

*Trust, Distrust and Commitment*¹

KATHERINE HAWLEY
University of St Andrews

Trust is intriguing. Trust is valued—the alleged crisis of trust is a famine not a flood—yet misplaced trust can be dangerous. Trust is needed when we lack knowledge, yet we have most trust in those we know best. Trust bestowed can be both an honour and a burden; distrust is rarely welcome. Practical reasons to trust can outstrip the evidence, yet counter-evidence can make it impossible to trust. Recent philosophical work on trust has emphasised its importance to both epistemology and ethics, asking whether reasons to trust someone must be based on reasons to think her trustworthy (e.g. Hinchman 2005, Faulkner 2007, McGeer 2008, Hieronymi 2008).

But this work, and the literature it builds upon, is curiously lopsided. To understand trust, we must also understand distrust, yet distrust is usually treated as a mere afterthought, or mistakenly equated with the absence of trust. In this paper I offer an account of both trust and distrust. Roughly speaking, to trust someone is to rely upon that person to fulfil a commitment, whilst distrust involves an expectation of unfulfilled commitment.

Whilst I advocate this commitment theory, my main concern is to demonstrate the importance of providing a unified account of both trust and distrust, either through commitment or otherwise. We need to understand both trust and distrust if we are to understand the different ways in which trust can go wrong, the reasons why both trust and distrust are sometimes unwanted, the nature and limitations of trustworthiness, and the difference between unpredictability and untrustworthiness.

I will adopt two common assumptions. First, trust is primarily a three-place relation, involving two people and a task: you may trust me to look after your children, to keep a secret, or to tell the truth (Holton 1994, Jones 1996, Hardin 2002, Hieronymi 2008). We do sometimes speak of simply trusting someone, and I will discuss this generalised trust in section 7.2. Second, trust involves expectations about both competence and willingness: when you trust me to look after your children, you take it that I am capable of childcare, and that I will exercise that capability as required.

1 Trust Is Not Mere Reliance

Following Annette Baier, philosophers usually distinguish genuine trust from mere reliance (Baier 1986, 234; Hieronymi 2008, 215; Holton 1994, 2–3; Jones 1996, 14;

Jones 2004, 4; McLeod 2006, 3; O'Neill 2002, 15; Potter 2002, 3–4; Pettit 1995, 205). For example, we often rely upon inanimate objects but we do not grant them the rich trust we sometimes grant one another; inanimate objects can be reliable but not genuinely trustworthy. Moreover our reactions to misplaced trust differ from our reactions to misplaced reliance. Suppose I trust you to look after a precious glass vase, yet you carelessly break it. I may feel betrayed and angry; recriminations will be in order; I may demand an apology. Suppose instead that I rely on a shelf to support the vase, yet the shelf collapses, breaking the vase. I will be disappointed, perhaps upset, but it would be inappropriate to feel betrayed by the shelf, or to demand an apology from it.

Inanimate objects can be relied upon without being trusted. And there are circumstances in which people are relied upon without being trusted. Suppose you regularly bring too much lunch to work, and leave the excess for others to eat. Suppose you do this because you're bad at judging quantities, not because you're keen to feed your colleagues. I rely on you to provide my lunch: I anticipate that you will do so, and I don't make alternative arrangements. But this reliance should not amount to trust: you would owe me no apology if you ate all the food yourself, and I ought not to feel betrayed by this, even if I felt disappointed (and hungry). Likewise 'Kant's neighbors who counted on his regular habits as a clock . . . might be disappointed with him if he slept in one day, but not let down by him, let alone had their trust betrayed' (Baier 1986, 235). Our attitudes to people acting reliably under threat are sometimes also cited in this context.

So there is a distinction between trust in a rich sense—trust which can be betrayed—and mere reliance. But distinguishing trust from mere reliance clashes with ordinary usage: we might talk of trusting a shelf to hold a vase. Anticipating these concerns, some distinguish two types of trust—normative and predictive (Hollis 1998, 10), or affective and predictive (Faulkner 2007, 880)—rather than distinguishing trust from mere reliance. I will stick with the less theory-laden 'trust' and 'mere reliance', whilst acknowledging that we can use the language of trust for mere reliance.

Why focus on a distinction which is not consistently marked in ordinary language?

The distinction is important because trust, not mere reliance, is a significant category for normative assessment. Trust, unlike mere reliance, is connected to betrayal. Moreover trustworthiness is clearly distinguished from mere reliability. Trustworthiness is admirable, something to be aspired to and inculcated in our children: it is a virtue in the everyday sense, and perhaps in the richer sense of virtue ethics too. Mere reliability, however, is not. A reliable person is simply predictable: someone who can be relied upon to lose keys, or succumb to shallow rhetoric, is predictable in these respects, but isn't therefore admirable. Even reliability in more welcome respects need not amount to trustworthiness: when you reliably bring too much lunch, you do not demonstrate trustworthiness, and nor would you demonstrate untrustworthiness if you stopped.

