
Article:
Faulkner, P. (2014) The Practical Rationality of Trust. Synthese, 191 (9). pp. 1975-1989. ISSN 0039-7857

https://doi.org/10.1007/s11229-012-0103-1

The Practical Rationality of Trust

Paul Faulkner

University of Sheffield

1.

'Trusting' can describe both an attitude and an action. I can trust you to do something like turn up on time by arranging to meet you in a place where it would be very inconvenient to have to wait, and by adopting a certain attitude towards your turning up on time when making this arrangement. Trust *as an action* is the act of depending on someone or something doing something; trust *as an attitude* is then the attitude one takes towards this dependence. These two aspects are constitutively connected: one cannot identify *acts of trusting* independently of the *attitude of trust* they are done with. So to describe an act as trusting goes someway to explaining it. The aim of this paper is to make sense of this explanation.

This paper argues that in describing an action as trusting we can say something about its motivation: the truster *acts out of trust*. And acting out of trust need not be acting for a specific end. This is to agree with Michael Stocker when he states that

> teleological considerations are not sufficient for understanding many significant sorts of acts. To understand them we must refer to their source, or *arche*, not simply their end, or *telos*. More exactly, character and other elements *we act from or out of* are critical for identifying, evaluating, and explaining many significant sorts of acts.[1]

This, I think and hope to show, is true for certain acts of trust. So the first target of this paper is the Humean account of practical rationality. Here I hope to develop and apply Stocker's point: the rationality of acting out of trust is not adequately described in teleological terms. The bigger ambition is then to offer a contrasting theory of how acting out of trust, and so cooperation more generally, can be rationally explicable. This paper starts with the critical project: showing that certain facts about trust tell against the Humean theory of practical rationality. This

[1] Stocker (1981), p.747.

occupies sections two to five. Sections six to nine then develop an alternative account of the practical rationality of trust.

2.

Consider, then, a situation where one party trusts, or depends on another doing something and does so with an attitude of trust. Such situations are everyday occurrences and include handing money over for goods, taking a taxi, waiting for a friend at a pre-arranged location, jointly carrying a heavy object, rowing a boat, asking for directions and so on. These cases can be schematically represented as cases where a trusting party *A* trusts a trusted party *S* to do something ϕ. And two things *can* be true of such cases.

First, we can decide to trust; we can, for instance, decide to ask a passer-by for directions, and then choose to act on what they tell us. A trusting party *A* can decide to trust *S* to ϕ. Of course, this is not always the case. Sometimes one is forced to depend on another doing something where given the choice one would not. In such cases one's attitude towards one's dependence would not be one of trust and the act of depending would, correspondingly, not be one of trusting but of merely relying.[2] However, this distinction makes clear that often we do have options and that trusting is something that one can decide to do. Richard Holton gives the following example.

> Suppose you run a small shop. And suppose you discover that the person you have recently employed has just been convicted of petty theft. Should you trust him with the till? It appears that you can really decide whether or not to do so.[3]

[2] Compare Baier (1986).
[3] Holton (1994), p.63.

The shopkeeper trusts her new employee to be honest and demonstrates this trust by leaving him in charge of the till. Moreover, that this could be an act of trusting and not merely one of relying is shown by imagining this contrary. Suppose that the shopkeeper cannot shake the belief that her new employee will again prove dishonest. And suppose that she makes this belief, and her reluctance to depend on his honesty, manifest while handing over charge of the till. In this case, her new employee could well resent her suspicion while acknowledging its empirical ground. The content of this resentment is that the shopkeeper should trust him, or at least that she should trust him given that she has chosen to rely on him. But this presupposes that the shopkeeper had the liberty to take a trusting attitude towards leaving him in charge of the till. The first fact about trust is that we have such a liberty; it is that a trusting party can decide to trust.

Second, we can see another's depending on us acting in someway as a reason to act that way; we can, for instance, view a tourist's request for directions as a reason to direct him truly. A trusted party *S* can view *A*'s depending on his ϕ-ing as a reason to ϕ. Again, and of course, this is not always the case. *A* might depend on *S* ϕ-ing simply because *A* knows that *S* will ϕ and, on the same grounds, might know that his depending on *S* ϕ-ing will form no part of *S*'s motivation to ϕ. Equally, *A*'s dependence could motivate *S* but it might motivate *S* not to ϕ. This might be the case in a prisoner's dilemma type situation where *A*'s silence, if it were known, could be seen by *S* as a reason to confess.[4] However, if a trusting party's dependence can be viewed as a reason to act untrustworthily, then, in other circumstances, it can equally be viewed as a reason to act trustworthily. Moreover, that we can view things this way is shown by the attitudes we can display as trusting parties. Suppose we've arranged to meet at a place close to your work for an after work drink. And you don't show. Then just as I am about to give up you turn up, but only because this is your route home and you are passing by. When I discover this, the initial annoyance at your tardiness will be replaced by a stronger feeling of resentment.

[4] See Luce and Raiffa (1957).

What is resented is not that you did not show up as planned, since you did, but that your reason for showing up was not what I think it should have been. To understand my resentment, one has to presuppose that it was open for you to view my waiting at this place as a reason for turning up. The second fact about trust is that we can view things this way; it is that a trusted party can see a trusting party's dependence as a reason for behaving as the trusting party expects.

