both Jones and Holton admit that although trust itself is not primarily constituted by belief, it does play a

decisive role in certain beliefs being formed and fostered. Viewing someone with an attitude of optimism or adopting a stance of trust towards them leads one, in normal circumstances, to consider them trustworthy. But then we might ask, what is the rational status of this belief that arises out of trust? Jones tells us our optimism makes us blinkered, while Holton's practical stance does not admit of external justification, so these beliefs which result from trust are not going to be well-supported. If we understand Evidentialism as a kind of *ideal* which requires believers to take steps to ensure that their beliefs are well-supported by evidence (by broadening one's evidential base, etc.), then it seems that the beliefs which follow the trust of Holton and Jones are inevitably going to fall short of this Evidentialist ideal. The point here is that to attempt to save trust itself from the charge of irrationality while at the same time condemning many of the epistemic states and practices of trustors seems like a hollow defence of our everyday epistemic practices from the charge of irrationality.

2.2. Rejecting (2): non-evidentialist epistemologies of trust

The second way that the Evidentialist objection can be countered is by denying (2), the premise that the primary norm governing the propriety of cognitive attitudes is Evidentialism. We propose to treat both Judith Baker's account of trust as rational commitment and Benjamin McMyler's construal of trust as a second personal attitude as examples of this approach.[18]

Baker's strategy is best viewed as involving an indirect rejection of (2), since her primary target is not Evidentialism *per se,* but a broader conception of rationality in which it is usually embedded. This conception divides justificatory reasons neatly and exclusively into two varieties according to whether they provide support for actions or beliefs: 'theoretical' reasons justify beliefs and are governed by truth-directed norms, whilst 'practical' reasons support actions and are governed by goal-directed norms. Baker takes aim at the sharpness of the purported division between reasons for action and belief in this orthodox conception, contending that truth-directed attitudes can be supported by goal-directed considerations because – to put it bluntly – truth can be a goal.

Her focus is on what she dubs 'rational commitments', which are cognitive attitudes that are under the control of the will to some degree and are supported by goal-directed considerations. Consider, for example, the cognitive attitudes of individual scientists who stubbornly stick by their hypotheses in the face of counter-evidence, or those of the skilled defence advocates who represent only their clients' interests in judicial proceedings. Inasmuch as such scientists take their hypotheses to be true and such lawyers presume their clients to be innocent, they have formed and maintained cognitive attitudes, but these attitudes would seem to be illegitimate by Evidential standards. Nevertheless, both the scientific enterprise and the judicial system are well served in their truth-seeking efforts by the practice of adopting such attitudes. Similarly, Baker contends, attitudes of trust are rational commitments: they are cognitive attitudes which, despite being illegitimate by Evidential standards, serve an important epistemic purpose, since they facilitate interpersonal interactions characterised by openness and honesty. As she puts it:

> ''if a result […] of one's trust is that barriers to honesty are removed and the other person is open with us, then trust in their veracity will be merited and end-directed rationality will not be opposed to truth-directed rationality'' (Baker 1987:13).[19]

Baker's rejection of the regimented picture of rationality thus serves to undermine Evidentialism by rejecting its claims to sovereignty in the realm of belief. Thus trust can be

shown to be rational despite flouting Evidential norms, because it serves to promote 'honest' and 'open' interactions, and in this way serves the goal of truth.

It is unclear in Baker's article just how trust serves a truth-directed goal. One reading (suggested by the quotation above) is that attitudes of trust tend to bring about the trustworthiness which they, as cognitive attitudes, were taking to already obtain. But this reading cannot be correct, since an attitude of trust can as easily prompt exploitation and deceit instead of trustworthiness, and, if this is so, then Baker's solution appears either naïve or unable to properly serve its vindicatory function.[20] On an alternative reading (suggested by the analogy with scientist and lawyer), trust serves a truth-directed goal more indirectly, such as by being a necessary mechanism in a larger enterprise concerned with truth. But this too is problematic, since it does not obviously amount to a defense of trust as theoretically rational. Consider a parallel problem for the case of the scientist and the advocate: just because *what they do* serves an important function in a truth-seeking enterprise, it does not follow that *what they believe* is rationally defensible.

