

Trust in the guise of belief

Article (Accepted Version)

Booth, Anthony Robert (2018) Trust in the guise of belief. *International Journal of Philosophical Studies*, 26 (2). pp. 156-172. ISSN 0967-2559

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/73701/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

*Trust in the guise of belief*¹

2nd Place Prize in Annual Robert Papazian Competition at IJPS

Anthony Robert Booth

Abstract

What kind of mental state is trust? It seems to have features that can lead one to think that it is a doxastic state (cf. Adler 1994, Baier 1986, Hieronymi 2008, Keren 2014, McMyler 2011) but also features that can lead one to think that it is a non-doxastic state (cf. Baker 1987, Faulkner 2011, Jones 1996, McLeod 2011). This has even lead some philosophers to think that trust is a unique mental state that has both mind-to-world and world-to-mind direction of fit (Holton 1996), or to give up on the idea that there is a univocal analysis of trust to be had (Hardin 2004, Simpson 2012). Here, I propose that 'trust' is the name we give to mental states that we would think of as beliefs if belief was to be thought of in 'pragmatist' terms (that is, as a state posited primarily to explain agents' actions) and belief resists 'pragmatist' treatment. Only such an account, I argue, can univocally account for all the diverse features of trust. As such, I also propose that the explanation of trust provides us with a case for understanding the limitations of a comprehensively 'pragmatist', or 'Neo-Wittgensteinian' conception of the mental.

1. A 'Mixed' Mental State Pragmatism

What is trust? What kind of mental state is it? Is it a belief-like (doxastic) state, an affective, emotive state, or perhaps a kind of a stance (perhaps a 'participant stance' which comes along with evidential standards for judging the trustworthiness of a fellow person²)? There currently does not seem to be any real consensus as regards the answer to this question. Some think that

¹ With many thanks to Boudewijn de Bruin, Gail Leckie, Marco Meyer, Michael Morris, Mahon O'Brien, Alex Oliver, Rik Peels, Sarah Sawyer, Rowland Stout, Chris Thompson, Rik Peels, Jens Van 't Klooster, and an audience at the University of Southampton Philosophy Research Seminar, for their helpful comments on earlier drafts of this paper.

² As well as involving thinking it appropriate to have certain reactive attitudes on discovery of a betrayal of trust.

trust is (partly or wholly) a belief (e.g. Adler 1994, Baier 1986, Hieronymi 2008, Keren 2014, McMyler 2011); some think it is a non-doxastic mental state such as an emotive state (e.g. Baker 1987, Faulkner 2011, Jones 1996, McLeod 2011), or a non-doxastic disposition to rely on a trusted party (Kappel forthcoming), and some think it is a kind of a stance, a “participant stance” (very influentially Holton 1996).

The reason for the disagreement is that trust has properties that make it look like a doxastic state and properties that make it look like a non-doxastic state, to the point where people even have given up on thinking that there is a univocal account of trust to be had (cf. Hardin 2004, Simpson 2012) or to suggest that trust has what Michael Smith (1987) argued no mental state could have – both mind-to-world *and* world-to-mind direction of fit (Holton 1996).

I want to propose what I am tentatively going to call the ‘guise of belief’ thesis about the metaphysics of trust. That is, I want to propose that trust is the name we give to mental states that we would think of as beliefs if belief was to be thought of in what I will call ‘pragmatist’ terms³. We ought not to think of belief in fully ‘pragmatist’ terms, because belief is involuntary. But other mental states (I think trust is one, and intention another) rightly deserve pragmatist treatment. Let’s call the thesis I want to defend ‘the trust in the

³ By ‘pragmatist’ I don’t really have in mind the sort of view usually associated with, say, C.S. Pierce as I will explain (I think there is a relation between the view I call ‘pragmatism’ and the latter, but I will not have the space to explore that relation).

guise of belief thesis' - *TG* for short. I want to argue that *TG* is the only theory of trust that can account for *all* the diverse features of trust that have putatively been identified, and that this should warrant further investigation of the view. Here is a first stab at defining *TG*:

TG *S* trusts that someone else *S'* Φ -s iff *S* voluntarily p-lieves that *S'* Φ -s.⁴

The definition above is hardly very informative unless what 'p-lieves' is defined⁵. Here is what I have in mind:

P-lief *S* p-lieves that *p* iff ascribing a belief that *p* to *S* is the proximate best way to explain *S*'s internal and/or external behaviour.

And now I need to explain what I mean by the 'proximate best' explanation. First, though, let me say something about the general 'Neo-Wittgenstienian' 'pragmatist' account of what are our mental states underlying this. I think that Bob Stalnaker summarises the view nicely with this:

Rational creatures are essentially agents. Representational states should be understood primarily in terms of the role that they play in

⁴ Background assumption: trust is a three-place relation - "person *B* is trustworthy with respect to person *A* and action Φ (or domain of interaction *D*)" (Keren 2014, p. 2595). This is the standard view in the literature on trust; however see Domenicucci & Holton (forthcoming) for a dissenting view.

⁵ I am not going to say anything here about what the content of the p-lief is; what it is to be 'trustworthy'. Here I follow the literature in bifurcating between discussions on 'trust' and 'trustworthiness'. However, I find the account of trustworthiness in Hardin 2004 attractive.

the characterization and explanation of action...and, according to this picture, our conceptions of belief and attitudes pro and con are conceptions of states which explain why a rational agent does what he does.⁶ (Stalnaker 1987, p. 4).