With these two facts about trust in hand it is possible, I would argue, to outline the following account of trust. Here I borrow from what I have said about trust before.[5] On one understanding of trust, to trust is simply to make a judgement of reliability in a situation of dependence. Thus Hollis remarks,

> we trust one another to behave predictably in a sense that applies equally to the natural world at large. I trust my apple tree to bear apples, not oranges. It trust its boughs to bear my weight, if they look strong and healthy. I trust my reliable old alarm clock to wake me tomorrow, as it did yesterday.[6]

To say that *A* trusts *S* to ϕ on this understanding is just to say that *A* depends on *S* ϕ-ing and expects *S* to ϕ. Trust in this sense I have called *predictive*. Predictive trust is a species of belief: the expectation that the trusted will do something is just the prediction or belief that they will do so. To trust is to show a willingness to depend, and so take a risk, because of this belief. Remove the belief and you have simply reliance, and remove the risk and you just have a prediction about how things will turn out. We often trust people to do things in this sense.

However, the expectation characteristic of trust can also be normative. It can concern the trusted party's motivations or reasons for acting, as when I expect you to show up to our pre-arranged meeting. So understood, trust is still a matter of dependence and expectation. It is still a hybrid notion being composed of the factual matter of dependence coupled with the subject's attitude towards this. But the

---

[5] XXX

[6] Hollis (1998), p.10.

expectation is now that the trusted party have certain motivations and act on the basis of these. On this understanding to say that *A* trusts *S* to ϕ is to say that *A* depends on *S* ϕ-ing and expects this to motivate *S* to ϕ. The expectation here is a matter of thinking that *S* should see *A*'s dependence as a reason to ϕ, and should ϕ for this reason. Trust in this sense I have labelled *affective* because of its associations with various reactive attitudes.

What I labelled the two facts about trust are then respectively a fact about affective rather than predictive trust, and a fact about the trustworthiness that is expected in affective trust. That is, subject *A* can decide to trust *S* to ϕ in the affective sense of trusting *S* to ϕ because *A* can commit to acting and thinking in the ways characteristic of affective trust. And in doing so what *A* will expect of *S* is that *S* be trustworthy in the sense that *S* will then see *A*'s depending on his, *S*'s, ϕ-ing as a reason to ϕ. What I would now like to argue is that these facts about this more normative sense of trust do not fit the Humean account of practical rationality.

3.

We can decide to trust and we can see another's depending on us acting in a certain way as a reason for acting in that way. These two facts about trust pose a problem for Humean explanations of action, or so I want to argue in this section. Let me take the first fact first.

Consider an act of trusting in the affective sense, where *A* trusting *S* to ϕ in this sense will consist of *A* depending on *S* ϕ-ing and expecting this to motivate *S* to ϕ. The question is: what explains acts of trust thus identified? On the Humean account our doing things is to be explained in terms of our motivating reasons. Motivation starts with some 'propensity', 'aversion', 'pro-attitude' or *desire* that something be the case. Our motivating reasons are then our desires in conjunction with our means-end beliefs about how to fulfil them, thus Michael Smith's definition:

> R at t constitutes a motivating reason of an agent X to φ iff there is some ψ such that R at t consists of appropriately related desire of X to ψ and a belief that were she to φ she would ψ.[7]

So *A*'s act of trusting *S* to ϕ would be explained by reference to *A*'s desire for *S* to ϕ, or desire for some good that *A* believes follows *S* ϕ-ing, and *A*'s belief that *S* will ϕ. This explanation of acts of trust in terms of desires and means-end beliefs is good for acts of trust that are grounded in trust as a predictive attitude. Where *A*'s trusting *S* to ϕ expresses no more than the belief that *S* will ϕ, all that is needed to motivate *A*'s trust is some desire for *S* to ϕ or for the goods that follow this.

However, when it comes to acts of affective trust, the problem is that *A* can trust *S* to ϕ without the means-end belief – or belief that *S* will ϕ. To illustrate this take the case where I am waiting to meet you so we can go for a drink after work but supplement this case as follows. We are in the process of a messy divorce, and you were nothing but unreliable in the plans you made with me in the past, so I have no confidence you will show up today. This is not to suggest I believe that you will *not* show up; I might have no confidence you will show but if I believed that you will not show it would be wrong to describe my waiting as an act of trusting. In this case it would be something else, my making a point, for instance. Exactly what is implied by the attitude of affective trust is something I will return to shortly, but the point to be made here is that in affectively trusting *S* to ϕ *A* need not believe that *S* will ϕ. Moreover, this is crucial to our being able to decide to trust. *A* can decide to trust *S* to ϕ *just because* trust need not involve the belief that *S* will ϕ. We can decide to trust because the expectation constitutive of trust can be normative rather than merely subjective. And whilst we can choose to hold others to expectations, we cannot choose to believe that someone will behave one way rather than another. *A* cannot decide to believe that *S* will ϕ. But if *A* can decide to trust *S* to ϕ, this belief cannot

[7] Smith (1994), p.92. (I've changed the agent and action place-holders for clarity).

form part of the explanation of *A*'s act of trust. So the first fact about affective trust – that we can decide to trust – does not fit with teleological explanations of action.