Perhaps this last complaint is unfair, given that Baker's whole point to unseat this very bifurcation between action and belief. Much thus rests on the success of her arguments against the bifurcated conception of rationality. A key claim is that the regimented picture is inadequate to account for these 'rational commitments', attitudes which seem to span the divide between action and belief, and whose justification might accordingly span the divide between practical and theoretical reasons. However, the mere notion of a rational commitment alone seems innocuous for a proponent of the orthodox view, who can quite simply re-analyse any rational commitment into two components – a belief and a sticking-to-belief – each with its own justification and rational status. On the orthodox view, then, we might concede that there is some instrumental justification in the scientist's *stubbornness*, the advocate's *bias* and the trustor's *faith*, but insist nevertheless that, since not one of them has a belief standing in the right sort of relation to evidence, they are all theoretically irrational. Baker is explicit that her discussion is programmatic and sketchy, and a fuller argument is required to reveal the inadequacies of this orthodox rejoinder.

McMyler's work on trust can also be viewed as mandating an *indirect* rejection of (2). Like Baker, McMyler views trust as a distinctive sort of attitude, one that is both fully cognitive and yet not justified in the way in which beliefs are regularly taken to be justified. Whilst Baker holds that trust is a kind of 'commitment' which may be justified on instrumental grounds, McMyler sees trust as a special kind of second-personal attitude, justified in virtue of relations of authority and responsibility existing between trustor and trustee.

To see what he means by this, consider two ways in which our not-so-paranoid-anymore parent might express their new found trust in a babysitter:

(i) I trust Beth to look after the kids.
(ii) I trust that Beth will look after the kids.

As McMyler notes, there appears to be a difference in meaning here, since the two expressions seem to have differing truth conditions. But this difference cannot be one of propositional content of the attitude, since in both cases the only plausible that-clause held true by the speaker is 'that Beth will look after the kids'. McMyler's suggestion for making sense of this intuitive discrepancy in meaning is that the attitude of trust differs across these two expressions in the way in which it is directed upon that propositional content. In (ii), trust features as a 'third-personal attitude', aimed directly upon a state of

affairs (Beth looking after the kids), while, in (i), trust features as a 'second-personal attitude', in that it is directed on that state of affairs only mediately, via a person (Beth). What it means to trust in this second-personal way is thus something more than simply a matter of forming a belief regarding the trustworthiness of the one trusted. It is a matter of forming that belief on the authority of the one trusted; of holding the one trusted responsible for one's trusting belief; and of being disposed to defer challenges to that belief to them.

For McMyler, the cognitive attitude of trust is based on considerations that speak in favour of the truth of a proposition involving the trustworthiness of the one trusted, but the considerations in question are distinctive second-personal reasons, rooted in the authority of the one trusted. In allowing such reasons for trusting beliefs, McMyler is rejecting what he terms the ideal of 'Epistemic Autonomy', which contends that a rational agent alone is solely responsible for all of their beliefs. In contrast, on McMyler's account, part of the justification for the agent's trusting belief is something for which she is not responsible: when challenged she is entitled to 'pass the epistemic buck' to the one trusted. It is in virtue of this rejection of Epistemic Autonomy that his account is also, implicitly, an indirect rejection of Evidentialism. This is because Evidentialism is, at base, an individualistic thesis: rational propriety in the theoretical realm is a matter of whether or not individual believers stand in the right sort of relationship to the evidential base relevant to their judgements. Though other people can feature within this base, their participation is not irreducibly personal in the way that McMyler's account permits.[21]

One challenge to McMyler's account stems precisely from this rejection of Epistemic Autonomy. The challenge is to say how the relations of authority and responsibility between trustee and trustor bear on the truth of what the trustor holds true in trusting, and hence how these facets of the relationship can provide justificatory support *of an epistemic kind* to the trustor's trusting beliefs. Responding to this worry, McMyler draws heavily on an analogy with practical reason. Just as a command provides a *prima facie* reason for one to act as commanded on the authority of the person issuing the command, so too does being invited to trust provide a reason to believe what one was invited to trust in on the authority of the trustee. Arguably, however, this analogy breaks down, because the notion of authority in the theoretical realm is markedly different from the notion of authority operative in the practical case – a fact that can be glimpsed when one considers that practical but not theoretical authority can be 'wielded'. Practical authority is active: a matter of having socially or materially conferred power to make it the case that the world be thus-and-so. In contrast, theoretical authority is reflective: an expert can tell us how things are, but has no special ability – in her capacity as theoretical expert – to change it or demand that it change.