According to this 'pragmatist' picture, there cannot be any fully 'private' mental states, in the sense that having a mental state must be associated with some actual or possible action(s) (whether a further mental action, or otherwise): since, as Anscombe puts it, if having a mental state didn't "guarantee" some action, "then what is there left for it to be but a combination in a vacuum?" (Anscombe 1957, p. 52). The idea, that I am labelling 'pragmatist' here is the reductive one that belief (and mental states in general) are nothing but states that explain action.

I think the question of whether to treat belief in this pragmatist manner is orthogonal to the question of whether belief is a 'dispositional' (e.g. Braithwaite 1967, Schwitzgebel 2002) or 'occurrent' (e.g. Bogdan 1986) mental state (and as such is compatible with the failure or behaviourism). I think this because a pragmatist account of belief leaves it an open question whether it is more explanatory to think of belief as a representational, occurrent state or a dispositional state⁷.

More controversially, however, I think that it is (also) an open question whether we should be pragmatists about *all* our mental states. That is, I take

⁶ See also Lewis' "Rationalization Principle" (Lewis 1974).

⁷ As I said the sense of 'private' I was referring to earlier is just about it's guaranteeing some mental or physical action, and not whether it's a representational or 'occurrent' state.

the following question to be open: are there some mental states that resist pragmatist treatment (while some do not)? If we answer in the affirmative, these mental states would then be essentially private occurrences, 'bombinating' – but not always causing a shift – among our actions (mental or otherwise). I think that belief is such a state: belief is not *merely* a state postited to explain action. And here is an extremely simple argument for why:

- (1) If belief is rightly to be thought of along pragmatist lines, believing at will is possible.
- (2) Believing at will is not possible.
- (3) Belief ought not to be thought of along pragmatist lines.

I think that we cannot believe at will, along with most people writing on the issue.⁸ When I say that 'believing at will is not possible' I mean that we cannot generate belief that *p* either simply by intending to believe that *p* or by doing something that will directly cause us to believe that *p* (cases where I make it the case that *p* in order to believe that *p* notwithstanding, c.f. Feldman 2001). My case for (1) I borrow from Hedden 2015. His argument, however, is the obverse of mine - something like as follows:

- (1) If belief is to be thought of along pragmatist lines, believing at will is possible.
- (4) Belief ought to be thought of along pragmatist lines.
- (5) Believing at will is possible.

As Hedden puts it:

⁸ C.f. Peels forthcoming for an outlier position.

According to [a] Pragmatist picture of belief, whether an agent has a given belief is determined in part by how she acts. In particular, whether an agent has a given belief is determined in part by how well attributing to her that belief would rationalize and explain her actions. Since an agent has voluntary control over her actions, and hence has control over one of the key factors that determine what she counts as believing, she thereby also has voluntary control over what she believes. (Hedden 2015, pp. 495 – 496).

Clearly, as they say, one person's *modus ponens* is another's *modus tollens*!⁹ I endorse Hedden's defence of (1): that under pragmatism whether an agent has a belief is determined in part by how well attributing her that belief explains her actions, and agents have voluntary control over their actions.¹⁰ However, his defence of (4,5) seems to me to be predicated on the idea that endorsing pragmatism about mental states is an *all or nothing* affair – that one cannot be a pragmatist about some mental states but not others. And this is the idea I here seek to deny. I want to show that denying it is not hopeless by showing that a 'mixed' mental state pragmatism can help us make sense of trust.

So, returning to my definition of p-lief, recall:

⁹ Please note, however, that I am here only really *articulating* not arguing for my view. The argument for the view will come in later sections. One could reply here on Hedden's behalf with the suggestion that his version of the thesis that we have doxastic control is compatible with at least *some* of the ways people have of being doxastic involuntarists. Alston 1989 has some very persuasive arguments in defence of the view that *no* version of the theses that we have doxastic control is true (and his is a *very* comprehensive treatment of all the different ways one could take that thesis). I am - for the purposes of this paper - relying on the success of Alston's paper, but only in terms of how I think the view is most easily or best understood, *at first blush!* As I re-iterate, the arguments I marshal in support for the view comes in the later sections.

¹⁰ I think Hedden does a great job fencing off potential worries with this. I refer the reader to his paper if she has such worries.

P-lief S p-lieves that p iff ascribing a belief that p to S is the proximate best way to explain S 's internal or external behaviour.

I can now explain what I mean by 'proximate best' explanation. By 'proximate best' explanation, I mean an explanation where we need to imagine the absence of an undercutting defeater in order for it to be the 'best'. A proximate best explanation would have been the best had a given undercutting defeater not obtained. The relevant undercutting defeater here is that we have reason to think that belief resists pragmatist treatment. Because belief is involuntary (and thus possibly private) the 'best' explanation as to S 's mental and non-mental actions might involve ascription of a belief that p to S , but it might nevertheless be possible that S does not believe that p . In other words, knowledge that pragmatism does not apply to belief works as a kind of undercutting defeater for the putative explanation.

This is *not* to say that beliefs do not explain action, but they would if they were voluntary. Rather, the claim is that belief cannot be *nothing but* a state posited to explain action because it is involuntary.¹¹ Now, some philosophers think that belief has a constitutive norm or aim such that belief's voluntariness is ruled out on conceptual grounds (e.g. Boghossian 2003, McHugh 2012, Setiya 2014, Shah & Velleman 2005, Steglich-Petersen 2009¹²). This might make it look impossible to even imagine what 'belief' would look like were it not involuntary, such that we cannot conceive of the

¹¹ Thanks to Rowland Stout for suggesting this clarification.