Consider now the second fact about trust: that *S* can see *A*'s depending on his ϕ-ing as a reason to ϕ. The problem this poses for a Humean account of action explanation starts with the fact that ϕ-ing might be of some cost to *S*. It might be that *S* has no desire to ϕ and *A*'s depending on his ϕ-ing need not change this. Thus, if *S* is motivated in the end to ϕ, this must be because there is some package of rewards and sanctions that *S* is sensitive to. It does not matter how this package is described, so for the sake of argument consider Russell Hardin's account of what motivates trustworthy behaviour.[8] According to Hardin, *S* will be motivated to be trustworthy to the extent that *S*'s interests "encapsulate" *A*'s interests so that there is a convergence of interest.

> The typical reason [that interests are encapsulated] is that the relationship is ongoing in some sense and that the trusted would like for it to continue. This is the unifying element for encapsulated interests: *the desire for the relationship to continue – for whatever reason, from merely financial interests, to deeper emotional ties, to reputational effects on other relationships.*[9]

So the further motivating desire which trumps *S*'s desire not to ϕ is the desire for his relationship with *A* to continue (where this desire will then be explained in terms of further desires, be they financial reward, emotional connection and so on).

Putting this into Smith's principle then yields: 'R at t constitutes a motivating reason of agent *S* to ϕ iff *S* desires to continue his relationship with *A* and believes that ϕ-ing is a means to achieving this.' However, this is not quite right as it stands. The problem is that *S* might be ignorant of the nature of his desires, or where his best interests lie. Or *S* might fail to recall what he knows to be the case: that not ϕ-ing will be detrimental to the continuation of his relationship with *A*. This problem

---

[8] See Hardin (1996), Hardin (2002) and Hardin (2006).
[9] Hardin (2006), p.31.

can then be described by analogy with theoretical reasoning. One cannot explain *S*'s belief that *q* by observing that *p* entails *q* and *S* believes that *p*. Entailment is merely a logical relation. What is needed for an explanation of *S*'s belief is that *S realise this entailment and so infer* that *q*. Similarly, one cannot explain *S*'s action by observing that *S* desires good relations with *A* and believes that were she to ϕ, she would achieve this. What is also needed is that this belief and desire be "put together" by S.[10] And this is what Smith's principle states, viz. 'R at t consists of an appropriately related desire of X to ψ and a belief that were she to φ she would ψ' Consequently, G.F. Schuler observes,

> Smith's putting-together point is really just the thought that desire-belief explanations of action rely for their explanatory force on the fact that the agent whose action is being explained is engaged in some practical reasoning.[11]

Thus it must be possible – maybe only in after the fact reflection – for *S* to be able to explain his ϕ-ing in terms of the desire for a continued relationship with *A* and the belief that ϕ-ing was a means to achieving this.

The problem, if Schueler's point is correct, is that the Humean account of motivating reasons *must capture how we think about our reasons for trusting* if it is to have any explanatory force. And when our trust is best described in the normative sense characterised *we do not think about things in this way*. It may be that the desire for good continued relations with *A* motivates *S* to ϕ *at some level*. But where *S* thinks as the trustworthy person would, *S* will not think about ϕ-ing in terms of this desire. Rather, if it is this desire that is doing the motivating, it will do so by causing *S* to view *A*'s depending on his ϕ-ing as itself a reason to ϕ. Thus, if *S* is asked why he ϕ-ed, despite the cost, the simple reply '*A* depended on my doing so' or '*A* trusted me to' would be appropriate. Thus, it misdescribes how the trustworthy

[10] "[I]n order for a desire and belief to constitute a motivating reason the agent must, as it were, put the relevant desire and belief together." Smith (1994), p.92.
[11] Schueler (2009), p.109.

person thinks of things to claim that their response to another's dependence is premised on the desire for the continuation of trusting relations. This point is slightly delicate because the motivating desire could be for the continuation of the form of the relationship *viz*. for it to be one of trust. But this motivation receives expression in seeing acts of depending *as themselves reasons* for action.

Jon Elster makes this point, I think, in claiming that trustworthiness is *essentially a by-product*.[12] He cites Montaigne who gives a case much like the case of the shopkeeper in section two: his trusting his servant with "full charge of [his purse] without supervision".[13] The servant, Montaigne observes, "could cheat me just as well if I kept accounts, and, unless he is a devil, by such reckless trust I oblige him to be honest."[14] However, this obligation would vanish if Montaigne's reason for giving his servant full charge of his purse *was simply to make the servant behave well*. To say that trustworthiness is an essential by-product, is then to say that "it cannot be realized by actions motivated only by the desire to realise [it]".[15] The reason for this, I suggest, is that in trusting his servant, Montaigne makes manifest a positive presumption about his servants trustworthiness; whereas Montaigne would make manifest that he believes the quite contrary proposition that his servant is not trustworthy, if his trust were merely a means to an end. Thus, if Montaigne's trust is to be self-fulfilling, as Montaigne hopes, then his servant cannot take it to be merely a means to an end. Rather, he must take Montaigne's trust as expressing a positive presumption about his trustworthiness. And if the servant fulfils this presumption he will act in the way Montaigne expects of him, which is just that he will see Montaigne's depending on his honesty as a reason to be honest. And this will be his reason for being honest not the desire for the instrumental good that follows from his continued employment. So the second fact about trust – that we can see

[12] Elster (2007), p.351.
[13] Montaigne (2004), p.1078.
[14] Montaigne (2004), p.1079, cited in Elster (2007), p.350.
[15] Elster (2007), p.86.

another's depending on us acting a certain way as itself a reason to act that way – does not fit with teleological explanations of action either.