Much of McMyler's book comprises of attempts to respond to this worry, most often claiming that what drives the rejection of the parallel between the epistemic and practical cases is an unmotivated, tacit adherence to the ideal of Epistemic Autonomy. But even if one accepts the parallel, there seems to be another problematic disanalogy between the cases of commanding and trusting. It is a central feature of commanding that there is an explicit act of one person undertaking responsibility for what is thereby authorised. Absent such an explicit undertaking, it does not seem legitimate to defer responsibility to the authoriser. (For example, the defence 'I was just following X's orders' fails if it transpires that X never explicitly ordered anything, even if it can be shown that X may have intended to so order). Yet, whilst there may some cases of trust where there has been an explicit act by the trustee of authorising the trust ('Trust me, I'm a doctor'), such explicit invitations to trust are often absent. Our parent may trust Beth to look after the kids even

when Beth has not explicitly stated she would, or when there is no standing babysitting arrangement, but then it is hard to say in what way the parent can be thought of as trusting on Beth's authority.

To restrict cases of trust only to those where there is an explicit invitation to trust unduly restricts the range of cases in which trust is taken to be warranted. To avoid this, McMyler suggests that even where there is no explicit invitation, there is some *implicit authorisation* based on the prior relationship between trustor and trustee which provides warrant. Yet this notion of 'implicit authorisation' is dubious. This can be seen by reflecting on McMyler's primary case of authority in the theoretical realm, that of X telling Y that p, which McMyler treats as an instance of X authorising Y to believe that p on her (X's) say-so, and thus not of Y making up her own mind on the matter.[22] Suppose that X never explicitly told Y that p, but that Y attributes to X the act of (implicitly) telling her (Y) p, perhaps on the basis of background knowledge gleaned from their long-standing relationship. In such a case, Y must not only attribute the belief that p to X, but also the authorisation for Y to believe p, since McMyler is clear that there is a substantial and epistemically-significant difference between merely possessing a belief and authorising another to take that belief up. But when the authorisation to believe that p is attributed by Y to X, this seems to be precisely a case of Y coming to a conclusion herself and not of deferring to X's authority, which is why deferring a challenge of p to X when X has not explicitly undertaken responsibility for p is, in normal circumstances, inappropriate. Thus, even if one concedes the more general claim that second-personal reasons may be suitable for justifying cognitive attitudes, its application to cases of trust that do not involve an explicit act of authorisation by the one trusted is troublesome.

2.3. Rejecting (3): Evidentialist accounts of trust

A different sort of response to the Evidentialist objection is to admit that trust is essentially a cognitive attitude and that, as such, it should live up to Evidentialist norms governing such attitudes, but deny (3), the premise that the attitude of trust conflicts with Evidential norms. Whilst some may deny (3) by simply rejecting the description of the epistemically distinctive ways in which trustors form and maintain their beliefs outlined above, we will focus on approaches that accept that trust involves such distinctive epistemic practices but deny that these conflict with Evidentialist norms, since – it is claimed – there are good Evidential reasons mandating such distinctive epistemic practices.

An obvious starting point here is the observation that in cases where that conflict appears most acute there is typically a prior relationship between trustor and trustee. So, for instance, when I persist in trusting my friend's claims to innocence despite widespread condemnation and mounting evidence against her, perhaps I do so simply because I know her a lot better than others do. It is not, as Jones implies, that I am 'blinkered' and so unable to perceive the considerations which would impugn the trustworthiness of my friend. Quite the opposite: in virtue of our on-going relationship, I am epistemically *privileged* with regards to the track-record and character of my friend. What this means is that my apparently peculiar epistemic practice, involving, e.g. resistance to apparent counter-evidence, can in fact be readily explained from within the Evidentialist framework, in terms of this differential distribution of evidence.

