

'Trusting-to' and 'Trusting-as': A qualitative account of trustworthiness

Joshua Kelsall

To cite this article: Joshua Kelsall (2022): 'Trusting-to' and 'Trusting-as': A qualitative account of trustworthiness, Inquiry, DOI: [10.1080/0020174X.2022.2075456](https://doi.org/10.1080/0020174X.2022.2075456)

To link to this article: <https://doi.org/10.1080/0020174X.2022.2075456>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 24 May 2022.



Submit your article to this journal [↗](#)



Article views: 412



View related articles [↗](#)



View Crossmark data [↗](#)

'Trusting-to' and 'Trusting-as': A qualitative account of trustworthiness

Joshua Kelsall^{a,b}

^aUniversity of Stirling, Stirling, UK; ^bPhilosophy, University of St Andrews, St Andrews, UK

ABSTRACT

Philosophical accounts of trustworthiness typically define trustworthiness as an agent being reliable in virtue of a specific motivation such as goodwill. The underlying thought motivating this view is that to be trustworthy is to be more than merely reliable. If motivational accounts are correct, this is a problem for non-motivational accounts of trustworthiness, as motivations are not required for trustworthiness. In this paper, I defend the non-motivational approach to trustworthiness and show that the motivational approach is inadequate. I do this by making a novel distinction between *trusting-to* and *trusting-as* relations. A *trusting-to* relation is a relation in which a trustor 'X' trusts the trustee 'Y' to do something. *Trusting-as* relations are an overlooked relation implicit in all *trusting-to* relations. They describe the social relationship that holds between X and Y. I will argue that *trusting-as* relations determine whether any specific motivations are required for trustworthiness *trusting-to* relations. Thus, I show that acknowledging *trusting-as* relations enables us to provide a satisfactory explanation of the motivation intuition without making specific motivations constitutive features of trust.

ARTICLE HISTORY Received 28 October 2021; Accepted 22 February 2022

KEYWORDS Philosophy of trust; trustworthiness; motivations; commitments

1. Introduction

Philosophical accounts of trustworthiness typically define trustworthiness as an agent being reliable in virtue of a specific motivation (Baier 1986; Hardin 2002; Jones 1996). For example, on Baier's account of trustworthiness, an agent is trustworthy if they are reliable in virtue of bearing their trustors goodwill (Baier 1986, 234–235). These *motivational accounts* of trust are appealing because they explain the intuition that there is more

CONTACT Joshua Kelsall  jjck1993@outlook.com, jjck@st-andrews.ac.uk  University of St Andrews, 5 The Scores, St Andrews KY16 9AL, UK

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

to being trustworthy than being merely reliable. I call this the *motivation intuition*. In contrast are *non-motivational accounts* of trustworthiness such as Hawley's commitment account (2014, 2019). A criticism of non-motivational views, which Hawley herself acknowledges, is that they seem ill-equipped to explain the intuition that trustworthiness is more than mere reliability (2019, 19–20). In this paper, I argue that this criticism of non-motivational accounts rests on a mistake about the nature of trust.

I will argue that the mistake is due to a narrow analysis of trust as a three-place relation of the kind: 'x trusts y to Φ '. I call these *trusting-to* relations because they are relations in which the trustor 'x' trusts the trustee 'y' to do something ' Φ '. I will argue that implicit in every trusting-to relation is what I call a *trusting-as* relation. The trusting-as relation refers to the particular social relationship that holds between x and y. A paradigm example of such a relationship is friendship. I will argue that it is these implicit trusting-as relations that determine whether any specific motivations are required for trustworthiness three-place trust relations. In consequence, I will show how acknowledging trusting-as relations enables us to provide a satisfactory explanation of the motivation intuition without making specific motivations constitutive features of trust.

This paper has three aims. (1) To defend the non-motivational approach to trustworthiness. (2) To highlight the inadequacies of the more popular motivational approaches. (3) To highlight an overlooked feature of trust, the trusting-as relation, which is not only implicit in all trusting-to relations but has a significant bearing on both what it means to trust in any standard three-place trust relation, and to be trustworthy.

This paper is structured as follows. In Section 2, I provide the impetus for advancing a motivational account of trustworthiness. I will do so by introducing cases in which a trustee in a trust relation is reliable, but only in virtue of a morally suspect motivation. Such cases are meant to trigger the motivation intuition that to be trustworthy requires one to have a morally praiseworthy motivation. In Section 3, I introduce trusting-as relations, distinguish them from trusting-to relations, and show how they determine (1) what it means to trust a trustee and (2) what it means for a trustee to be trustworthy. In Section 4, I demonstrate how trusting-as relations allow supporters of non-motivational accounts of trustworthiness to explain the motivation intuition without making specific motivations constitutive features of trustworthiness. In Section 5, I will argue that a significant implication of recognising trusting-as relations is that motivational accounts of trustworthiness of any stripe will fail to provide an adequate analysis of trustworthiness.

2. The motivation intuition

In this section, I motivate the intuition that specific motivations are constitutive features of trustworthiness. In thinking about what it means to be trustworthy, one necessary element is reliability. If I tell you that a colleague is a trustworthy administrator, then part of what I express is that you can rely on them to carry out their administration duties. If I say that someone is a trustworthy truth-teller, then part of what I express is that they reliably tell the truth. Despite this, philosophers argue that trustworthiness is more than mere reliability (Hawley 2012; 2014; Baier 1986; McLeod 2015). One can motivate this view by looking at cases in which an agent is reliable and yet seemingly not trustworthy. If we accept our intuitions in these cases, then it is argued that trustworthiness is reliability plus something else, and it is this something else that is used to distinguish being trustworthy from being merely reliable.

One popular type of account of this extra something is a motivational account of trustworthiness. On a motivational account, to be trustworthy is to be reliable in virtue of having a specific motivation. Different motivational accounts require different motivations, but perhaps the most popular motivational account is Baier's *goodwill* account (Baier 1986; 1991). Subsequent philosophers have amended certain elements of Baier's view, but still, take goodwill to be the motivation that makes an agent trustworthy (Jones 1996; McGeer 2008). For the purposes of this paper, we'll take goodwill to be our paradigm motivational account, so I will briefly explain it before explaining why we might think that trustworthiness requires a specific motivation in the first place.

Baier argues that when we trust others we do not simply rely on their more or less dependable psychological habits, but instead, we depend on their goodwill towards us (1986, 234–235). If x bears y goodwill, then x cares about y 's interests intrinsically, that is *because* they are y 's interests. If you make good on someone's trust because it will further your self-interest in some way, then your motivation is not one of goodwill. Thus, while you may be reliable you are not trustworthy. If you are motivated to be trustworthy in virtue of the fact that someone is counting on you, then you demonstrate that you care about that person's interests intrinsically, which is to have goodwill. Thus, you are both reliable and trustworthy.