4.

A good response at this juncture would be to note that it is possible that some other desire and means-end belief motivates *A*'s trusting *S* to ϕ, and so rationalises *A*'s trust. What this desire could be is suggested by the divorce and shopkeeper cases. The shopkeeper can trust her new employee to be honest and not steal despite her evidence to the contrary; and the estranged husband waits despite his pessimism. And what seems to motivate in these cases is something like the desire to do the right thing. The shopkeeper feels that she ought to trust her new employee with the till, despite his history. And the husband feels that he ought to trust his wife to turn up, and so work on the presumption that she will, despite her history. However, if there were norms of trust like this, a teleological explanation of acts of affective trust would be simple. Put generally, *A* could trust *S* to ϕ even if *A* had no belief that *S* will ϕ or desire for *S* to ϕ because *A*'s trust could be motivated by the desire to conform to the norm dictating that one ought to trust and the belief that this is a situation where the norm applies. This resolves the apparent conflict with the first fact about trust: *A* can decide to trust *S* to ϕ because doing so need not require *A* believing that *S* will ϕ. Hypothesising norms of trust also resolves the putative conflict with the second fact about trust. To see how it does so requires some consideration of what the content of these norms of trust would be.[16]

Let me say that a trusted party *S* is trustworthy in a circumstance where *A* trusts *S* to ϕ if and only if *S* fulfils the expectation that *A* had in trust. This gives two notions of trustworthiness corresponding to the two notions of trust. A trusted party *S* is trustworthy, in a circumstance defined by *A*'s (affectively) trusting *S* to ϕ, if

[16] See XXX

and only if *S* sees *A*'s depending on his, *S*'s, ϕ-ing as a reason to ϕ and ϕs for this reason. The *norm of trustworthiness* then prescribes that if *A* depends on *S* ϕ-ing, *S* has a reason to ϕ and, other things being equal, *ought to* ϕ for this reason. That there is such a norm is then shown by the fact that *A* is liable to resent *S* if *S* doesn't ϕ or doesn't ϕ for this reason and it is this norm which provides the content of *A*'s resentment. What *A* feels in resenting *S*'s untrustworthiness is that *S did have* a reason and *ought to have acted on it*. This is the reason described by the norm, which is meant to prescribe behaviour irrespective of subjective motivation. Thus the shopkeeper would not resent her employee's theft any the less, though maybe she would curse her naivety, if she learnt that it was planned from the start.

The paired *norm of trust* would then prescribe, in circumstances where *A* could choose to trust *S* to ϕ, that *A* ought to trust. That is, other things being equal, *A should act on the presumption* that if he depends on *S* ϕ-ing, *S* will be motivated by this to ϕ; and if *S* ϕs in a context where it is salient that *A* depends on *S* ϕ-ing, then, other things being equal, *A ought to explain S*'s ϕ-ing *on the basis of the presumption* that *S* ϕs in response to his, *A*'s, need for *S* to ϕ. And that there is such a norm is equally shown by appropriateness of the characteristic emotional reactions to the breach of this norm. Thus, in circumstances where *A* could choose to trust *S* to ϕ, we could imagine *S* resenting *A*'s not doing so; the new employee, for instance, might understand but nevertheless resent the shopkeeper installing CCTV and focusing the camera on the till.

Thus the second fact about trust – that *S* can view *A*'s depending on his ϕ-ing as a reason to ϕ - is merely the claim that *S* can be guided by the norm of trustworthiness.

5.

The presence of norms is indicated by the reactive attitudes that cluster around their violation. Consideration of the reactive attitudes that find their place within

contexts of trust then shows, I think, that there are norms of trust with the content just described. So this defence of the teleological explanation of acts of affective trust is good up to this point: it is right to suppose that there are such norms of trust. However, these norms do not figure in explanations of action in the way proposed by the teleological account. This account does not adequately capture how norms of trust guide behaviour; and this amounts to saying that the second fact about trust remains problematic for the teleological account.

According to the teleological account, the norms of trust and trustworthiness guide behaviour in the following way: taking the norm of trustworthiness, the proposal is that *S* reasons, 'in these circumstances (where *A* depends on me ϕ-ing) the norm dictates that I ought to ϕ, I want to follow this norm, so I should ϕ'. That is, the desire to follow the norm combines with the belief that these are circumstances where the norm applies to yield a reason for doing what the norm prescribes. This is what Paul Boghossian calls the 'intention view' of norm following.[17] And this view, Boghossian follows Kripke in arguing, cannot provide a *general* account of norm following because it makes norm following inferential when inference is itself as instance of norm following.[18] However, this bigger issue aside for the moment, the problem is that the intention view misdescribes how norms of trust guide behaviour.