2 Distrust Is Not Mere Nonreliance

Philosophical accounts of trust standardly attempt to explicate this distinction between trust and mere reliance. But it is a mistake to theorise trust without considering distrust. Distrust is not mere absence of trust: I rely upon the shelf to support the vase. For the reasons given above, I do not trust the shelf. But nor do I distrust it.

Distrust is not mere absence of trust. Moreover, distrust is not even mere absence of reliance. Like trust, distrust has a normative dimension. The distinction between trust and mere reliance shows in our different reactions to misplaced trust (betrayal) and misplaced reliance (disappointment). Similarly, the distinction between distrust and mere nonreliance shows in our different reactions to misplaced distrust and misplaced nonreliance. If I discover that I have wrongly distrusted you, appropriate reactions include remorse, apology, and requests for forgiveness. In contrast, if I take my car to be unreliable, then discover that it is after all reliable, then remorse would not be appropriate. I might regret some missed opportunities, but that's all.

Not relying on an inanimate object does not amount to distrusting it. Moreover, not relying upon people need not amount to distrusting them. My colleagues have never bought me champagne, so in particular I do not rely upon them to buy me champagne next Friday. But it would be wrong, even offensive, to say that I distrust my colleagues in this respect—after all, they have not offered to buy me champagne next Friday, and there is no social convention that they do so. If they did buy me champagne unexpectedly, I ought to be grateful, but I need not feel remorse about my earlier decision not to rely on them. Indeed, it would be bad manners for me to suggest in retrospect that I should have trusted my colleagues to buy me champagne, or to apologise for my earlier non-trust.

Whilst I do not rely upon my colleagues to buy me champagne, it would nevertheless be wrong to distrust them in this respect. Not because they can after all be trusted to buy me champagne, but because neither trust nor distrust in this respect is appropriate. Recall the case in which I rely upon you to supply my lunch, without trusting you to do so, and could feel disappointed but not betrayed if you stopped. Likewise, I do not rely upon my colleagues to buy me champagne, but I do not distrust them in this respect; the fact that they do not buy me champagne does not indicate any degree of untrustworthiness.

So distrust is richer than mere nonreliance, just as trust is richer than reliance. Just as we should distinguish trustworthiness from mere reliability, we should distinguish untrustworthiness from mere unreliability: colleagues who do not buy me champagne are unreliable in this respect, but not thereby untrustworthy. Just as there is a middle ground between trust and distrust, there is a middle ground between trustworthiness and untrustworthiness—in the clearest case, inanimate objects merit neither trust nor distrust, so they are neither trustworthy nor untrustworthy. Thus 'untrustworthy'—i.e. 'meriting distrust'—is not simply the complement of 'trustworthy'. 'Distrustworthy' would be more literal, but I will resist the neologism.

3 Reliance and Nonreliance

Many philosophers have distinguished trust from mere reliance; I have distinguished distrust from mere nonreliance. But what is it to rely upon someone to do something? I will adopt Richard Holton's view, that to rely on someone to X is to act on the supposition that she will X: this can fall short of believing that she will X, though it is incompatible with outright belief that she will not X (Holton 1994). Relying is not always a matter of belief, its justification conditions are not purely epistemic, and, unlike believing, relying is sometimes a matter of direct choice.

In this sense, relying on someone to do something needn't mean putting your fate in their hands. You can rely upon me to bring enough food for everyone at the picnic whilst nevertheless bringing plenty of food yourself, because you don't want to seem ungenerous: you're acting on the supposition that I will bring lots of food, and indeed this partly explains the large quantities you bring along. Thus reliance in this sense needn't imply risk or vulnerability. For me, this is one attraction of Holton's account; others disagree, but I cannot examine this issue here.

Distrusting someone with respect to X involves not relying upon her to X, rather than relying upon her not to X: distrust does not require confident prediction of misbehaviour. Extending Holton's framework, nonreliance involves not acting on a particular positive supposition, rather than acting on the corresponding negative supposition.

There are good questions to ask about degrees of reliance, degrees of reliability, and the fact that suppositions are sometimes idle wheels: when I don't need your help, it may make no practical difference whether or not I act under the supposition that you would help if asked. Again, I do not have space to explore these issues here, but I am consoled by the fact that similar questions arise for many, perhaps all, different accounts of trust. Moreover, our understanding of these grey areas between reliance and nonreliance can only be enhanced by paying proper attention to distrust alongside trust, as I recommend.