Why suppose that trustworthiness requires a specific motivation such as goodwill? Why isn't being trustworthy simply a function of a person's reliability? The underlying thought is that trustworthiness is something

praiseworthy in a morally significant sense while reliability is not. We can strengthen the thought by looking at different kinds of cases that appear to support it. In the first case, an agent is reliable but because their motivation is morally suspect, we are reluctant to call them trustworthy. The second and third cases contrast two reliable agents: one well-motivated and one ill-motivated. By *well-motivated* I mean that they have a positive motivation in the moral sense, while *ill-motivated* means that they have a negative motivation in the moral sense. In this case, the well-motivated agent is slightly less reliable than the ill-motivated one, yet intuitively we would still want to say that the ill-motivated agent is not more trustworthy than the well-motivated one.

False Friend: John believes that Charlie is one of his closest friends. Because of this, John shares a personal secret with Charlie and Charlie promises that he will keep the secret. Unbeknownst to John, Charlie does not actually like John at all, and his sole motivation for keeping the secret is because he wants to curry favour with John, who is wealthy and well-connected. If John lacked these qualities, then Charlie would not keep the secret.

False Friend is a case of reliability. Charlie can be relied upon to keep John's secret, but it is less clear that it is a case of trustworthiness. Suppose that John discovered Charlie's true motivations. Not only would he be deeply hurt by Charlie's falseness, but he would be reluctant to trust Charlie in the future. Given that our aim in trusting is primarily to trust the trustworthy¹ we can take this as indicative of John no longer taking Charlie to be trustworthy. If we think that John is correct in his assessment, then we have a case where an agent is reliable but not trustworthy. Therefore, if a theory of trustworthiness is to explain this case it will have to include more than mere reliability. The goodwill account could explain the intuition that Charlie is reliable but not trustworthy. Accepting this account, one would say that because Charlie is motivated by the desire for reputation and not motivated by an intrinsic concern for John and his interests, he lacks goodwill toward John and so he is not trustworthy even if he is reliable.

Now let's consider two further cases that support the motivation intuition:

Reliable Minion: Count Dracula has a faithful servant, Renfield, who will do anything for him. Renfield is motivated by fear of Dracula and desire for

¹This isn't always the case, as therapeutic trust shows. Though it is open to debate whether therapeutic trust counts as genuine trust. For analysis of therapeutic trust and critical discussion see (Horsburgh 1960; McGeer 2008; Hieronymi 2008).

immortality. His fear and desire are so strong that his motivation to be reliable can never be defeated. Therefore, he will always strive to do whatever his master commands.

Fallible Friend: Dr Van Helsing is the faithful friend of Mina and Jonathan Harker. He will do anything that he can to help them in their fight against Dracula. Nevertheless, Van Helsing could be persuaded under the right circumstances to turn his back on his friends. This scenario is highly unlikely, however, and Van Helsing's motivation is generally strong enough that he will nearly always stand by his friends.

The difference between Renfield and Van Helsing in these cases is a difference between what McGeer and Pettit would call the 'durability' of their trustworthiness (2017). McGeer and Pettit understand trustworthiness as a capacity to respond to other people's dependence, and they note that this capacity has two components: durability and dependability (McGeer and Pettit 2017, 16–19). The durability of trustworthiness describes the resistance one has to disrupters to one's capacity for trustworthiness. If I am initially responsive to your trusting me but easily swayed by fear or a bribe, then my trustworthiness is not very durable. The dependability of trustworthiness describes the number of situations where I am positively responsive to your trusting me at all. The difference can be illustrated by the example of a strongman who is terrified of mice. The strongman has a dependable capacity to lift heavy weights since he has the muscle and motivation to do so when called upon in most situations. However, his capacity can be disrupted in situations where mice are present since fear overtakes him, and he will run away. In both *Reliable Minion* and *Fallible Friend*, our trustees are dependably trustworthy because they would both respond positively to the dependence of their respective trustors. However, Renfield's trustworthiness is more durable because it cannot be defeated, whereas Van Helsing's trustworthiness is less durable because it can be defeated. If all there is to trustworthiness is being reliable, then Renfield is more trustworthy than Van Helsing.

If we think that something is going awry in these cases, that Van Helsing is somehow more trustworthy than Renfield despite his fallibility, then we can use motivations to explain why. On Baier's goodwill account, to be trustworthy is to be reliable in virtue of goodwill. Van Helsing has that goodwill. He would not count as maximally trustworthy because he is, after all, fallible. Nevertheless, he would count as more trustworthy than Renfield because on a goodwill account, Renfield would fail to be trustworthy altogether because he lacks the motivation of a trustworthy

person. Thus, Renfield is reliable, but he is not more trustworthy than Van Helsing even though the latter is less reliable.

In this section, I have presented three cases which motivate the motivation intuition and, in consequence, motivational accounts of trustworthiness. In *False Friend* we had a case where a trustee is reliable but ill-motivated. In *Reliable Minion* and *Fallible Friend*, we had two contrasting cases in which the more reliable agent is ill-motivated, and the less reliable agent is well-motivated. In all the cases, there seems to be something wrong with calling the ill-motivated but reliable agent trustworthy, and especially in cases such as *Reliable Minion* and *Fallible Friend*, where the ill-motivated agent turns out to be more trustworthy than the well-motivated but fallible agent. The lack of a trustworthy motivation can be used to explain what is going wrong in these cases. Using Baier's goodwill account we were able to show that the ill-motivated agent is not trustworthy despite their reliability, because they lack the motivation of a trustworthy agent.

The strength of motivational accounts such as Baier's is that they can explain why agents who are reliable but ill-motivated are not trustworthy.² The problem with a non-motivational account is that they seem incapable of doing this. I'll conclude this section with a brief example case to illustrate the point. I will look at the main competing account of trust to the motivational accounts: Hawley's *Commitment Account* of trustworthiness (Hawley 2014; 2019).

On Hawley's account, both trust and trustworthiness are forms of reliance and reliability, but instead of relying on or being reliable in virtue of having a specific motivation, one relies on, or is reliable in virtue of keeping, one's commitments. To be trustworthy for Hawley involves both the ability to keep the commitments that one incurs but also a judiciousness in selecting one's commitments (2019, 73–74). A trustworthy person is one who has the intention and competence to keep their incurred commitments, and the wherewithal to avoid making commitments that they will not be able to keep. However, what is not required on the account is that a trustworthy person has any specific motivation when making and keeping commitments. In consequence, the commitment account seems incapable of explaining the

²To some extent this is, of course, dependent on which motivation is taken to be the trustworthy motivation. As we'll see in Section 3, Hardin's alternative motivational account of trust is actually incapable to make these distinctions because of the motivation that is chosen. The point here then, is not that all specific motivational accounts can solve this problem, but rather that, in taking a motivational approach in general, one has the tools to solve the problem. A non-motivational account on the other hand, as we will see with Hawley's commitment account, appears unable to solve the problem.

cases introduced in this section. Charlie is trustworthy because he meets his commitment to keeping John's secret, irrespective of his selfish motivation. Renfield is more trustworthy than Van Helsing because Renfield will be more likely to keep the commitments he makes due to the durability of his reliability. Hawley acknowledges this motivational objection to her commitment account in her work, though she argues that motivational accounts are susceptible to their own objections which her account can avoid (2019, 16–20). In this paper, I won't be assessing these objections, nor will I be defending the commitment account outright. Instead, my aim is to show that motivations are not constitutive features of trustworthiness, and that non-motivational accounts can explain the motivation intuition without making specific motivations constitutive features of trustworthiness.