This line of thought would need to be further developed to provide a satisfactory response to the Evidentialist objection. As it stands it is unable to account for cases of trust which seem to flout evidential norms but where there is no on-going relationship

between the parties – for instance, as might be the case if our parent organised her baby-sitter via an agency. Also, the notion that the 'knowledge' we have of our friends could potentially serve as evidence relevant to the propriety of our ongoing trusting belief is contestable. Such knowledge, like our persistent trust, might itself be the result of a blinkered and biased view, and likewise subject to the charge of irrationality.[23] Finally, even if we admit such knowledge as evidence and restrict our view to favourable, friendship-type cases, appeal to the greater quantity of evidence at the disposal of friends does not quite seem to account for the peculiar ways in which trustors are required by their trust to *approach* that evidence. Arguably, a hyper-informed but disinterested observer does not have the same epistemic duties of the trusting friend, duties which go beyond finding merely coherent interpretations of the facts, to extending charity, or the 'benefit of the doubt', towards one's friends.[24]

A related Evidentialist strategy, which fares better on these counts, has been developed by Philip Pettit.[25] His account hinges on a distinction between trustworthiness and 'trust-reliability': *trustworthiness* is identified with laudable traits such as loyalty, prudence and virtue, while *trust-reliability* is a matter only of the likelihood of somebody meeting the expectations constitutive of a trustor's trust. While trustworthiness supports trust-reliability, so too does a certain, much less laudable trait, which he calls 'esteem-seeking'. Pettit's key claim is that the fact that people generally wish to be well thought of by others (that they seek esteem) is something which can be sensibly exploited by trustors in furtherance of their own private goals.

To show this, Pettit points to the dynamic nature of trusting encounters, in which trustors' signalling of trust tends to create a reason for trustees not to disappoint. When trust is signalled, trustees comprehend that they are considered to be trustworthy and thus held in some sort of esteem, even if, in fact, they are only considered trust-reliable on account of their esteem-seeking nature. Since they will wish for this perceived esteem to be maintained, and perhaps for onlookers to form similarly good opinions of them, this creates a powerful reason for trustees to live up to the expectations of trustors. The upshot of this is that trust will be sensible even when there is some evidence which appears to throw the trustworthiness of a trustee into doubt. All that matters for trust to be rational is that trustors have good reason to think their trust will not be disappointed – i.e. to think that those they trust will be trust-reliable – and it turns out that there is a quite general fact about human psychology which provides just such a reason.

In assessing Pettit's view it is worth considering in more detail the nature of this 'reason' – or as Pettit more often calls it, this 'mechanism' – which allegedly makes acts of trust sensible even when the trustor has no prior evidence that the trustee is trustworthy. Recall that the notion of rational justification in which our discussion is framed is internalist, such that what it means for a certain reason to rationalise a trustor's trust is for them to be able to reflectively recognise that reason as the reason for their trust. In contrast, Pettit's presentation of cunning trust is more amenable to an externalist notion of rationality: the paper is called 'The Cunning of Trust', not 'The Cunning of Trustors', and Pettit treats the fact that there is a reason (the regard-seeking psychological mechanism) for trust, whether or not trustors know about it or base their attitudes upon it, as sufficient grounds for making trust generally rational However, we need not beg the question of the internalist framework of rationality preferred here in order to pose a problem for Pettit. To see this, consider the cunning trustor: a person who (perhaps having read Pettit's article) *happens to* trust on the basis of what they recognise to be the presence of the regard-seeking psychological mechanism of a trustee. Is the cunning trustor really trusting the trustee? It seems peculiar to say that by deceptively exploiting

another person's subpersonal mechanism in order to get them to behave in the way we want, that we could be trusting them.