The problem is that the teleological account of norm following produces behaviour that 'conforms' to the norm, but this is not sufficient for following the norm. If *S* ϕs only because he reasons that this is what he should do, then *S*'s behaviour 'conforms' to the norm of trustworthiness but it is not a case of *S* acting trustworthily, and so actually following this norm. This is because trustworthiness, in the thicker of the two senses identified, requires that one acts for certain reasons. Consequently, the norm of trustworthiness does not merely prescribe a course of action; it also prescribes what one's reasons for this action should be. This norm then guides behaviour through being *internalized*: being guided by the norm is a

[17] Boghossian (2008), p.485.
[18] See Boghossian (2008), pp.490-493, and Kripke (1982).

matter of thinking about the circumstances to which the norm applies in terms of the norm.[19] In the circumstances defined by *A* depending on *S* ϕ-ing, and where the norm of trustworthiness is internalized, *S* will see *A*'s dependence as a reason to ϕ. Insofar as it is internalized this norm then captures the way the *S* will think about this circumstances, viz. he will think it is one wherein he has a reason to ϕ.

There are two points being made here: following these norms of trust requires that one's conformity be grounded on specific reasons; and following these norms is not a matter of making any inferences but of seeing certain things as reasons. Let me take these points in turn.

First, following the norms of trust and trustworthiness requires that one act for specific reasons, or that one's behaviour be explained in a certain way. Otherwise put, these norms prescribe actions – trusting or being trustworthy – whose identity has both behavioural and attitudinal components. Thus the shopkeeper does not really trust her new employee with the till if she installs CCTV to watch him. Trust requires that one takes the risk of depending on another doing something on the basis of positive presumptions about their motivations.[20] In installing the CCTV the shopkeeper might be thinking about her new employee's motivations, and might be thinking about them in order to make a prediction as to the probability of him stealing if he knew he were under observation, but if she decides to 'trust' him on the basis of this assessment, she is not trusting him in the thicker affective sense prescribed by the norm. Equally, supposing that the shopkeeper really does trust her new employee with the till, this employee does not respond trustworthily if his only reason for not stealing is the mistaken belief that the shopkeepr has installed CCTV. In this case, the ex-convict's way of acting is that of the trustworthy person but he does not act trustworthily because he does not have the requisite motivations.

---

[19] See Sripada and Stich (2006).

[20] Trust involves "reliance on [another's] goodwill towards one, as distinct from their dependable habits", Baier (1986), p.234.

This point should be familiar from H.A. Prichard's discussion of moral motivation.[21] One could motivate someone to behave in the same way as a moral person by showing this action to be in that person's interest, but this reason for acting morally is not the reason that would motivate the moral person. Moreover, *any* attempt to substantively motivate someone to behave in the same way as the moral person – to offer motivation for an action beyond it being the moral thing to do – would, Prichard argued, face the same issue. The parallel is that acting as the trustworthy or trusting person would similarly require one have certain motivations. This then connects with the second point: how is it that following these norms of trust, and trusting or being trustworthy, is not a matter of inferring that this is the trusting or trustworthy thing to do and desiring to do this?

Consider the schematic case of *A* trusting *S* to ϕ. The expectation constitutive of *A*'s trust, when this is not the mere prediction that *S* will ϕ, is that *S* will see *A*'s depending on his ϕ-ing as a reason to ϕ and be moved by this reason. The prescription of the norm of trustworthiness, and what *A* expects of *S*, is then that *S* will ϕ because he sees things this way. Following this norm then involves *S* ϕ-ing *because S* sees *A*'s depending on his ϕ-ing as a reason to ϕ. It is a matter of *seeing things as reasons and being moved by them*. This perceptual metaphor should be taken seriously. The point is that, in the ordinary and central cases, a reason for action is delivered by seeing features of the circumstance of trust as reasons for acting, rather than by seeing these same features as grounds for the instantiation of a norm that gives one a reason for acting. A complete explanation of why the trustworthy person ϕ-ed in certain circumstances is then given by stating that he saw a reason to ϕ in these circumstances. The possibility of this explanation then suggests a solution to the sceptical problem for rule following: this problem falsely assumes that internalising a norm is a matter of forming a general intention to follow the norm. This is what Boghossian called the intention view, and it is the view presupposed by teleological action explanations. However, at least for the norms of

[21] Prichard (1912).

trust and trustworthiness, internalization is rather a matter of learning to see things in a certain light.[22]

6.

At this point I'd like to return to Stocker's claim that to understand certain acts "we must recur to their source, or *arche*, not simply their end, or *telos*. More exactly, character and other elements *we act from or out of* are critical for identifying, evaluating, and explaining many significant sorts of acts."[23] This is true, I think, for certain acts of trust; it is true for acts of trusting and being trustworthy in the affective sense. Having argued that we cannot give a satisfactory teleological explanation of such acts, I would now like to make a positive proposal about how they should be understood.

What explains *A*'s act of trusting is *A*'s attitude of trusting: in trusting *S* to ϕ, *A acts out of trust*. To offer such an *out of*, or archaeological, explanation of action, Stocker argues, is: (1) to descriptively identify the act; (2) to explain and justify the action; and (3) to offer an explanation that is not reducible to acting for some end or purpose.