3. The difference between 'Trusting-to' and 'Trusting-as'

In this section, I provide the foundations for my non-motivational explanation of the motivation intuition. To do this, I must turn our attention away from analyses of trustworthiness and towards analyses of trust. Analyses of the latter often influence the former insofar as trustworthiness is generally understood in terms of an agent's being fit for or, as the name implies, 'worthy' of trust. For example, if trust is defined as reliance on goodwill, then (roughly speaking) trustworthiness is understood to be reliability in virtue of goodwill. In this section, I make a distinction between 'trusting-to' and 'trusting-as' relations. While trusting-to relations are often analysed in the literature, trusting-as relations have been overlooked despite being implicit in all trusting-to relations. I will show how trusting-as relations contribute to the meanings of what it means to trust and be trustworthy. Then, in Section 4, I will show how trusting-as relations allow us to give a non-motivational explanation of the motivation intuition.

Trusting-to relations refer to the typical three-place analysis of trust: x trusts y to Φ (Hardin 2002, 9; Holton 1994, 67; McGeer and Philip 2017, 159; McCraw 2015, 416; Baier 1995; Faulkner 2015, 2007; Jones 1996; Zagzebski 1996, 2012). In such relations, ' x ' stands for the agent who trusts, ' y ' stands for the agent trusted by x , and ' Φ ' is the thing that y is trusted to do. Although some analyse trust as a two-place or even one-place relation (Domenicucci and Holton 2017; Faulkner 2015), the dominant approach is to analyse trust as a three-place relation. In my view, analysing trusting-to relations taken in isolation is problematic for two reasons. Firstly, it

decontextualises trust from the actual contexts in which trust occurs. This in turn leads to the second problem, which is that in decontextualising trust we are incapable of giving a complete analysis of what it means to trust and be trustworthy in specific cases.

Every three-place trust relation involves at least two agents. A trustor and trustee.³ Just as in the case of trustworthiness where trust is identified as reliability plus some extra features, so too is trust understood as a form of reliance plus some extra features. Trust is typically taken to have two distinct features that distinguish it from mere reliance. Firstly, when we trust, we have heightened expectations of our trustees such that we expect them to be more than merely reliable (Hawley 2012, 5). Secondly, when we trust, we are disposed to react to the failing of our trust with feelings of betrayal, or gratitude in the cases where our trust succeeds (Hawley 2012). It will be important for understanding trusting-as relations to understand how different accounts of trust explain these two things, so I will now show how the goodwill and commitment accounts of trust do this.

On Baier's goodwill account, the heightened expectations that we have of our trustees are that they will bear us goodwill (1986, 234). We expect our trustees to care intrinsically about us and our interests, and we expect them to be motivated out of this concern for us when making good on our trust. When trust fails, we feel betrayed because the trustee reveals that they do not actually bear goodwill towards us, that we are of no intrinsic concern to them, and that our dependence on them isn't enough to motivate them to make good on our trust. Margalit, who adopts a similar goodwill approach to trust, captures this harm of betrayal in terms of a revelation that the trustee does not bear the trustor goodwill with the following:

The true insult in betrayal is the discovery by the betrayed that he wasn't at all on the mind of the betrayer. Indeed, what was done was not directed against him but instead was done with utter disregard for him. The maddening indifference is what hurts ... It is not the betrayed's interests that are ignored. He is ignored ... the betrayed realizes that he is not the significant other and not special. (Margalit 2017, 112)

However, why suppose that it is the expectation that someone would bear you goodwill that causes betrayal? On Hawley's commitment account, you do not expect your trustee to bear you goodwill, you

³We can of course talk about group trust, but for the sake of simplicity, I will focus on trust relations that occur between two individual agents as trustor and trustee.

expect them to keep a commitment that they have made. Regardless of the trustee's motivations, it is the breaking of a commitment that results in a betrayal. Suppose that a Christian vows to give a beggar charitable aid in his hour of need. Suppose also that this Christian does not bear the beggar any goodwill; she will be charitable because she sees it as her Christian duty to do so. The hour of need arrives, but the Christian turns the beggar away leaving him in dire straits. On the commitment account, this is betrayal because the Christian made a commitment which she subsequently broke. The harm done here in betrayal is the harm of leading the trustor to believe that they could count on the trustee, only for the trustee to fall short when the time comes. Defining trust in this way still allows us to distinguish mere reliance from trust. Inanimate objects cannot make commitments, so they cannot betray and thus cannot be trusted. Equally, you can rely on a person to do something without necessarily trusting them. Baier's own example of Kant's neighbours (1986, 234), who relied on the regularity of his walks to tell the time, could not complain of betrayal if he failed to be on time because Kant did not commit to being their timekeeper. The heightened expectations of trustors are also different from the expectations of those who merely rely. On a commitment account, you expect your trustee to recognise that they have a commitment to make good on your trust and make good on that commitment. When you rely on a person or object, you merely need to predict that they or it will be reliable.⁴

Any satisfactory account of trust will explain the difference between trust and mere reliance. It will do so by explaining why trust seems to involve specific expectations that are not present in cases of reliance and explain why failed trust results in feelings of betrayal. I will show in this section that one can neither provide a full account of the heightened expectations of trustors nor the conditions for betrayal without referring to trusting-as relations. But first, we need a complete picture of what trusting-as relations are and how they differ from the standard three-place trusting relations.

All trust relations of the form 'x trusts y to Φ ' exist in the context of a meta-relationship. By this, I mean that whenever trust is established, it is established within the context of a broader social relationship holding between the trustor and trustee. In these meta-relationships,

⁴Some define the difference in expectations between trust and mere reliance as a difference between normative expectations and having merely predictive expectations (Jones 2004). The thought being that when you trust, you think that your trustee ought make good on the trust, whereas when you rely, you merely predict that the relied upon object/person will act as expected.

the trustor and trustee play a specific social role with respect to each other, and these social roles then shape the heightened expectations of the trustor towards the trustee, and they set some of the conditions for the trustee's being trustworthy. These social roles and meta-relationships are, simply put, the typical social roles and relationships that we find ourselves in. Social roles include being a friend, lover, acquaintance, colleague, teacher, pupil, stranger, politician, doctor, man, woman, etc. The meta-relationship describes the interplay between the social roles played by the trustor and the trustee. In some cases, the social roles will be mutual, and thus the meta-relationship will be the same no matter who is the trustor or the trustee. For example, if the social role played by x and y is that of friends, then the meta-relationships arising in the trust will be 'x trusts y as a friend' or 'y trusts x as a friend'. On the other hand, sometimes the interplay between the social roles will be different. For example, the social roles played by Dracula and Renfield as master and slave, respectively. Thus, when Dracula trusts Renfield, he is trusting Renfield as a slave, whereas when Renfield trusts Dracula, he trusts him as his master. In general, the thought is that for any iteration of 'x trusts y to Φ ', there is an implicit trusting-as relation that, when included gives us 'x trusts y to Φ , as a Ψ '.