¹² For arguments that belief is merely contingently involuntary, see Booth 2017 and Bennett 1990.

counterfactual situation relevant for ascription of p-lief¹³. Crucially, however, what the doxastic constitutivist accounts above (unanimously claim to) rule out on conceptual grounds is that one can believe something as a result of an intention to believe it: that an *intention to believe that p* itself can cause a belief that *p*. However, if pragmatism about belief were true, then the sense in which we could believe at will is that by acting in a certain way we could directly make it the case that we have one doxastic attitude over another – it would not be the intention itself that has the causal power. As Hedden puts it:

For all I have said, it might be that in some cases, by acting one way rather than another, you can make it the case that you have one set of beliefs rather than another, even though it is impossible for you to form an intention to believe *p* and have that intention cause you to come to believe *p*. (Hedden 2015, p. 508).

As such, my claim that we cannot believe at will (in a way that is incompatible with doxastic pragmatism) should be taken to describe a contingent fact about belief in this world.

Also note that p-lief is not the attitude that has been sometimes called ‘acceptance’. This is because, at least on certain prominent accounts, acceptance is ‘context-specific’ (e.g. Bratman 1992, Cohen 1989, Frost-Arnold 2014). Roughly, one accepts that *p* only if one takes oneself to have good reasons to act in light of *p* in a given context, where ‘good reasons’ can denote non-epistemic reasons (so ‘acceptance’ can be voluntary where belief is not).

Here is Bratman explicating the notion:

¹³ Thanks to [anonymous] for bringing this point to my attention.

The three of us need jointly to decide whether to build a house together. We agree to base our deliberations on the assumption that the total cost of the project will include the top of the estimated range offered by each of the sub-contractors. We facilitate our group deliberations and decisions by agreeing on a common framework of assumptions. We each accept these assumptions in this context, the context of our group's deliberations, even though it may well be that none of us believes these assumptions or accepts them in other, more individualistic contexts. (Bratman 1992, p. 7).

As Frost-Arnold (2014) notes, for instance, if one were asked to place a bet on the cost of the house, one presumably would not take the highest sub-contractor estimates for granted in one's calculations. Like belief, however, p-lief is not context-dependent in the way that acceptance is meant to be above. Further, it is possible that one accept that p (and plan to act on p) even if one ends up not acting in light of p , while p-lieving that p guarantees some action (mental or otherwise) relevant to p . Though one of course could be a pragmatist about acceptance and hold that accepting that p must guarantee *some* action(s) relevant to p (such that ascribing acceptance is the best explanation of one's actions), but need not necessarily involve acting in a way such that p is a premise in one's practical deliberations.

The obvious objection is to wonder whether p-lief is really sufficient for trust. Does Sam trust Adam when he p-lieves him to be cheating on him, for example? I think the concern is allayed simply by recalling two things about

the account: first, not just any p-lief is sufficient for trust. The p-lief has to be voluntary, and concern the actions (or a domain of interaction) of another subject. Second, on the account, trust is a three-place relation. Sam does *not* trust Adam, if by that we mean that Sam does not trust Adam *to be faithful to him*. But I do not think there is anything infelicitous about saying that Sam trusts Adam *not to be faithful* – especially when we bear in mind that Sam trusts voluntarily, and so possibly against his evidence.

My argument for *TG* is that it can univocally account for *all* the different properties of trust, where all other accounts seem to fail.¹⁴ The relevant properties are the following:

Seemingly Non-doxastic Properties

- a. We can trust at will.
- b. We can trust against our (subjective) evidence.
- c. We can trust therapeutically.
- d. Our trust can be betrayed; trust is not reliance.
- e. Trust does not survive too much reflective scrutiny.

Seemingly Doxastic Properties

- f. Trust that *S* will ϕ is incompatible with belief that *S* will not- ϕ .
- g. Trust in *S*'s testimony that *p* is incompatible with belief that not-*p*.
- h. Rational trust can generate (epistemically) justified belief.

¹⁴ Of course some people deny that some of these putative properties are genuine properties, I will do something to parry their objections in what follows. However, given that there is clearly no consensus here, the fact that there is a theory that can account for all these diverse (seemingly incompatible) features should be thought to be something people working on trust should take on board. This is the result I seek.

In §2 I show that *TG* can account for *fg*; In §3 that it can account for *abcde*. Throughout I compare *TG* with its competitors, and argue that it's the best theory that can account for all of *abcdefg*. Since *TG* is an instance of what I called a 'mixed' mental state pragmatism, its vindication should also support the genus position it is a species of. I end by considering what the right metaphysics of trust tells us about what are our reasons to trust (and thus what it says about *h*).

2. Doxasticism about Trust

Here is why I think people have sometimes been drawn to the idea that trust has at least some belief-like properties, or involves belief (even if it is not to be fully identified with it): Trust looks to be sensitive to counter-evidence. When you trust someone, you seem to represent them as having the properties of being *trustworthy*. And thus on attaining damning evidence that that the trusted party is not in fact trustworthy, our trust disappears. That's what explains, for instance, why we don't tend to trust people or institutions that have violated our trust, as per when we stop trusting a Chancellor of the Exchequer when we know she's recently committed fraud. To illustrate with a case from Richard Holton, suppose you are playing a game where you are required to let yourself fall and be caught by a fellow player. It seems appropriate to say here that you are required to *trust* the fellow player to catch you. But supposing you have overwhelming evidence that your fellow player won't catch you: Can you still trust your fellow player? It seems you cannot. Your having evidence against the fact that they will catch you seems to militate against (if not completely destroy) the possibility of trust. Further,

it seems infelicitous to say, for instance, that Amanda trusts James to feed her cat while she is on holiday while Amanda believes that James will not feed her cat. It seems infelicitous to say, for instance, that Roger *trusts* his teacher when she says that Henry VII lost the battle of Bosworth when he believes that Henry VII won that battle. And what can explain this other than the fact that trust has a cognitive, belief-like component?