This last point suggests that some aspects of Pettit's account will strike many as just plain wrong, given the widespread contention that trust is the expression of what the trustor sees as a worthy interpersonal relationship between two parties. First, the interpersonal relationship is typically conceived by the trustor to be such that both parties think well of each other: the trustor holds the trustee in some degree of esteem, whilst the trustee is assumed to have at least some minimal degree of goodwill to the trustor.[26] In contrast, Pettit openly concedes that the disposition to seek regard is not one that is treated as worthy and places no limitations on the perceived motivations of the trustee. Second, the interpersonal relationship is typically conceived to require a certain degree of transparency of intent between trustor and trustee. In stark contrast, Pettit's account requires a *lack* of transparency between the parties in a trusting encounter: if the trustee were aware of the cunning trustor's reason for trusting (i.e. their identification within the trustee of a somewhat deplorable trait of regard-seeking) then the trustee would be less likely to be trust-responsive and thus the trustor would not have that reason to trust. Third, as noted at the outset, the attitude of trust is typically seen to preclude precisely the kind of calculated weighing up of considerations for trust-reliability that the cunning trustor displays.[27]

Pettit himself acknowledges that there are a variety of different forms of trust, and thus could allow that these three characteristics just noted characterise one variety of trust relations, whilst maintaining that his account helps explain the rationality of another variety of trust that does not have these characteristics. It is, however, precisely the variety of trust with these characteristics that is usually assumed to generate the tension between trust and rationality being explored here.

3. Accepting the Tension

Suppose we do not overturn any of the premises in our master argument, and we accept, moreover, that these premises lead, in a way that is more or less valid, to the conclusion, (C), that trust is not a rationally appropriate attitude. If so, one is in a position that could be described as *accepting the tension* between trust and rationality as set out in the argument. By way of conclusion, we propose to finish our discussion of the question 'Is it rational to trust?' by tentatively distinguishing between two different ways of *accepting this tension*: the first involves endorsing the conclusion of the argument, whilst the second does not.

3.1. ENDORSING (C)

Let us begin by considering the prospect of happily endorsing (C). We should note that (C) does not merely state that trust is an *epistemically* irrational attitude, leaving open the possibility of trust being rationally appropriate in some sense other than the epistemic. Given (1), that trust is here conceived as essentially a cognitive attitude, and (2), that such attitudes must meet the standards of Evidentialism if they are to be rational, it seems that once we when we have considered all things epistemic, we have considered all things relevant to the rationality of trust. So, if one does not overturn any of the premises in our master argument, it looks as though we must conclude not just that trust fails to be a rational attitude in terms of such-and-such norms, or from such-and-such perspective, but that it is not rational *simpliciter*.

Earlier we noted that (C) is unwelcome, for it clashes with the generally positive appraisal of trust implicit in everyday discourse. Given this clash, it may be tempting for someone to conclude that it is not our positive appraisal of trust which should give way, but our positive appraisal of rationality.[28] According to this, the puzzlement experienced in contemplating our master argument resembles the bafflement experienced by Spock in Star Trek, when the response so obvious from his arch-rational, Vulcan perspective seems to go against the intuitive response favoured by everyone else, both on the Enterprise and viewing at home. If trust is a psychological state that lies at the heart of our human interpersonal relationships and social structures, while rationality is an abstract ideal recognised by nobody besides part-Vulcans (and dispassionate theoreticians), it seems to matter not a jot that trust might fail to be rational. Following this tempting line of thought, one can freely embrace (C) without this having any revisionary implications for the everyday positive attitude towards trusting, as long as one is prepared to endorse the sentiment that there is more to life than rationality.

We do not propose to give this way of inoculating (C) from revisionary implications much room, since we simply reject the conception of rationality that drives it. On this conception, the norms of rationality are viewed as if they were some externally derived standard, something imposed from without in a manner that leaves room for the possibility of a permanent rupture between the dictates of such norms and what we internally recognise as the right thing to do. In contrast, we have assumed throughout that the notion of rationality is one from which we cannot be alienated, as something internal to our very self-conception, and which underpins our capacity to interpret others as being, like ourselves, interpretable and rational. On this conception, the possibility of acknowledging that some attitude is irrational, whilst nonetheless at the same time claiming that it is the right thing to do, simply makes no sense. If so, returning a negative response to our opening question involves accepting the revisionary implications of (C), not least by ceasing to accord the attitude of trust the high regard it currently has in in everyday discourse.