It is my view that no three-place trust relation is without an implicit trusting-as relation. Perhaps the most likely counterexamples to my view are cases of anonymous trust such as we find in 'trust game' experiments. In these experiments, participants play an anonymous game in which they may either trust one another, cooperating to maximise their shared rewards, or one player can defect and retain a reward for themselves while the duped player gets nothing (Ostrom and Walker 2003). I would argue that there is still a meta-relationship at play in this game; both players are anonymous strangers. In fact, researchers sometimes use this very relationship to make predictions about the outcome of such trust games, for example, that because of the anonymity of the relationship, participants will be less likely to trust one another and more likely to defect (Hardin 2003, 88). Thus, not only does there seem to be an argument that even in anonymous cases, the very anonymity creates a meta-relationship of strangers but this meta-relationship is taken to have important implications for the outcomes of players' behaviour.⁵ In any case, for the purpose of undermining the motivational

⁵Trusting-as relations may be used to justify a specific trusting-as relation, or even a lack of one. The trust-game between strangers is an example of how people might use their relationship as strangers to justify a lack of trust to cooperate in the game. On other hand, a friendship might be used to justify a

accounts of trust, it will not undermine the argument here even if it turns out that there is some small number of trust relations which do not have meta-relationships. It will be enough to undermine the motivational accounts so long as one accepts that in most instances of trust there exists a meta-relationship between the trustor and the trustee, and that when they occur, it is these meta-relationships that determine whether any specific motivations are required, rather than trust itself. In the unlikely situation in which there was a trust-relation without a meta-relationship, motivations would simply be irrelevant.

One might object to the notion of trusting-as relations as distinct from trusting-to relations. Of course, trusting someone as a friend is more complex than trusting someone to look after a handbag, but just as Hardin (2002, chp.3) argues that two-place trust is a shorthand way of talking about multiple three-place trust relations, so too might trusting-as relations be helpful shorthand for talking about multiple trusting-to relations. In that case, when I say, 'I trust you as a friend' we can translate the statement into 'I trust you to perform actions x, y and z'. I have two lines of defence against this argument. Firstly, it isn't clear that one could reduce one's being trusted as a friend to a specific set of actions. Indeed, trusting someone as a friend might involve more than expecting your friend to do certain things; you may also expect them to do things for certain reasons. For example, given that friendship is often defined as requiring a mutual goodwill between friends (Telfer 1970–1971; Annas 1977; Annis 1987; Cocking and Kennett 1998; White 2001; Helm 2021) then we might argue that a part of being a trustworthy friend is acting out of goodwill.^{6,7} This response doesn't get us very far, however, since one can just rebut with the argument that we can still reduce trusting-as relations to include specific motivations as well as specific actions.

positive trusting-to relation. Suppose someone asks John why he trusts Charlie to keep his secret. John might reply 'because he is my friend'. It is worth noting however that these justifying relations can be defeated by other considerations. For example, if John knows that Charlie is a compulsive blabbermouth, then even if John does trust Charlie in general as a friend, this won't justify trusting Charlie to keep a secret, since it will be defeated by the consideration that Charlie is a blabbermouth. Equally, suppose that Susan is someone who has only ever had extremely successful experiences of trusting strangers; she might then take stranger trust to be a sound justification to trust the other players in the trust game to cooperate. Thank you to an anonymous reviewer for bringing this point to my attention.

⁶The term 'goodwill' has been added; in fact, the more general way that this mutual goodwill is defined is as a mutual caring of friends for their own sake, which is a form of goodwill, since to have goodwill is to care about someone for their own sake.

⁷This isn't to say that this is the only condition for friendship of course, but to say that friendship requires things that go beyond actions.

My main response to the reductive account of trusting-as relations is to accept that a reductive account may be possible. I can do this because the essential point about trusting-as relations is not whether they can be wholly distinguished from trusting-to relations, but rather that these trust relations are a kind of higher order trust relation, which have significant impacts on lower order three-place trust relations. The idea is that when John trusts Charlie to keep a secret, this three-place trust relation is embedded in a second-order trust relation, a trusting relation between friends. And this second-order trust relation has an impact on what John expects of Charlie. Suppose again that to trust someone as a friend is to trust that one's friend will act out of goodwill. In that case, John trusts Charlie to keep his secret out of goodwill because they are friends. And when John discovers that Charlie is wrongly motivated, this causes the trust relation to break down. John no longer trusts Charlie to keep his secret, not because Charlie is incapable of keeping his secrets (he does keep it), but because Charlie has failed to keep it for the right reasons. He has failed to satisfy the implicit trusting-as relation (being trusted as a friend) in which the three-place trust is embedded.

Now contrast *False Friend* with our *Christian* case. Suppose now that the Christian does help the beggar, but that she does it out of duty and not because of any goodwill towards the beggar. The beggar knows that the Christian bears him no goodwill, but he knows that she is devout and therefore motivated to keep her vow out of duty and her devotion to God. It seems that the beggar can trust the Christian. He does not trust her as a friend, he trusts her as a Christian. What this means is that the beggar expects the Christian to act in accordance with the dictates of her faith and be motivated by that faith. The beggar does not need to expect the Christian to bear him any goodwill to trust her.

False Friend and *Christian* demonstrate how different meta-relationships shape both the normative expectations of trustors and play a role in determining whether the trustor feels betrayed when trust is let down. In *False Friend*, the meta-relationship is one of friendship and this means that John expects Charlie to act out goodwill. Thus, if John discovers that Charlie is not motivated by goodwill, the trust relation breaks down and John is betrayed. However, in *Christian*, the beggar trusts the Christian as a Christian, which means that he expects the Christian to be motivated by a sense of Christian duty. Thus, for an analysis of trust to give a full explanation of the normative expectations of trustors and

an explanation of what it means for trust to result in betrayal, one must look at the trusting-as relation implicit in the trusting-to relation.

Having focused in this section on trust rather than trustworthiness, I will now bring things full circle and explain what trusting-as relations mean for an analysis of trustworthiness. The meta-relationships in which trust occurs play a partial role in determining what it means for an agent to be trustworthy in trust relations. It does so in the same way that the object of trust in a three-place trust relation partially determines what it means to be trustworthy. In *False Friend*, the object of trust is secret keeping. Thus, in order to be trustworthy, Charlie needs to be capable of keeping a secret. If he was a compulsive blabbermouth, he would fail to satisfy John's trust even if he otherwise meant well. The object of trust typically sets the competence conditions for being trustworthy. Insofar as we trust people to do certain things, a part of what we expect of our trustees is that they will be capable of doing those things (Origi 2004, 64). However, this is not the complete picture. For when we consider trusting-as relations, the trustee needs to do more than be merely competent in order to be trustworthy. In *False Friend*, John is trusting Charlie as a friend. A condition of acting as a friend is that one acts out of a sense of goodwill. Thus, to be trustworthy as a friend, Charlie must not only keep the secret, but he must do so for the right reasons, reasons determined by the meta-relationship of friendship. Thus, to be trustworthy in *False Friend*, Charlie must be (1) capable of keeping a secret and (2) motivated to keep that secret out of goodwill. There is a third condition for trustworthiness, which is fixed by the notion of what it means to trust. For example, if we hold a commitment view of trust, then a part of what it means to be trustworthy is that one will make good on one's commitments. Now that we know what trusting-as relations are and how they interact with trust and trustworthiness, I can show how they can be invoked to explain the motivation intuition without making motivations constitutive features of trust or trustworthiness.