In other words, only a doxastic account of trust (one where trust is belief or involves belief or has a belief-like component) can explain the following:

- f. Trust that *S* will ϕ is incompatible with belief that *S* will not- ϕ .
- g. Trust in *S*'s testimony that *p* is incompatible with belief that not *p*.

To see why, consider how non-doxastic accounts attempt to explain *f* and *g*. In particular, consider Richard Holton's very influential account according to which trust is a higher-order *stance* (a "participant" stance) with respect to the evidential standards our lower-order beliefs about the objects (individual people, groups, or institutions) of our trust are subject to, this combined with reliance on the individual (trust is not mere reliance, according to Holton, as per *d*). His idea is, roughly, that though we can't choose our beliefs (our beliefs are not under direct voluntary control) we can choose our higher-order epistemic stance toward them¹⁵. So that when we don't have a belief that our trusted friend is *not* trustworthy, we can choose to have a more lenient epistemic stance toward him, such that we can believe that our friend is innocent of a crime even when others would fail to believe it. And further we can choose whether or not we rely on him.

¹⁵ Explicitly echoing here van Fraassen's "stance voluntarism" (cf. van Fraassen 1982, 2002).

In other words, trust that S will φ is incompatible with belief that S will not- φ , but it is not incompatible with an absence of doxastic attitude toward whether S will φ , suspension of judgement as to whether S will φ , or indeed belief that S will φ . Holton in fact uses the case of the game above also as his opening example, motivating the view that we can trust *at will*. Remove the belief that the fellow player *will not* catch me, and it seems that we can *decide* whether or not we trust. This, as I mentioned, is feature a about trust that I think needs explaining. And if a is true, as we'll discuss in the next section, then features b , c and d need explaining too. And the reason that we can trust at will for Holton is that trust is a particular *stance*, a stance we can take up at will (absence certain constraints)¹⁶ and that we can choose on who we rely.

But the problem for Holton is whether he has really explained f and g . Holton claims he wants to avoid the conclusion that he has really given us a "recipe" for believing something we want to believe at will; a recipe like the following: "Find someone who will tell you it is true and trust them to speak truly and sincerely" (Holton 1994, p. 12). If his account has the consequence that belief is voluntary, then he thinks the account is deficient (since he thinks that belief is not voluntary). He says two things in his defence as regards his account entailing doxastic voluntarism: first, that there are limits on whom we can trust to do what, and "one of the constraints on *trusting* you to do something is the lack of a belief that you will not do it" (Holton 1994, p. 12, my emphasis). So I can't acquire (at will) the appropriate epistemic stance toward my friend (I can't choose to trust them, in other words) if I believe that my

¹⁶ As I mentioned, its being a *participant* stance involves the participants thinking it appropriate to have certain reactive attitudes on discovery of a betrayal of trust.

friend is untrustworthy. I can't trust him to give me back my book if I believe that he won't.

But the question here precisely is the following: how can trust can have such a constraint – be sensitive to whether I do *not* believe the person will be trustworthy – if trust is not a belief (or at least involves belief)? Just asserting that the constraint applies does not explain it¹⁷.

Second, Holton tells us that the recipe won't allow us to “directly, and immediately” acquire beliefs¹⁸, since (to put it bluntly) we'll have to find a friend to tell us the thing we want to believe. But Holton alludes to the idea that we can in principle “trust ourselves” and that at times we must. Why can't we tell ourselves what we want to hear, and then decide to trust ourselves? Won't that fit the formula? Nothing in Holton's account seems to preclude it, unless we have reason to not to believe that we will act in accord with our trust (but then how could we trust ourselves?). Nor does it help to appeal to the reliance part of Holton's account, as per:

When I rely on someone to do something, I work this reliance into my plans: I plan on the supposition that they will do it. But if I believe that they won't do it, then that too is something I should work into my plans. So if I rely you to do something but believe that you won't, I am led to incoherence: I have to work into my plans both the supposition

¹⁷ A related point is made by Keren: “supporters [of Holton's account] must explain why trusting a speaker to speak knowledgeably and sincerely, without believing that she will, should result in believing what she says” (Keren 2014, p. 2598); cf. also Hieronymi 2008.

that you will do it, and the supposition that you will not. (Holton 1994 p. 72)¹⁹

The problem is that according to Holton I can voluntarily change my higher-order epistemic stance (part of my taking a 'participant stance') toward whether someone will do something. This is why for Holton trust is not *just* reliance. So suppose that if I took a more lenient higher-order epistemic stance toward whether someone will do something, I will drop the belief that they will not do it. If this is known to me, then there is nothing that prevents me from trusting you to do something even though I believe you will not do it. All I need to do is modify my higher-order attitude so that I drop the belief that you will not do it: I am not 'led to incoherence'. Put differently, if part of why we can trust at will is that we can modify our higher-order epistemic stances at will, then we will have more voluntary control over trust than *f* and *g* predict – since having control over higher-order epistemic stances will mean having control (at least sometimes) over dropping first-order beliefs that *not-p*.