3.2. RETHINKING THE TENSION

One way, then, of accepting the tension is to bite the bullet by accepting (C) and its revisionary implications. There is, however, a very different way of accepting the tension, which is rooted in – rather than demanding revision of – our everyday experiences of trusting.

To see this alternative, one needs to focus on an assumption which has been underpinning our approach to question 'Is it rational to trust?' from the outset. We have assumed throughout that the tension between trust and rationality is a kind of a conceptual puzzle that requires resolution, where resolution is to be achieved either by showing an error in what seems to generate the tension (here represented as premises in an argument), or by revealing the tension itself to be less problematic than originally thought (the option, here, of accepting the conclusion of an argument). The tension (note: not the conclusion) is thus treated in a manner akin to the *absurdum* of a *reductio*, as something which both can and must be eliminated through theoretical reflection on the way in which it is represented and the grounds which appear to generate it. In a strict *reductio*, there is precisely one possibility that cannot be countenanced: that the tension might be genuine and that the *absurdum* might be ineliminable. In deciding to formulate the tension between trust and rationality in parallel terms, we have thus shielded from view the possibility that the tension itself is genuine and ineliminable.

Reflection on the phenomenology of trust provides some motivation for thinking that we should take seriously the possibility that the tension here is indeed genuine and ineliminable, for the experience of trusting is often vexed. Consider again our trusting parent: we have hitherto failed to note the extent to which her decision to trust the babysitter may be accompanied by ongoing distress, generated precisely by her (perhaps implicit) recognition that maintaining her attitude of trust stands in conflicts with certain demands of rationality. And, arguably, the fact that she maintains her trust (if she does) in the face of these 'rational doubts' makes her trust all the more commendable. We would not want to say that every case of trust need be a vexed one, but if a connection can be drawn between the vexedness of trust and its positive appraisal in everyday discourse, then a theoretical approach to framing the tension such as the one pursued here appears misguided. For if trust is most commendable precisely when it is fraught with tension – when it is enacted in the face of 'rational doubts' – then any attempt to eliminate the tension between trust and rationality in order to vindicate the positive appraisal of trust in everyday discourse is bound to fail.

These cursory reflections suggest the possibility of another way of viewing the tension between trust and rationality which we might call 'practical' rather than 'theoretical', and which is not directed at resolution but which acknowledges that the tension is part of trust itself. The tension between trust and rationality is not, on this practical view, a conflict of the familiar philosophical kind, one which is assumed from the off to be 'mere appearance' and then, through a process of philosophical reflection, shown to be so. Rather, it is something constitutive of trust itself, and thus something which individuals who trust must live and wrestle with. In this sense, the tension is indeed something to be accepted, although this is not done by accepting (C), but by appreciating that the tension cannot be eradicated without at the same time eradicating a central feature of the attitude of trusting itself.

It would be too quick, however, to use these cursory reflections on the phenomenology of trusting to dismiss the potential explanatory insights of the theoretical approach into the tension between trust and rationality, and this is so even if one is prepared to accept that trusting is typically treated as most commendable when vexed. This is because the practical approach to the tension far too briefly gestured at in the previous paragraph is not the only way to take account of the observation that a trustor's sensitivity to the tension between trust and rationality is itself a critical feature of many central cases of trust. For, rather than seeing this phenomenological observation as putting paid to any theoretical attempt to resolve the tension between trust and rationality, we could see it as simply introducing an additional explanatory constraint on an adequate theoretical account of trust. The constraint could be put thus: an adequate account of trust must not just explain why trust is not, after all, in conflict with the norms of rationality, but must also explain why it appears that it is.[29] An account of trust that fulfils both of these explanatory obligations will have resolved the tension between trust and rationality while still remaining faithful to the phenomenology of trust, since what we earlier called the vexedness of trust could arise from the *apparent* tension between trust and rationality. This suggests that the theoretical approach we have taken here in this article might not be fundamental misguided after all; it may just be rather unhealthily focussed on only one of the explanatory tasks involved in tackling the tension between trust and rationality.