4. Explaining the motivation intuition

The motivation intuition is an intuition about what it means to be trustworthy. It is the intuition that being trustworthy requires something more than being merely reliable and that this missing something has something to do with the motivations of trustworthy persons. Thus, philosophers make specific motivations necessary conditions for

trustworthiness to resolve the problem. In this section, I argue that we can give an alternative and plausible account which does not require us to make motivations constitutive features of trustworthiness. I will do this by returning to the three cases discussed in Section 2 and explaining the cases in the terms of my trusting-as account of trust.

Let's begin with *False Friend*. In the previous section, I argued that the conditions for trustworthiness are fixed by (1) the object of trust, (2) the implicit trusting-as relation and (3) whatever is implied by the theory of trust. For our purposes, I will say that our theory of trust is the commitment account since it is an example of the non-motivational approach to trust that I am defending. In *False Friend*, Charlie satisfies condition (1) and possibly (3). He satisfies (1) because he keeps the secret. He appears to satisfy (3) because he doesn't break his commitment. However, he fails to satisfy (2) because he lacks the correct motivation. He is being trusted as a friend and therefore he needs to keep his commitment in virtue of goodwill and not self-interest. Incorporating the trusting-as relation into the commitment account, we might say that Charlie consequently does not keep his commitment either. If what it means to trust Charlie as a friend for John is that Charlie keeps his commitment out of goodwill, then Charlie does not actually keep his commitment at all. That this is relevant to the trust relation can be shown by referring to the way that John reacts when he discovers that Charlie is not acting as a friend should, that he is motivated to keep the secret for purely selfish ends. Clearly, John's expectations in trusting Charlie to keep his secret involved something greater than expecting him to be reliable. He also expected Charlie to keep his secret in virtue of their friendship. Supposing that this amounts to trusting John to act out of goodwill. Since Charlie does not act out of goodwill, he is not trustworthy.

One might argue that the commitment account cannot incorporate trusting-as relations, that it is somehow ad hoc to sneak in the relation to the account. Suddenly motivations have become relevant to an account of trust that is non-motivational. I have two responses to this line of objection. Firstly, the commitment account has not become motivational. The defining feature of a motivational account is that they make *specific* motivations necessary features of trust and trustworthiness. The defining feature of a non-motivational view is that they do not do this. The commitment account does not require specific motivations; rather, the motivations come from the trust-as relations, and in different trusting-as relations different motivations will be needed. Perhaps in some trusting-as relations, no specific motivation will be required at all for an

agent's trustworthiness. In any case, the point is that whether motivations are required for trustworthiness depends on the trusting-as relation, not on the theory of trust itself.

My second response to the worry that this feels ad hoc is to point out two things. Firstly, trusting-as relations do in fact seem to exist, and I have already presented arguments that they do have an impact on our expectations as trustors and the conditions for trustworthiness in Section 3. Secondly, incorporating these into the commitment account is no more ad hoc than incorporating specific competencies from the objects of trust is. If John relies on Charlie to keep a secret, then a part of what it means to be trustworthy is that Charlie has the capacity to keep secrets. If John relies on Charlie to keep his secret as a friend, then a part of what it means to be trustworthy is that Charlie acts out of goodwill. If we accept that trusting-as relations exist, then the role that they play in a given instance of trust is not significantly different from the role played by the object of trust. Just as the latter makes specific competencies necessary for trustworthiness, so does the former make specific motivations necessary for trustworthiness. There seems to me no reason in principle why, if the commitment account (or a non-motivational account generally) can incorporate the latter why it cannot also incorporate the former. A non-motivational account can explain the motivation intuition in *False Friend*. In this case, John was trusting Charlie as a friend. Thus, in this case, a condition for trustworthiness was that Charlie act out of goodwill. He did not do this, so he fails to be trustworthy.

Now let's look at the *Reliable Minion* and *Fallible Friend* cases and see if trusting-as relations help us explain why the fallible friend is more trustworthy than the reliable minion, even though the former is less reliable than the latter. If trust is reduced to mere reliability, the worry in cases such as these two is that a reliable but ill-motivated agent may turn out to be more trustworthy than a well-motivated but fallible agent. If we introduce motivations into our accounts of trustworthiness, we can fix this problem. Since Renfield doesn't bear Dracula goodwill, he is not trustworthy. Since Van Helsing does bear Mina and Jonathan Harker goodwill, he is trustworthy, though his trustworthiness is not maximally durable. To cast some initial doubt on this picture before I explain my own solution, I would suggest that, just as it seems extreme to say that Van Helsing is less trustworthy than Renfield, it equally seems extreme to say that Renfield is not trustworthy at all. Renfield is clearly devoted and loyal to Dracula, has promised him eternal loyalty, and while he

might well be motivated by fear and desire for immortality, it seems that if Renfield did fail Dracula, as he does in Stoker's novel, then this would count as a betrayal. Dracula certainly sees it that way, crushing Renfield's head against the floor and murdering him in an act of vengeance. After all, Renfield has obligated himself to Dracula's cause and has led him to believe that he will aid him in that cause. Thus, a failure is more than a failure to be reliable, it is going back on something that you led your trustor to believe that you would do. One might respond here with *so what?* Dracula is evil and so it is a good thing that Renfield broke his trust. One might even argue more strongly that Renfield cannot have an obligation to do something morally wrong in serving Dracula and that if this is so, it's not clear how he could betray Dracula since he did not fail to do anything that he should have done. I agree with the moral argument being made here but disagree that this means that Renfield cannot betray Dracula. Codes of honour, loyalty and trust relations exist among thieves as much as they exist among the morally praiseworthy (Lenard 2005: 366; Gambetta 1988). What makes a betrayal occur is that you actively encourage someone to depend on you, and then you let them down. In cases involving immorality, like *Reliable Minion*, it may well be that betraying your evil master is the morally right thing to do. But it's being the morally right thing to do doesn't mean that no betrayal has taken place. We could strengthen the point by considering a case where the trustor and trustee are both moral agents, but the trustee has to break the trust due to some other purpose. Suppose Van Helsing is in a trolley problem-like situation where he has to choose between saving Jonathan or Mina Harker. He cannot save both of them, though he has vowed to protect both of them. Let's suppose that he chooses to save Mina over Jonathan, and let's suppose that this is the morally correct action. Does this mean that he hasn't betrayed Jonathan by allowing him to die? It might be that we shouldn't blame Van Helsing because of the situation, but it still seems that Jonathan (in his spirit form at least) could justifiably complain of betrayal here.