According to *TG* trust is a *p*-lief. And one *p*-lieves that *p* just in case ascribing belief that *p* is the proximate best way to explain a certain set of actions. It is clear that a straight doxastic account can explain *f* and *g*. Can *TG* explain them? It can, since if we already *know* that *S* believes that *not-p*, then an explanation of *S*'s actions that involves a belief that *p* cannot be the best explanation of *S*'s actions. It cannot even be the best *proximate* explanation, since even if belief *was* rightly to be thought of in pragmatist terms, then *S*'s

¹⁹ Thanks to [anonymous] for discussion.

belief that *not-p* would come with a set of actions that we know the positing of belief that *not-p* best explains.

To illustrate: when I trust that my friend is innocent of a crime, I can choose to act in such a way that my actions would be explicable only if the “belief” that my friend is innocent is attributed to me. But belief cannot be attributed in such a pragmatist manner (since we cannot believe at will), so we call this state *trust* (identical to what I’ve called *p-lief*) not *belief*. I could not trust that my friend is innocent of a crime if I believed that he is not innocent. That’s because if I believed that he was not innocent, my actions would either be such that ascribing “belief” that he was innocent would not be the best explanation of those actions, or else we would know something about me that would make it the case that ascribing “belief” that he was innocent would not be the best explanation of my actions.

I think the same issue arises, incidentally, for accounts according to which trust is (at least sometimes or partly) the attitude of acceptance that we discussed in opening (an especially interesting such account is found in Frost-Arnold 2014). According to Frost-Arnold, trust as acceptance must account for the intuition that trust that *p* is incompatible with the belief that *not-p*:

...the epistemic constraints on acceptance nicely accounts for this intuition; one cannot trust *B* to ϕ if one takes oneself to have overwhelming evidence that *B* will not ϕ , because one cannot accept that *p* if one takes oneself to have overwhelming evidence that *p* is not even approximately true. Frost-Arnold 2014 p. 1968.

But why think that acceptance has to have this constraint if acceptance is not a belief? Surely we can decide to act on known falsehoods (and we can have non-epistemic reason to accept a falsehood), so the fact that acceptance can directly lead to action (while this might distinguish it from supposition) is not enough to guarantee the applicability of the constraint. For example, I can accept that – despite the overwhelming evidence²⁰ – I will get better from my illness if accepting that I will get better will lead me to a course of action that will increase the probability (however small) that I will get better. And surely our acceptances can be incompatible with each other, at least across contexts; as Frost-Arnold admits: “sometimes we adopt a cognitive state of acceptance, which is not subject to the ideals of consistency and coherence across contexts” (Frost-Arnold 2014, p. 1965). If even acceptances need not be consistent (across contexts) then it seems hard to explain why acceptances need to be consistent with beliefs (which are not context dependent). Again just asserting that the constraint applies does not explain why it does.

3. Non-Doxasticism about Trust

The central problem with doxastic accounts of trust is that they cannot account for the claim that we can at times trust at will, in a way in which we cannot believe at will. This is data-point *a* which I mentioned any theory of trust needs to explain. And if it is the case that we can trust at will, then (as I’ve mentioned) an account of trust will also have to explain the following:

²⁰ Evidence good enough for knowledge.

- b. We can trust against our (subjective) evidence.
- c. We can trust therapeutically.
- d. Our trust can be betrayed; trust is not reliance

b is a corollary of *a*, since if we could not but trust in line with our (subjective) evidence, then trust would not be under any voluntary control²¹. We can only trust therapeutically – that is, trust someone we deem currently less than fully trustworthy such that they become trustworthy – if we can choose to trust. Slightly more controversially, since this assumes something like an ‘ought’ implies ‘can’ principle, it seems wrong to say that our trust can be betrayed (in some deontic sense of ‘betrayed’) if trust is not something we can control. A further feature of trust that makes trust look be something other than belief, but is not a corollary of *a*, is the following (also previously mentioned):

- e. Trust does not survive too much reflective scrutiny.

By ‘reflective scrutiny’ I mean here the kind of reflection aimed at settling (for the subject doing the reflection) the question as to whether to trust²². As Annette Baier puts it: “Trust is a fragile plant, which may not endure inspection of its roots, even when they were, before the inspection, quite healthy” (Baier 1987, p. 260). These two claims seem incompatible, for example: I trust James to return my book tomorrow. I am going to spend the evening wondering whether James is going to return my book tomorrow.

²¹ Here is a nice articulation of *b* from Paul Faulkner: “[T]rust need not satisfy either a positive or a negative evidence condition: it need not be based on evidence and can demonstrate a willful insensitivity to the evidence. Indeed there is a tension between acting on trust and acting on evidence that is illustrated in the idea that one does not actually trust someone to do something if one only believes they will do it when one has evidence that they will” (Faulkner 2007, p. 876).

²² This would be compatible with the subject reflecting on the evidence in order to find the best evidence in order to persuade *others* to trust. Thanks to [anonymous] for this point.

The most explicit doxasticist attempt to give an account of trust which can explain these features is Keren's. His view, which has aspects in common with Zagzebski 2012 and McMyler 2011, is the following:

A trusts B to phi only if A believes that B is trustworthy, such that in virtue of A's belief about B's trustworthiness, A sees herself as having reason to rely on B's phi'ing without taking precautions against the possibility that B will not phi, and only if A indeed acts on, or is responsive to, reasons against taking precautions. (Keren 2014, p. 2609).