So, is it rational to trust? Our goal here has not been to answer this question, but to set out a way of understanding the tension between trust and rationality to which it alludes. We started by attempting to capture the tension in terms of a formal argument,

and then proceeded to explore some of the ways in which resolutions have been attempted in the more recent philosophical literature on trust, construing each of the authors under discussion as trying to undermine one part of this formal argument. By the end of the discussion, it emerged that this theoretical construal may be inadequate in at least one important respect: by being focused exclusively on eliminating the tension, it fails to acknowledge and account for the role that the tension itself plays in the phenomenology of trust. [30]

Short Biographies

Jeremy Wanderer is Senior Lecturer in Philosophy at the University of Cape Town, with research interests at the confluence of issues in moral psychology, epistemology and philosophy of language. He is the author of 'Robert Brandom' (McGill-Queens UP, 2008) and co-editor (with B. Weiss) of 'Reading Brandom: On Making It Explicit' (Routledge, 2010).

Leo Townsend is a Tutor in the philosophy department at the University of Cape Town, where he is also pursuing graduate research into issues relating to norms and reason.

Notes

⋆ Correspondence: Department of Philosophy, University of Cape Town, Private Bag X3, Rondebosch, Cape Town 7701. Email: leo@tenderscan.co.za

[1] That trust just is such an expectation is common in much of the social scientific literature on trust. See, e.g. Gambetta (1988).
[2] See Marušić (2011) for a recent overview.
[3] The essays collected in Dougherty (2011) provide a sense of recent discussions of these issues.
[4] For example: Feldman (2000:688–91).
[5] For example: Hall and Johnson (1998).
[6] Cf. Booth (2007).
[7] For example Baker (1987); McGeer (2008:240); Townley (2011:23–43).
[8] *Pace* McGeer (2002) & Hertzberg (1988).
[9] Baier (1986:235), as discussed and modified in Holton (1994) and Fricker (2012).
[10] The label 'non-cognitivist' is adapted from Becker (1996). Neither Jones nor Holton, whose accounts we discuss in this section, apply the label to their accounts, nor (we suspect) would they embrace it, since both see trust as fully integrated in cognition. As the ensuing discussion makes clear, the tag fits for our purposes nonetheless. An explicit example of this strategy is Lahno (2001). See also Faulkner (2011).
[11] Jones (1996).
[12] Holton (1994).
[13] Holton's (1994) discussion begins by drawing a parallel between the concepts of 'trust' and 'faith'. As an anonymous referee has pointed out, it is possible to generate the same kind of tension between faith and rationality as we have done here for trust. The question of just how faith and trust are related (and the question of the relationship between these and the concept of 'hope') are interesting questions worthy of further exploration. (For some discussion of faith in this regard, see See Audi (2011) and Buchak (2012). For hope, see McGeer (2008) and Martin (2011)).
[14] Jones (1996:23).
[15] Holton (1994:7).
[16] Hieronymi (2008).
[17] Again, see Hieonymi (2008).
[18] Baker (1987); McMyler (2011:Chapter 4). As per our assumption of epistemic internalism, we will not explore more direct routes of rejecting Evidentialism, such as reliabilist accounts of trust (e.g. McLeod (2002)).
[19] In a more recent (as yet unpublished) paper, Baker (ms.) argues that trust serves an epistemic function beyond just encouraging honesty between friends: trust allows us to enter into the sort of dialogue with others needed to, in the terminology of Bernard Williams's (2002) *Truth and Truthfulness*, 'steady the mind' and 'reach one's thoughts'. In short, trust is needed in order to properly form attitudes, including cognitive attitudes, and in this sense trust serves an epistemic function far more basic than the one envisaged in her earlier (1987) paper.
[20] Cf. Jones (1996:15).