Ultimately, it seems that trust and trustworthiness operate independently of morality. In such cases of immorality, the immoral trustee may have good moral reason to betray trust, but the trustor can still justifiably complain of a betrayal if the trustee reneges, even if that betrayal was morally required of the trustee. Motivational accounts such as Baier's, which moralise trust by making it about goodwill seem to struggle to explain this amorality. This doesn't rule out motivational accounts altogether, however, since one need not have a moralised motivational

view. Nevertheless, it helps our purposes to cast some doubt on whether Baier's view is as successful at explaining the distinction between these cases as we initially supposed. In the last section, I will consider Hardin's non-moralised motivational account of trust and show that even these kinds of account struggle to explain the difference between the cases, albeit in the other direction.

Now that I have cast doubt on Baier's motivational explanation of these cases, let's look at my non-motivational explanation. On my account, I am forced to accept the view that there is a sense in which Renfield is more trustworthy than Van Helsing. By stipulation, Renfield is more durably trustworthy than Van Helsing is of his friends. This is not something that can be changed even factoring in trusting-as relations. However, the recognition of trusting-as relations does allow us to make a significant distinction between Renfield and Van Helsing, which allows us to explain both the sense that (1) Renfield is trustworthy to a degree *and* (2) that Van Helsing's trustworthiness is more significant than Renfield's.

The meta-relationship between Van Helsing, Jonathan and Mina, is a relationship of three friends or comrades. This is the context in which Jonathan and Mina trust Van Helsing to protect them. They trust him as a fellow friend and comrade. The meta-relationship between Renfield and Dracula is that of minion or slave, to boss or master. When Dracula trusts Renfield, he trusts his slave to obey his commands. Within the confines of these meta-relationships, Renfield turns out to be more trustworthy than Van Helsing since he is more likely to satisfy the trust of his master than Van Helsing. His trustworthiness, as we already argued, is more durable than Van Helsing's. However, now that we recognise the existence of trusting-as relations, it turns out that there is also an important sense in which Van Helsing is more trustworthy than Renfield. Renfield cannot be trusted by Dracula in the same way that Van Helsing can be trusted by Jonathan and Mina. Because Renfield only acts out of fear and desire for immortality, and not out of goodwill, Dracula can trust Renfield as neither friend nor comrade. On the other hand, because Van Helsing does bear goodwill towards Mina and Jonathan, he can be trusted as a friend and comrade, even if we acknowledge that this trust is fallible in extreme circumstances.

One might object with the argument that this difference doesn't matter. At the end of the day, Renfield is still more durably trustworthy than Van Helsing. Moreover, just as Renfield can't be a trustworthy friend, neither can Van Helsing be trustworthy in the way that Renfield is trustworthy. This objection misses the point, however. The significant

difference between Van Helsing and Renfield is not a difference of quantity, but of quality. I accept that in terms of getting things done, Renfield is more trustworthy than Van Helsing, but at the same time, the trust relation holding between Van Helsing and his comrades, the way in which Van Helsing is trustworthy, is of a richer, more substantive kind than the way in which Renfield is trustworthy. A trustworthy friend is a person who is trustworthy because they genuinely care about you. They see you as someone who is of intrinsic worth, not simply a means to an end. They don't ask for payment for being trustworthy, they don't harbour dark motivations underneath. The fact that you are counting on them is all they need to be motivated to make good on that trust. Dracula might trust Renfield, and Renfield may be trustworthy, but Dracula knows that ultimately Renfield's devotion is grounded in other desires and fears. Renfield does not care about Dracula's interests because they are his interests; he cares about keeping his skin and becoming immortal. For Dracula this might be all he needs, but for someone who isn't a demonic monster, we might consider this to be an inferior kind of trust relation. It's of lower quality than the relationship of trust holding between Van Helsing and his comrades. If we think this, then the trusting-as relation allows us to make distinctions between different qualities of trust. We can say, for example, that Van Helsing's trust relation with his comrades is of a richer quality than the one that exists between Renfield and Dracula. Moreover, we can say that being a trustworthy friend, even if it is fallible, is greater than being a trustworthy slave or minion, in the sense that the former is qualitatively superior. This is because trust between friends is a relation of trust predicated on mutual goodwill and fellow feeling, while trust relations predicated on extrinsic desires or fears are relationships in which the trustee does not really care about the trustor or their interests.

There are two additional arguments I can give to support the notion that being a trustworthy friend is of a richer quality than being a trustworthy minion. Firstly, consider *False Friend*. When John discovers that he cannot trust Charlie as his friend, that Charlie is not a trustworthy friend, he is hurt and betrayed. Nevertheless, he could still trust Charlie to keep his secret since Charlie is motivated to do that. If this trust were of a richer quality, however, we would expect John to be elated at the revelation. But he is not. In losing the ability to trust Charlie as a friend, he has lost access to a trust relation that is of a richer and finer kind.

A second argument in favour of the quality of trust is that the richer forms of trust, such as friendship trust, in practice, will generally

provide the basis for more durable trust relations than other relations not predicated on goodwill. If an agent is motivated out of goodwill, then they are motivated to make good on your dependence for no other reason other than that you are depending on them. This makes for a potentially highly durable form of trust, since there are fewer extrinsic conditions on the motivations of the trustee. On the other hand, if a trustee is motivated out of some other motive, as Renfield is motivated out of fear and desire for immortality, then the trustor has to do work to ensure that these fears and desires still provide the motivational impetus for trustworthy behaviour. Dracula must make sure that Renfield fears him, and he must always keep that promise of immortality lingering. If he softens his approach, or if he refuses to grant Renfield immortality, then Renfield will not be trustworthy. Thus, one argument for the richer quality of forms of trustworthiness involving goodwill such as friendship trust is that they in general make for more durable trust relationships. At the very least, it is more desirable as a trustor to know that your trustee sees your dependence on them as a reason for action, rather than seeing them as someone who needs further extrinsic reasons for making good on trust.

In conclusion, recognising the existence of trusting-as relations allows us to explain the motivation intuition without making motivations constitutive features of trustworthiness. On a non-motivational account such as the commitment account, one condition for trustworthiness is fixed by the concept of trust. On the commitment view, one expects one's trustee to recognise that they ought to make good on their commitment and to keep the commitment. Since people make commitments to do certain things, one also expects that one's trustee will have the competence to satisfy whatever it is they have committed to do. And, since people make commitments in the context of meta-relationships, trustors also expect their trustees to keep their commitments for the right reasons. If I am trusting a friend to keep their commitments, then I trust them to keep their commitments out of goodwill. If I am trusting my minions, I expect them to keep their commitments out of fear and desire for my praise. In *False Friend*, John trusted Charlie as his friend, and this meant that to be trustworthy, Charlie had to act out of goodwill. Since Charlie does not do this, he betrays John's trust. In *Reliable Minion* and *Fallible Friend*, Renfield is more trustworthy as a minion than Van Helsing is trustworthy as a friend. However, being a trustworthy friend is something that Renfield can never be. Moreover, being trustworthy as a friend is a finer thing than being trustworthy as a selfish, cowardly minion. Thus, in

relegating the role of motivation to the trusting-as relation rather than to trust and trustworthiness, non-motivational views, so long as they can incorporate the trusting-as relation, can give a plausible explanation of the motivation intuition without making specific motivations constitutive features of trust or trustworthiness. Thus, the objection that non-motivational accounts cannot explain the motivation intuition does not hold.