Keren calls this a "preemptive reasons account" of trust. The thought is, very briefly, that – drawing on Raz's (1990) account of preemptive reasons for action – the reasons generated by trust are analogous to those generated by an authoritative command. This in the sense that an authoritative command gives us (according to Raz) a higher-order reason *against* acting for certain other reasons. In the case of trust, these latter reasons are reasons that prevent one from "taking precautions" against the possibility that the trusted party will not act as one "trusts", according to Keren. In other words, if you trust James (in Keren's analysis, you believe that James is trustworthy), you also fail to take precautions against the possibility that your belief that James is trustworthy is false – that is, you refrain from collecting and reflecting on available evidence relevant to that proposition.²³

I want to now show how *TG* fares better than the Preemptive Reasons Account in explaining *abcde*. I take it that a non-augmented doxastic account

²³ A similar account is offered both by Zagzebski (2012) – especially as regards speaker-trust – and McMyler (2011).

is going to have trouble explaining *abcde* and that Keren's account qua doxastic account can explain *f* and *g*. I have argued that *TG* can explain *f* and *g* in the last section.

Let's take *e* first.

e. Trust does not survive too much reflective scrutiny.

Keren's account can explain *e* because too much reflective scrutiny is a violation of the pre-emptive command.

TG can explain *e* because too much reflective scrutiny about whether *p* is incompatible with p-lief about *p*. If belief was to be determined on purely pragmatist grounds, then the fact of your constantly searching for new evidence as to whether *p* is best explained by an attitude of not yet believing that *p*.

Both accounts can explain *e*. The contest here is tied. So let's move on to *b*.

b. We can trust against the evidence.

The preemptive reasons part of Keren's account is precisely designed such that it can explain *b*. We can trust against the evidence on this account because we have *higher-order* reason that commands us to ignore any first order-evidence we might attain against *p* (and/or to fail to obtain such evidence).

The trouble is whether we might think that an account of trust must explain a slightly stronger claim:

b' We can trust against our total evidence.

By ‘total evidence’ I mean evidence that includes ‘higher-order’ evidence, or epistemic reason. *B'* is a corollary of *a* since if we can trust at will, then we can trust for reasons that go against epistemic reasons, whatever the order.

TG can explain *b* and *b'*. This is because we can trust at will according to *TG*. We can trust against the (total) evidence by acting in such a way that the best explanation of our actions is to ascribe a p-lief to us.

I think that gives *TG* an advantage over the preemptive reasons account. And I think it does so even if explaining *b'* is not a genuine desideratum for an account of trust. That’s because the way that the preemptive reasons account explains *b* is via an attempt to appeal to high-order evidence, and not via its ability to explain *a* and *thus* that it can explain *b*.²⁴ In fact Keren’s account cannot explain *a* – Keren himself does not attempt to. He does address *c*, however.

c. We can trust therapeutically.

Here he rules that the notion of therapeutic trust²⁵ is mistaken, or that it is far from obvious that there is such a phenomenon when taken to be different from merely acting as if one trusts. He asks us to consider the standard example of parents who trust their teenage daughter to look after the house in order to promote trustworthiness in her. He thinks the example does not give us a clear-cut example of therapeutic trust (where the latter involves absence of belief that the trusted party is trustworthy). First, because he thinks it is not clear whether we are not ascribing belief (or some degree of belief) to the

²⁴ Keren and Holton’s accounts have this feature in common.

²⁵ Frost-Arnold (2014) identifies what she calls “corrective trust” and “coping trust” as species of therapeutic trust. If she’s right that there are sub-versions of therapeutic trust, that further vindicates my claim that (*c*) and (*a*) are intimately connected.

parents when we ascribe trust to them. Second, because, to the degree that we think that the parents do fail to have the belief that the daughter is trustworthy, we might legitimately wonder whether they really *trust* their daughter, “or merely act *as if* they do” (Keren 2014, p. 2610).

Keren’s objection takes the form of a challenge here. He does not show that the notion of therapeutic trust is somewhat incoherent or unstable. Rather, he challenges us: present us with a putative case of therapeutic trust which cannot be analysed into either involving belief on the part of trusting party, or involving non-genuine trust – acting *as if* they trust. Take Holton’s game where you ‘decide’ to trust by allowing one self to fall and be caught by a fellow player behind you. In the absence of belief that the fellow player will catch you and you fall anyway, can you be really be said to genuinely trust, or merely to act as if you trust?

I think that *TG* can answer this challenge by showing how it relies on the faulty assumption that if the reason why one is predicating trust is to explain *actions* then the person on whom we are predicating trust cannot be *genuinely* trusting, but merely “acting as if one trusts”. And it’s enough to answer the challenge to allay Keren’s objection to therapeutic trust. Suppose that we are pragmatists about belief. If we are, then the line between acting as if one believes and genuine belief is going to involve a difference of degree, not kind. That is, doxastic pragmatists could make a distinction between merely acting as if one believes and genuine belief. However, to make sense of the phenomenon of acting as if one believes, they will appeal to the fact that one’s actions do not require the ascription of belief in order to make them rational.