For the case of acting out of trust, I have already argued (1) and partially argued (3). I have argued (1) in claiming that acts of trust and acts of being trustworthy are necessarily done, in part, for certain reasons. The correct identification of the act requires reference to the reasons that motivate it. And I've partially argued (3) in claiming that in trusting *S* to ϕ *A*'s motivation is not best captured in terms of the desire to follow a norm prescribing trusting, and *A*'s motivation need be neither a desire for *S* to ϕ nor a desire for any goods that follow *S* ϕ-ing. To fully argue (3) all plausible candidate purposes would need consideration.

---

[22] McDowell makes the same point about moral norms. See McDowell (1995), pp.100-101.
[23] Stocker (1981), p.747.

However, there is, I think, only one further plausible candidate: *A* trusts *S* to ϕ for the end of securing *S*'s trustworthiness. And this end, I've followed Elster in claiming, is self-defeating. To secure his servant's trust Montaigne's trusting his servant with his purse cannot have this as its end; trustworthiness, in Elster's term cannot but be a by-product of trust. To secure it, Montaigne's trusting his servant with his purse must be an act done out of trust, rather than an act done for an end. Only in this case would it really be an act of trust. So the trusting, if it is to be such, must be explained as an act done out of trust.

This brings me to Stocker's (2), and the two questions are: *how does the archaeological explanation that A acts out of trust explain A's trust? And how does it justify A's trust?* I consider the explanatory question in section seven, and then the justificatory question in the final sections eight and nine.

7.

A key feature of acts of affective trust is that their explanation need not presuppose a belief about outcome. The is illustrated by the divorce case: the long suffering husband knows that his soon to be ex-wife has a history of unreliability, and this is evidence that she will leave him waiting today. So, short of further evidence, it would be unreasonable for him simply to believe that things will work out favourably. However, one could suppose that the husband has no further evidence and yet not attribute any naivety or false belief in explaining his trust. All that needs to be considered is how he, as a trusting person, will see things. Keeping this case in mind but using the more general schema, in trusting *S* to ϕ, *A* will perceive the situation defined by this act of trust as one wherein *S* has a reason ϕ. So other things being equal *A* will presume that *S* will be moved by this reason and so will ϕ. If this turns out to be true and *S* acts as *A* expects, *S* will have proved trustworthy. So in affectively trusting *S* to ϕ, *A* presumes that *S* will prove trustworthy just as in predictively trusting *S* to ϕ, *A* would believe this. This is not to suggest that trust

involves *A* reasoning to this conclusion but is rather to claim that in trusting *S* to ϕ, *A* makes this presumption. However, the presumption that *S* will ϕ renders *A*'s act of trust rationally explicable in the same way that the belief that *S* will ϕ would do so. Consequently, the act of trust is rationally *self-supporting*: it implies a presumption that gives a reason for trusting.[24]

It might seem odd that trust can bootstrap itself into rational explicability in this way. However this oddness should be lessened once it is clear that trust is both an attitude and an action and that what is being offered is an account of the interaction of these two aspects of trust. The dynamic by means of which reasons for trusting are generated can then be clarified by separating out the temporal stages wherein an act of trust follows a decision to trust. In giving this stage by stage account I borrow from Michael Bratman's theory of intention.

The intention to do something, Bratman claims, involves a commitment to act and reason. Suppose that at t0, *A* decides to trust *S* to ϕ, this decision involves *commitments to act and reason*. It implies a commitment to depend on *S* ϕ-ing at some later time t1 and a commitment to act as if *S* had ϕ-ed at t2, some later time still. In deciding to trust his wife to honour their plan to meet, for instance, the husband commits to leaving his house, going to the arranged meeting and waiting for his wife if she is not already there. And the decision to trust involves a commitment to reason in a characteristic way. Intentions extend into action through guiding practical reasoning, Bratman argues, and the decision to trust implies a commitment to thinking about the trust situation in a way that then provides the set of motivations rationally needed to extend the decision to trust into an act of trusting.

Intentions guide practical reasoning by determining the background against which this reasoning occurs, where this background is one of *acceptance* rather than belief. Acceptance is very much like belief: to accept that *p* is to reason and act as if

[24] That trust gives such a motivating reason I argue in XXX.

one believed that *p*. Acceptance is differentiated from belief in that one can accept something one believes to be false. And acceptance does not have the stability of belief: what is accepted either becomes belief, or is abandoned when the context of acting and reasoning ends.[25] The set of things accepted may then *bracket* certain beliefs or *posit* certain things that are not believed. For instance, if the cost of error is sufficiently high, it can make sense to bracket the belief that the probability of a bad outcome is very small and to posit its contrary for the purpose of deciding what to do.[26]

Similarly, in trusting his wife the husband brackets his belief that she will in all likelihood not show, and brackets those beliefs that give him his reason for thinking this, and thereby gives his wife the benefit of the doubt. The commitment to reason implied by a decision to trust is then a commitment to accept various things for the purposes of further practical reasoning. Thus, in deciding to trust *S* to ϕ, at t0, *A* accepts '*S* will be able to see that I depend on his ϕ-ing' and '*S* will see this as a reason to ϕ'. Antecedently to the decision to trust *S* to ϕ, *A* may or may not have believed this about *S*, but adopting the attitude of affective trust commits *A* to accepting these propositions about *S*. Having accepted this at t0, at t1, when the context is one of depending on *S* ϕ-ing, *A* is thereby committed to accepting, '*S* can see my dependence on his ϕ-ing', '*S* sees that he has a reason to ϕ', and so to accepting: '*S* will ϕ for this reason, other things being equal'. Then at t2, and assuming that *S* did ϕ, *A* is committed to accepting '*S* ϕ-ed, at least in part, because I depended on his ϕ-ing'.