[21] McMyler himself formulates Evidentialism more narrowly, such that his account falls within the Evidentialist camp thus conceived. See McMyler 2011:153–4.
[22] As this makes clear, McMyler's work on trust emerges out of his interests in the epistemology of testimony. The relationship between these topics has been the subject of much recent discussion, such as by Moran (2005); Faulkner 2011; Fricker (2012), in addition to McMyler's work.
[23] Baker (1987:5).
[24] Stroud (2006:515–8).
[25] Pettit (1995).
[26] Cf. Baier (1985:234).
[27] McGeer (2008:252).
[28] Stroud (2006:518–9), considering a parallel tension between rationality and friendship, puts this kind of response thus: 'if the canons of epistemic rationality… are incompatible with friendship, then… so much the worse for epistemic rationality'.
[29] The constraint mirrors the contention of some philosophers of mind that an adequate solution to the mind-body problem must, in addition to fulfilling its other explanatory tasks, explain the persistence of the problem itself. (cf. Lycan 1987:42).
[30] We would like to thank Judith Baker whose engaging seminar on trust at the University of Cape Town stimulated our interest in this topic.

Works Cited

Baier, A. C. 'Trust and Antitrust.' *Ethics* 96 (1986): 231–60.
Baker, J. 'Trust and Rationality.' *Pacific Philosophical Quarterly* 68 (1987): 1–13.
Becker, L. C. 'Trust as Noncognitive Security about Motives.' *Ethics* 107.1 (1996): 43–61.
Booth, A. 'The Two Faces of Evidentialism.' *Erkenntnis* 67.3 (2007): 401–17.
Dougherty, T. *Evidentialism and its Discontents*. New York: Oxford University Press, 2011.
Faulkner, P. *Knowledge on Trust*. Oxford: Oxford University Press, 2011.
Feldman, R. 'The Ethics of Belief.' *Philosophy and Phenomenological Research* 60 (2000): 667–95.
Fricker, M. 'Group Testimony? The Making of a Collective Good Informant.' *Philosophy and Phenomenological Research* 84.2 (2012): 249–76.
Gambetta, D. Ed. *Trust: Making and Breaking Cooperative Relations*. Oxford: Blackwell, 1988.
Hall, R. and C Johnson. 'The Epistemic Duty to Seek More Evidence.' *American Philosophical Quarterly* 35 (1998): 129–40.
Hertzberg, L. 'On the Attitude of Trust.' *Inquiry* 31 (1988): 307–22.
Hieronymi, P. 'The Reasons of Trust.' *Australasian Journal of Philosophy* 86.2 (2008): 213–36.
Holton, R. 'Deciding to Trust, Coming to Believe.' *Australasian Journal of Philosophy* 72.1 (1994): 63–76.
Jones, K. 'Trust as an Affective Attitude.' *Ethics* 107 (1996): 4–25.
Lahno, B. 'On the Emotional Character of Trust.' *Ethical Theory and Moral Practice* 4 (2001): 171–89.
Lycan, W. *Consciousness*. Cambridge, MA: MIT Press, 1987.
Martin, A. 'Hopes and Dreams.' *Philosophy and Phenomenological Research* 83 (2011): 148–73.
McGeer, V. 'Developing Trust.' *Philosophical Explorations* 5.1 (2002): 21–38.
——. 'Trust, Hope, and Empowerment.' *Australasian Journal of Philosophy* 86.2 (2008): 237–54.
McLeod, C. *Self-Trust and Reproductive Autonomy*. Cambridge, MA: MIT Press, 2002.
McMyler, B. *Testimony, Trust, and Authority*. New York: Oxford University Press, 2011.
Moran, Richard. 'Getting Told and Being Believed.' *Philosopher's Imprint* 5.5 (2005): 1–29.
Pettit, P. 'The Cunning of Trust.' *Philosophy and Public Affairs* 24 (1995): 202–25.
Stroud, S. 'Epistemic Partiality in Friendship.' *Ethics* 116 (2006): 498–524.
Townley, C. *A Defense of Ignorance*. Lanham, MD: Lexington Books, 2011.
Williams, B. *Truth and Truthfulness*. Princeton, NJ: Princeton University Press, 2002.