The existence of the possibility of pretending to believe (or acting as if one believes) raises the bar as to what is required in order for belief to be ascribed. But this does not mean that one's beliefs are not a function of one's actions (internal or external), for the pragmatist. But according to the account I'm defending here, p-lief (and not belief) is a function of one's actions. Belief resists pragmatist treatment, since it is involuntary. What we are ascribing the parent, when we say that she therapeutically trusts her daughter is p-lief, not belief (according to this account). To trust is thus a species of *acting as if* – one trusts when one's actions are so consistent with belief that p , that belief that p becomes necessary to preserve the ascription of rational action (put differently, the fact of our beliefs not being under our voluntary control puts a break on a comprehensive principle of charity which demands assuming rationality). Like with pragmatism about belief, the difference between acting as if one believes that p , and one acts as if that p is going to be one of degree. It is then not an objection to the account if we are inclined to judge that the parent acts *as if* he believes that his daughter is trustworthy, but does not genuinely believe that she is. If he acts *as if* he believes her to be trustworthy comprehensively enough, then he p-lieves that she is trustworthy. That is, he trusts her. And he can p-lieve that she is trustworthy, even when there is (known) absence of belief that she is trustworthy.

The fact that Keren's explanation of b does not depend on its explanation of a , and the account's concomitant failure to explain a means that the Preemptive reasons account will not be able to explain:

d Our trust can be betrayed; trust is not reliance.

It's perhaps no surprise then that Keren himself does not attempt to show that his account can explain *d*. According to *TG* trust is not mere reliance because it is not an involuntary state like belief. We *choose* to put our trust in people (sometimes against the evidence that they are trustworthy) and thus we can rightly feel resentment when our trust is betrayed. However, we can, of course, also choose to rely on others – James can choose to rely on his new bookshelf to house his books – so the difference between reliance and trust cannot be just down to our ability to trust at will. The difference will be down to the content of respective attitudes of trust and reliance. In the former and not in the latter is someone *S* represented as being trustworthy (bookshelves are just not the sort of thing that can bear the property of being trustworthy). Any doxastic account has this last feature, but since belief is involuntary, it won't be enough for it to explain why we can rightly feel betrayed by breaches of trust.

5. Conclusion: Reasons to Trust?

What does the metaphysics of trust tell us are the reasons of trust if *TG* is true? According to both Jones (1996) and Keren (2014), what makes trust justified is determined in large part by what is the correct metaphysics of trust. I guess this is because it is assumed that if trust is a belief, and only if it is a belief, then trust is made rational or justified by epistemic reason/evidence. This is to assume that there can be no non-epistemic reasons for belief, and no epistemic reasons for actions, and I'm not sure we can assume these claims for free. Nevertheless, assuming them to be true, if trust

is a p-lief then what kind of reason justifies p-lief? Or, put differently, what kind of reason do pragmatists about belief think justifies belief: practical reason or epistemic reason (evidence)? It seems to me that nothing prevents a pragmatist about belief from thinking that evidence solely justifies belief. Even if they end up accepting that there are times when we can believe at will (by acting in such a way that ascribing belief would be the only way to explain those actions) and as such that belief can be sensitive to non-epistemic considerations, beliefs arrived at in this way can be thought by the pragmatist to be always unjustified if they are at odds with the available evidence. As I mentioned in section 1, the fact that belief is a state whose ascription is there to explain our actions does not mean that belief is itself an action, solely evaluable in non-epistemic terms.

According to *TG*, however, one can trust against the evidence, and trust at will²⁶. If the proponent of *TG* holds that only evidence can justify p-lief, she will be forced to concede that such trust is always unjustified. And this seems wrong. Trust – against the evidence – seems not only possible, but justified on occasion. But under *TG* nothing rules out the possibility that the reasons for trust are *all things considered reasons* and that epistemic consideration can sometimes compete (sometimes winning, sometimes losing) with non-epistemic considerations in our deliberations as to whether to trust. Our judgement about whether one ought to trust (because determined by our judgement as to whether one ought to trust *all things considered*) is then determined by the particular pushes and pulls of the different kinds of reason

²⁶ That is, that we can at least sometimes trust at will (thanks to Rowland Stout for the clarification).

in any given situation. So one cannot give a general prescription as to what kinds of reason justify trust – the answer is simply that sometimes epistemic reasons do, and sometimes non-epistemic reasons do.

Can epistemically justified belief be generated by states that are justified by non-epistemic reasons? I personally see no reason why that cannot be the case. However, even if justified belief can be generated only by states justified by epistemic reason, this does not mean that *TG* is incompatible with *h* – since according to *TG* epistemic reasons are sometimes the reasons for trust, and so trust can generate epistemically justified belief, even if the latter can only be generated by epistemically justified states.

To conclude, then, I think that *TG* can handle all of *abcdgh* better than any currently proposed alternative. This means that *TG* can handle the central paradox with respect to trust, nicely summarised by Baker with this:

Someone might try to distinguish trust from genuine or full belief. Trust, on such a view, would be a watered down variant of belief, something like pretence or acting-as-if something were true. But this is to view trust as a non-serious form of belief. Whereas what one demands from one's friends is belief, not pretence, that one is innocent. And what some outsiders find amazing is just the fact that serious belief continues in the face of rising evidence against it (Baker 1987, p. 9).

What I take to be the basic insight underscoring *TG* is that one can claim that trust is something slightly different from ordinary belief, but yet is not just pretence or “acting-as-if” something were true: in a pragmatic picture of our

mental states we can still make a distinction between one's believing that *p*, and acting as if one believes that *p*. The difference between the two might not be a difference in *kind* but it is a difference nonetheless. As such, p-lief is no less "serious" than belief. It's a very serious and difficult²⁷ demand to make of our friends that they p-lieve us to be innocent, but when we heed it we can with effort maintain it in the face of rising evidence against it – no matter what some of our very private thoughts may be.