This set of commitments, of course, can be overturned by the evidence: in deciding to trust *S* to ϕ, *A* does not decide to trust come what may. But if any of these propositions were not accepted, it would cease to make sense to say that *A* trusted

---

[25] See Cohen (1992) and Bratman (1999), n.20, p.30.

[26] See Bratman (1999), p.29. Such cases, I think, make trouble for Hawthorne and Stanley's principle that you should "Treat the proposition that *p* as a reason for acting only if you know that *p*." Hawthorne and Stanley (2008), p.577.

because accepting this set of propositions is an expression of *A*'s attitude of trust and so a commitment of *A*'s decision to trust. This background of acceptance then specifies a way of thinking about trusting *S* to ϕ, which provides *A* with a motivation for so doing.

This is not to suggest that trusting someone to do something involves explicitly representing these accepted propositions in one's reasoning. The claim is rather that the acceptance of these propositions provides a rational basis for the act of trusting. And this partly defines how it is that the attitude of affective trust involves seeing things in a certain light.[27]

8.

Turning now to the issue of justification, the question is not how *A*'s attitude of trust motivates *A* to trust *S* to ϕ, but what, if any, justifying reason it provides for doing this? On the teleological account the answer is 'whatever reasons *A* can be seen to have on the basis of rational deliberation from what else *A* believes and desires'. This basis and process of rational deliberation, Bernard Williams has famously argued, must be taken broadly.[28] The basis cannot just be what is believed but must include relevant improvements to what is believed. Thus *A* might believe this petrol to be gin but that doesn't give him a reason to mix it with tonic and drink it. However, any improvement to *A*'s epistemic position, made by purging his belief set of falsehoods or supplementing it with truths, must be easily reached "otherwise one merely says *A* would have a reason to ϕ if he knew that fact".[29]

Equally, the process of reaching the conclusion that one has a reason to ϕ cannot just be a matter of means-end reasoning but could be an imaginative process, or indeed any process that could qualify as 'reasoning'. Thus Montaigne's servant

[27] For what else is involved see XXX, pp.185-6.
[28] Williams (1980).
[29] Williams (1980), p.103.

might decide not to take the opportunity of theft he is presented with, not because he reasons that this will secure his employment but because he is shamed by imagining the discovery of his theft. And whilst this presumes the desire to avoid the sentiment of shame, this desire gives the servant a reason to be honest by determining the nature of what the servant imagines rather than by supplying an end for a process of means-end reasoning.

However, this point about the broadness of rational deliberation does not distinguish justificatory from explanatory reasons: it is an observation about practical reasoning generally. To make this distinction between justification and explanation it must be possible to distinguish good and bad processes of rational deliberation, just as it is possible to distinguish good and bad bases for deliberating. If *A* had drank the petrol, his action would have been done for a reason and so explicable, but this explanation would fail to justify *A*'s action, and in this sense *A* had no reason to drink the petrol, since it would refer to *A*'s false belief that the stuff to which *A* added tonic was gin. Equally, it is possible to distinguish good and bad processes of rational deliberation: in logical reasoning one can go wrong, *and if there are norms of trust, then one can go wrong in one's thinking about the trust situation*.

Take a situation where *A* decides to trust *S* to ϕ – Montaigne's deciding to trust his servant or the shopkeeper deciding to trust the ex-convict, for instance – in such a case, if *A* goes ahead and does depend on *S* ϕ-ing, then the norm of trustworthiness prescribes that *S* should see this fact as a reason to ϕ in reasoning about what to. And the norm of trust prescribes that, other things being equal, *A*'s practical reasoning ought to be based on the presumption that *S* will see things this way. In so far as these norms have been internalized, rational deliberation need not start from what *A* or *S* desires but can begin with their perception of the situation they are in, where this perception contains within it a judgement about what ought to be done. This judgement stems from seeing things in terms of the relevant norm, where to see things this way is just to have internalized this norm. So *S* will see *A*'s

depending on his ϕ-ing as a reason to ϕ, whilst *A* will see the trust situation as one wherein *S* has a reason to ϕ. In then treating this fact as a reason *S* will conform to the norm of trustworthiness. Whilst in making this presumption in any further practical reasoning *A* conforms to the norm of trust. To reason differently would then be to fail to conform to these norms and so in terms of these norms, it would be to *go wrong* in one's rational deliberation in this trust situation.

9.

On one understanding, to treat X as a justifying consideration is to value X.[30] So in conforming to the norms of trust, one demonstrates a valuation of trust and trustworthiness. On another understanding, to say that *A* values X is to say that *A* would desire X *under conditions of full rationality*.[31] This raises the question: is rational deliberation in conformity with the norms of trust just a matter of reasoning as the fully rational person would? This question is complicated, and I do not feel able to offer a satisfactory answer. However, two considerations suggest a negative response.