5. The consequences of trusting-as relations

In this section, I want to consider the implications of trusting-as relations for motivational accounts of trustworthiness. I will argue that if we accept the existence of trusting-as relations, then motivational accounts of any stripe will be unable to provide a satisfactory account of trust in general. I will argue that instead, all motivational accounts of trust should be regarded as accounts of specific trusting-as relations, rather than trust relations generally.

In order to reject the motivational approach, I will discuss two specific accounts which offer the broadest range of interpretation. I will focus again on Baier's goodwill account, but I will also introduce Hardin's motivational account of trust, the encapsulated interest theory of trust. I do this because Hardin's account, unlike Baier's, is a broad account of trust that on the face of it seems like it might get around the problems faced by Baier's due to its breadth. However, I will show that ultimately, if we take trusting-as relations seriously, then motivational accounts prove to be unsatisfactory as theories of general trust and trustworthiness.

As we have seen, different trusting-as relations make a difference to the normative expectations of trustors, as well as to the conditions for trustworthiness. When we trust as friends, we expect our trustees to be motivated by goodwill, and thus acting out of goodwill is a condition for trustworthiness. When we trust someone to do their Christian duty, we expect them to be motivated by that duty, and thus acting from duty is a condition for trustworthiness. If both of these are instances of trust, then the problem with motivational accounts should already be apparent.

According to motivational accounts, a specific motivation is required in order to trust and in order to be trustworthy. On Baier's account, that motivation is goodwill. Thus, a goodwill account would be unable to explain why we are able to trust people to act out of a sense of duty, since acting out of duty does not necessarily mean acting out of goodwill,

and thus on the account, it does not count as trust at all, nor does it count as trustworthiness for a trustee to act out of a sense of duty over a sense of goodwill. It is a flaw of the goodwill account specifically that it is too narrow an account of trust to explain a number of situations in which we have heightened expectations of others, and rightfully feel betrayed when we are let down. Arguably I can place my trust in a politician, and I do not need to think she bears me any goodwill in order to trust her decisions. Let's say I trust a politician to legalise gay marriage, since this politician made such a promise in their manifesto. It seems that she doesn't bear me any goodwill. She doesn't know who I am. It seems that if I knew that she was only making this promise to improve her reputation, that if she reneged after getting elected that I could complain as a betrayal. Baier would claim that my complaint is invalid because I did not rely on the goodwill of the politician. But it is hard to see why goodwill matters. Surely what matters more is that the politician communicated to me and the voting population that we could count on her to do this thing if she got elected. And even if her motivations for making such a promise were self-interested, it seems that in failing to live up to this after encouraging people to count on her, she has betrayed them. For Baier, however, this would simply count as a case of being let down, of being disappointed. But to say this is to suggest that reliance on the politician to keep her promise is no different than relying on a car to get me to work; in both cases we simply expect our relied upon object to be reliable, and when we are let down we can only feel justifiably disappointed, but not betrayed. This to me seems to be an understatement.

One might of course suggest that the goodwill account is only useful for explaining trust in interpersonal relationships. However, it isn't clear that goodwill is even needed to explain trust in all our interpersonal dealings with one another. Jones gives the counterexample of stranger trust (Jones 2004, 4; Blackburn 1998). Do I need to presume that a stranger bears me goodwill in order to trust them for directions? Has a stranger who sends me to a dodgy part of town where I get mugged not betrayed my trust, regardless of their motivations or my presumptions about their motivations? The problem, then, with making specific motivations constitutive features of trust is that they make it difficult to offer a general account of trust that captures trust in different cases. The advantage of trusting-as relations is that the motivations that are required to satisfy a given trust relation are no longer determined by trust, but by the trusting-as relation, just as the kinds of competencies that are required to satisfy a given trust relation are determined by the object of trust.

However, perhaps the issue is not with motivational accounts per se, but specifically Baier's account, which provides a particularly narrow motivation as a condition on trust and trustworthiness. To show that the problem applies more generally, I will look at Hardin's encapsulated interest theory of trust. On this account, you trust someone when you judge that they will be reliable because they will encapsulate your interests among their own (2002, chp 2; 2006). This account explains why *Christian* is still an instance of trust. Even if the Christian does not act out of goodwill, she can still be trusted because the beggar believes that the Christian will encapsulate his interests insofar as they are required for the Christian to fulfil her duty. Likewise, the Christian can be trustworthy insofar as she encapsulates the beggar's interests, regardless of a lack of goodwill. Hardin's account also explains the gay marriage case. What I expect of the politician who promises to legalise gay marriage is not that she bears me personal goodwill, but that in this matter she will encapsulate my interest in legalising gay marriage among her own. Whether she does this out of self-interestedness or a sense of justice is irrelevant on the account.

The problem with Hardin's account is that it seems to explain the cases where Baier's account fails quite well, but it cannot offer a satisfactory explanation of the cases that Baier's account does explain well. Take *False Friend*. In *False Friend*, Charlie encapsulates John's interests among his own; thus John can trust Charlie and Charlie can be trustworthy. It doesn't matter that Charlie's reasons for encapsulating John's interests are selfish. The beauty of Baier's account is that it can explain why Charlie is not trustworthy. He is not trustworthy because he needs to have goodwill in order to be trustworthy. Hardin's account then, fails to provide a satisfactory account of trust in more personal trust relations, such as trust relations that occur between friends.

In conclusion, the disadvantage of motivational accounts of trust and trustworthiness is that they undergenerate. They can account for some range of trust relations but do not account for all trust relations. Recognising the meta-relationships in which trust occurs is important, because these meta-relationships impact what it means to trust and be trustworthy, and it is these meta-relationships, not trust itself, which determines whether we need to expect our trustees to have specific motivations to be trustworthy. If this is correct, then motivational accounts cannot provide satisfactory accounts of trust, because in making specific motivations constitutive features of trust, they take the motivational role that trusting-as relations play, which results in

undergeneration. The advantage of non-motivational accounts that incorporate trusting-as relations is that they do not do this, and thus, can provide stronger analyses of trust and trustworthiness.

6. Conclusion

In this paper, I defended the non-motivational approach to analysing trustworthiness and demonstrated the inadequacies of motivational approaches to analysing trustworthiness. I did this by highlighting an overlooked feature of trust relations, what I called ‘trusting-as’ relations. I defined a trusting-as relation as a meta-relationship of trust that determines some of the conditions for trustworthiness in a three-place trust relation. I argued that one condition that the trusting-as relation may set for trustworthiness is that the trustee act in virtue of a specific motivation, although this motivation is specific to the trusting-as relation and not determined by trust itself.