REFERENCES

- Adler, J. 1994: "Testimony, trust, knowing" *Journal of Philosophy* **91(5)** pp. 264 – 275.
- Alston, W. 1989: "The Deontological Conception of Epistemic Justification" in his *Epistemic Justification: Essays in the Theory of Knowledge* (Cornell University Press).
- Anscombe, G.E.M. 1957: *Intention* (Cambridge, MA: Harvard University Press) (this version published in 2000).
- Baier, A. 1987: "Trust and Anti-Trust" *Ethics* **96** pp. 231 – 260.
- Baker, J. 1987: "Trust and Rationality" *Pacific Philosophical Quarterly* **68** pp. 1 – 13.
- Bennett, J. 1990: "Why is belief involuntary?" *Analysis* **50(2)** pp. 87 – 107.
- Bodgan, R. 1986: "The manufacture of belief" in Radu Bogdan (ed.) *Belief: Form, Content, and Function* (Oxford: Clarendon Press).
- Boghossian, P. 2003: "The normativity of content" *Philosophical Issues* **13** pp. 31 – 45.
- Booth, A.R. Forthcoming: "Belief is Contingently Involuntary" *Ratio*.
- Braithwaite, R. 1967: "The nature of believing" in A.P. Griffiths (ed.) *Knowledge and Belief* (Oxford: Oxford University Press).
- Bratman, M. 1992: "Practical reasoning and acceptance in a context" *Mind* **101** pp. 1 -15.
- Cohen, L.J. 1992: *An Essay on Belief and Acceptance* (Oxford: Clarendon Press).
- Domenicucci, J. & Holton, R. forthcoming: "Trust as a Two-Place Relation" in

²⁷ A more serious demand to make than merely "accepting" that the friend is innocent (cf. Frost-Arnold) and so better qualifies for what Hieronymi (2008) calls the demand for "fully fledged trust" (see also McGeer 2008).

- Faulkner and Simpson (eds.) *New Perspectives on Trust*.
- Faulkner, P. 2007: "On telling and trusting" *Mind* **116** pp. 875 – 902.
- . 2011: *Knowledge on Trust* (Oxford: Oxford University Press).
- Frankish, K. 2007: "Deciding to Believe Again" *Mind* **116** pp. 523 – 527.
- Frost-Arnold, K. 2014: "The Cognitive Attitude of Rational Trust" *Synthese* **191** pp. 1957 – 1974.
- Hardin, R. 2004: *Trust and Trustworthiness* (New York: Russell Sage Foundation).
- Hawley, K. 2014: "Trust, Mistrust, and Commitment" *Noûs* **48(1)** pp. 1-20.
- Hedden, B. 2015: "Believing and acting: voluntary control and the pragmatic theory of belief" *Logos & Episteme* **6(4)** pp. 495 – 513.
- Hieronymi, P. 2008: "The reasons of trust" *Australasian Journal of Philosophy* **86(2)** pp. 213 – 236.
- Holton, R. 1994: "Deciding to Trust: Coming to Believe" *Australasian Journal of Philosophy* **72** pp. 63 – 76.
- Jones, K. 1996: "Trust as an affective attitude" *Ethics* **107(1)** pp. 4 – 25.
- Kappel, K. forthcoming: "Believing on trust" *Synthese*.
- Keren, A. 2014: "Trust and belief: a Preemptive Reasons Account" *Synthese* **191** pp. 2593 – 2615.
- Lewis, D. 1974: "Radical Interpretation" *Synthese* **27** pp. 331 – 44.
- McGeer, V. 2008: "Trust, Hope, and Empowerment" *Australasian Journal of Philosophy* **86(2)** pp. 1 – 18.
- McHugh, C. 2014: "Exercising Doxastic Freedom" *Philosophy and Phenomenological Research* **88(1)** pp. 1 – 37.
- McLeod, C. 2002: *Self-trust and reproductive autonomy* (Cambridge: Cambridge University Press).
- McMyler, B. 2011: *Testimony, trust and authority* (Oxford: Oxford University Press).
- Nickel, P. 2010: "Voluntary Belief on a Reasonable Basis" *Philosophy and Phenomenological Research* **81** pp. 312 – 334.
- Peels, R. 2015: "Believing at Will is Possible" *Australasian Journal of Philosophy* **93** pp. 524 – 541.
- Raz, J. 1990: *Practical Reasons and Norms* (Princeton: Princeton University Press).
- Setiya, K. 2008: "Believing at Will" *Midwest Studies in Philosophy* **32** pp. 36 – 52.
- Shah, N & David Velleman 2005: "Doxastic Deliberation" *The Philosophical Review* **114(4)** pp. 497 – 534.
- Simpson, T. W. 2012: "What is Trust?" *Pacific Philosophical Quarterly* **93(4)** pp. 550 – 569.
- Schwitzgebel, E. 2002: "A Phenomenal, Dispositional Account of Belief" *Noûs* **39(2)** pp. 249-75.
- Smith, M. 1987: "The Humean Theory of Motivation" *Mind* **96** pp. 36 – 61.
- Stalnaker, R. 1987: *Inquiry* (Cambridge, MA: MIT Press).
- Steglich-Petersen, A. 2009: "Weighing the aim of belief" *Philosophical Studies* **145** pp. 395 – 405.
- Thompson, C. (ms): "A Reductive Account of Trust".
- van Fraassen, B. 1984: "Belief and the Will" *Journal of Philosophy* **81** pp. 235 – 256.

- , 2002: *The Empirical Stance* (New Haven: Yale University Press).
- Zagzebski, L. 2012: *Epistemic Authority: A theory of trust, authority and autonomy in belief* (Oxford: Oxford University Press).