First, when we think about cases, and about the decision to trust or not these cases present, it does not seem that this decision is merely a matter of rationality. The shopkeeper, for instance, does not seem to be compelled by reason to trust her new employee, and nor does this employee seem to be under any rational obligation to respond positively to her trust. It is a good thing if she trusts and he proves honest. But she could not be accused of irrationality if she took precautions (ensured he was supervised, installed CCTV etc.) and he could not be accused of

[30] See Bratman (1996), p.39 and Bratman (2000).

[31] Smith (1994), p.156. For Smith, these conditions give a slightly more robust way of completing what *A* has a reason to do than that offered by Williams; that is, in order to be fully rational, *A* must have no false beliefs, have all the relevant true beliefs, and deliberate correctly.

irrationality if he behaved dishonestly. Any such accusations of irrationality seem to be, as Williams remarks, mere "bluff".[32]

Second, when thinking about whether to trust or to be trustworthy it is possible to consider these questions exclusively in terms of the thin notion of predictive trust and its paired notion of trustworthiness. This restriction is often characteristic of debates on the rationality of trust; and given this restriction, rationality will be entirely instrumental and determined by interest.[33] What I have argued is that to operate with this restriction is to miss the sense we find in trusting and trustworthy behaviour; it is to miss how we explain and justify our trusting and our being trustworthy. One consequence of this omission is that certain acts of trust which have a clear rationale, such as the shopkeeper's trusting her new employee with the till, are wrongly taken to be irrational. Another consequence is that certain 'trusting strategies' can be rationally justified even when they demonstrate little trust. For instance to operate a policy of tit-for-tat is not to behave as a trusting person because there should be as much shame in 'titting' as 'tatting'.[34] The consequence that bears on the present point is that it would be hard to argue that the subject who does think about things instrumentally, who does operate a policy of tit-for-tat, thereby *demonstrates a failure of rationality*. Operating a policy of tit-for-tat seems to be fully rational.[35]

However, such a subject does not conform to the norms of trust and trustworthiness. The rational deliberation that is prescribed by these norms of trust, therefore, does not seem to be simply a matter of reasoning as the fully rational person would. The norms of trust articulate our deliberative perspective, but this perspective thereby seems to embody something fuller than that of instrumental

---

[32] Williams (1980), p.111.
[33] This perspective on trust has been very influential. See, for instance, the contributions in Gambetta (1988).
[34] I argue this point properly in XXX.
[35] See Axelrod (1984).

reason; it seems to be an expression of our way of valuing things, or what McDowell called an ethical outlook.

(Author references omitted)

Axelrod, R. 1984. *The Evolution of Cooperation*. New York: Basic Books.
Baier, A. 1986. "Trust and Antitrust". *Ethics* 96:231-60.
Boghossian, Paul. 2008. "Epistemic Rules". *Journal of Philosophy* 105 (9):472-500.
Bratman, Michael. 1996. "Identification, Decision and Treating as a Reason". In *Faces of Intention - Selected Essays on Intention and Agency*, edited by M. Bratman. Cambridge: CUP.
———. 1999. "Practical Reasoning and Acceptance in a Context". In *Faces of Intention: Selected Essays on Intention and Agency*, edited by Bratman. Cambridge: Cambridge University Press. Original edition, Mind 1992.
———. 2000. "Reflection, Planning and Temporally Extended Agency". In *Structures of Agency*, edited by M. Bratman. Oxford: OUP.
Cohen, L.J. 1992. *An Essay on Acceptance and Belief*. Oxford: Clarendon Press.
Elster, Jon. 2007. *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*. Cambridge University Press.
Gambetta, D., ed. 1988. *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell.
Hardin, R. 1996. "Trustworthiness". *Ethics* 107 (1):26-42.
———. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.
———. 2006. *Trust*. Cambridge: Polity.
Hawthorne, John, and Jason Stanley. 2008. "Knowledge and Action". *Journal of Philosophy* 105 (10):571-590.
Hollis, M. 1998. *Trust Within Reason*. Cambridge: Cambridge University Press.
Holton, R. 1994. "Deciding to Trust, Coming to Believe". *Australasian Journal of Philosophy* 72 (1):63-76.
Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
Luce, R.D., and H. Raiffa. 1957. *Games and Decisions*. Mineola, N.Y.: Dover Publications.
McDowell, John. 1995. "Might There Be External Reasons?". In *Mind, Value and Reality*, edited by J. McDowell. Cambridge, MA: Harvard University Press.
Montaigne, Michel De. 2004. "On Vanity". In *The Complete Essays*, edited by M. A. Screech. London: Penguin Books.
Prichard, H.A. 1912. "Does Moral Philosophy Rest on a Mistake?". *Mind* 21 (81):21-37.
Schueler, G.F. 2009. "The Humean Theory of Motivation Rejected". *Philosophy and Phenomenological Research* 78 (1):103-122.

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
Sripada, C.S., and S. Stich. 2006. "A Framework for the Psychology of Norms". In *The Innate Mind: Culture and Cognition*, edited by P. Carruthers, S. Laurene and S. Stich. Oxford: Oxford University Press.
Stocker, Michael. 1981. "Values and Purposes: The Limits of Teleology and the Ends of Friendship". *The Journal of Philosophy* 78 (12):747-765.
Williams, B. 1980. "Internal and External Reasons". In *Moral Luck*, edited by B. Williams. Cambridge: C.U.P.