In showing that trusting-as relations set the motivational conditions for trustworthiness, I was able to then defend the non-motivational approach to analysing trustworthiness. A key objection to non-motivational accounts such as Hawley’s commitment account is that they cannot explain the motivation intuition; the intuition that being trustworthy requires more than being simply reliable. I motivated the intuition by showing that cases in which an agent is reliable but ill-motivated seem to be cases where we are reluctant to ascribe trustworthiness to the reliable agent. Motivational accounts explain these cases by making specific motivations constitutive features of trust and trustworthiness. On my non-motivational explanation, one need not look to the theory of trust or trustworthiness to determine whether a specific motivation is needed for trust or trustworthiness, but instead one ought to look at the implicit trusting-as relation instead. In doing so, I argued, one can explain the motivation intuition without making motivations constitutive features of trust or trustworthiness.

In introducing trusting-as relations, I was able to show a key inadequacy of motivational accounts of trust. The inadequacy is that such accounts cannot accommodate trusting-as relations. This is because it is the trusting-as relation that determines whether and which motivations are required for trustworthiness in trust relations, rather than trust or trustworthiness requiring specific motivations in all cases. Therefore, while motivational accounts may be helpful in understanding very specific kinds of trust relations, they will not be helpful in providing

general accounts of trust or trustworthiness. Whether or not non-motivational accounts such as Hawley's commitment account will be able to provide such a general account is a project for further research.

Acknowledgements

I am very grateful to Matthew Green for his insightful contributions to our discussions and his extensive comments on the paper. Special thanks are also owed to Katherine Hawley, Rowan Cruft, Niklas Becker and two anonymous reviewers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by generous funding from the Society for Applied Philosophy's Doctoral Bursary awards.

References

- Annas, Julia. 1977. "Plato and Aristotle on Friendship and Altruism." *Mind* 86: 532–554.
- Annis, David. 1987. "The Meaning, Value, and Duties of Friendship." *American Philosophical Quarterly* 24: 349–356.
- Baier, Annette. 1986. "Trust and Antitrust." *Ethics* 96 (2): 231–260.
- Baier, Annette. 1991. "Two Lectures on 'Trust': Lecture 1, 'Trust and Its Vulnerabilities' and Lecture 2, 'Sustaining Trust'." In *Tanner Lectures on Human Values*. Vol. 13, edited by G. B. Peterson, 109–174. Salt Lake City: University of Utah Press.
- Baier, Annette. 1995. "Trust and Anti-Trust." In *Moral Prejudices: Essays on Ethics*, 95–129. Cambridge: Harvard University Press.
- Blackburn, Simon. 1998. *Ruling Passion: A Theory of Practical Reasoning*. Oxford: Clarendon Press.
- Cocking, Dean, and Janet Kennett. 1998. "Friendship and the Self." *Ethics* 108: 502–527.
- Domenicucci, Jacopo, and Richard Holton. 2017. "Trust as a Two-Place Relation." In *The Philosophy of Trust*, edited by Paul Faulkner, and Thomas Simpson, 149–160. Oxford: Oxford University Press.
- Faulkner, Paul. 2007. "On Telling and Trusting." *Mind* 116 (464): 875–902.
- Faulkner, Paul. 2015. "The Attitude of Trust is Basic." *Analysis* 75 (3): 424–429.
- Gambetta, Diego, ed. 1988. *Trust: Making and Breaking Cooperative Relations*. Blackwell.
- Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.
- Hardin, Russell. 2003. "Gaming Trust." In *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research*, edited by E. Ostrom, and J. Walker, 80–103. New York: Russell Sage Foundation.

- Hardin, Russell. 2006. *Trust: Key Concepts*. Cambridge: Polity Press.
- Hawley, Katherine. 2012. *Trust: A Very Short Introduction*. Oxford: Oxford University Press.
- Hawley, Katherine. 2014. "Trust, Distrust, and Commitment." *Nous (Detroit, Mich)* 48 (1): 1–20.
- Hawley, Katherine. 2019. *How to Be Trustworthy*. Oxford: Oxford University Press.
- Helm, Bennett. 2021. "Friendship." In *The Stanford Encyclopedia of Philosophy*. Fall 2021 ed. edited by Edward N. Zalta. forthcoming, <https://plato.stanford.edu/archives/fall2021/entries/friendship/>.
- Hieronymi, Pamela. 2008. "The Reasons of Trust." *Australasian Journal of Philosophy* 86 (2): 213–236.
- Holton, Richard. 1994. "Deciding to Trust, Coming to Believe." *Australasian Journal of Philosophy* 72 (1): 63–76.
- Horsburgh, H. J. N. 1960. "The Ethics of Trust." *Philosophical Quarterly* 10 (41): 343–354.
- Jones, Karen. 1996. "Trust as an Affective Attitude." *Ethics* 107 (1): 4–25.
- Jones, Karen. 2004. "Trust and Terror." In *Moral Psychology: Feminist Ethics and Social Theory*, edited by Peggy DesAutels, and Margaret Urban Walker, 3–18. Maryland: Rowman & Littlefield.
- Lenard, Patti Tamara. 2005. "The Decline of Trust, The Decline of Democracy?." *Critical Review of International Social and Political Philosophy* 8 (3): 363–378.
- Margalit, Avishai. 2017. *On Betrayal*. Cambridge: Harvard University Press.
- McCraw, Benjamin. 2015. "The Nature of Epistemic Trust." *Social Epistemology* 29 (4): 413–430.
- McGeer, Victoria. 2008. "Trust, Hope and Empowerment." *Australasian Journal of Philosophy* 86 (2): 237–254.
- McGeer, Victoria, and Pettit Philip. 2017. "The Empowering Theory of Trust." In *The Philosophy of Trust*, edited by Paul Faulkner, and Thomas Simpson, 14–34. Oxford: Oxford University Press.
- McLeod, Carolyn. 2015. "Trust." In *The Stanford Encyclopedia of Philosophy*. Fall 2015 ed. edited by Edward N. Zalta. Accessed 9 November 2018. <https://plato.stanford.edu/archives/fall2015/entries/trust/>.
- Origi, Gloria. 2004. "Is Trust an Epistemological Notion?" *Episteme* 1 (1): 61–72. doi:10.3366/epi.2004.1.1.61
- Ostrom, Elinor, and James Walker. 2003. *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research*. New York: Russell Sage Foundation.
- Telfer, Elizabeth. 1970–1971. "Friendship." *Proceedings of the Aristotelian Society* 71: 223–241.
- White, Richard. 2001. *Love's Philosophy*. Maryland: Rowman & Littlefield.
- Zagzebski, Linda. 1996. *Virtues of the Mind*. New York: Cambridge University Press.
- Zagzebski, Linda. 2012. *Epistemic Authority*. Oxford: Oxford University Press.