4 Motives-Based Accounts of Trust Fail to Explain Distrust

Most philosophical accounts of trust are explicitly designed to explain the difference between trust and mere reliance. They should also enable us to explain the difference between distrust and mere nonreliance. There is a simple reason and a more complex reason why an account of trust should also furnish an account of distrust. Simple: the trust/reliance and distrust/nonreliance distinctions seem closely analogous, so our default expectation should be that the two distinctions are grounded in analogous ways. More complex: an account of trust should explain when trust is appropriate and when it is not. In a given situation, trust may be inappropriate either because distrust is appropriate or because neither trust nor distrust is appropriate. Understanding these possibilities will improve our understanding of how and when trust is appropriate.

But standard theories focus on the difference between trust and mere reliance, treating distrust as an afterthought.

Trust is standardly thought to involve reliance, plus some extra factor. Many theories identify this extra factor with a positive view of the motives of the trusted person. If I trust you to look after my vase, then I rely upon you to do so, and moreover I take it that you have the right kind of motive for looking after my vase. Different theorists disagree about what the ‘right kind’ of motive might be.

In this section, I will show how motives-based accounts of trust cannot readily explain the nature of distrust. I cannot survey every motives-based account of trust (McLeod (2006) provides a critical overview), but will focus on those offered by Russell Hardin (2002) and Karen Jones (1996). Both theories are well-developed, sophisticated, and prominent in the literature (though Jones herself later advocates a different account (2004)). Moreover they differ significantly from one another, turning on the trustee’s rational self-interest and other-directed goodwill respectively.

Hardin argues that when we trust someone, we expect the trustee to encapsulate our interests within her own because she has an interest in maintaining or strengthening her relationship with us. In trusting you to look after my vase, I take it that you will do so because you have incorporated my interest in preserving the vase amongst your own interests: looking after the vase is now in your own interest. In contrast, when I rely upon the shelf to hold the vase, I do not have any expectation about the shelf’s motives or interests, for I realise it has none.

Jones defines trust as

an attitude of optimism that the goodwill and competence of another will extend to cover the domain of our interaction with her, together with the expectation that the one trusted will be directly and favorably moved by the thought that we are counting on her. (1996, 1)

For Jones, it is important that optimism be an affective attitude, not (just) a belief, so that the justification conditions for optimism, and thus trust, are not purely epistemic. Reference to the trustee’s responsiveness to our counting on her permits a distinction between genuine trustworthiness and reliable benevolence.

What about distrust? On the motives-based model, we might expect distrust to involve nonreliance, plus a negative attitude regarding the motives of the distrustee. This negative attitude must go beyond expecting the distrustee to lack the motives required for trustworthiness. After all, inanimate objects lack the motives required for trustworthiness: they do not incorporate our interests amongst their own, and they do not act out of goodwill towards us. Yet we do not distrust inanimate objects, even when we decide not to rely upon them.

The same is often true in interpersonal situations. Consider the colleagues who do not buy me champagne. I do not rely upon them to buy me champagne; moreover I know that they have not incorporated my interest in drinking champagne amongst their own interests, and I am not optimistic about their goodwill in the champagne-buying domain. Yet it’s not appropriate for me to distrust my colleagues: they are not displaying untrustworthiness, and if they do surprise me with champagne, I needn’t feel remorse about not having trusted them. Neither trust nor distrust is appropriate in this context.

So if distrust is a matter of nonreliance plus a negative attitude towards the distrustee, this negative attitude must go beyond merely expecting the distrustee to lack certain positive motives. Perhaps distrust involves expecting the distrustee to act out of ill will towards us, or expecting the distrustee to have an interest in frustrating our interests. This would explain why we do not distrust inanimate objects (they are not actively working against us), and why I should not distrust my colleagues (they are not maliciously striving to deny me champagne).

But neither expectation of ill will nor expectation of attempts to frustrate my interests is necessary for distrust. After all, someone who lies and cheats to achieve her goals should be distrusted, even if she does not bother to bear either goodwill or ill will to others, and does not care about other people's interests.

Nor are gloomy expectations sufficient for distrust, even in combination with nonreliance. Suppose that a deeply honourable person campaigns to have me imprisoned for my crimes. I cannot rely on this person to help me, moreover I know that she bears me ill-will and is actively trying to frustrate my goals. But my attitude to her needn't amount to distrust, for she is straightforward and honest in her campaigning. (This doesn't mean that I should trust her; merely that I do not have grounds to distrust her.) Recall the connections between distrust, untrustworthiness, and remorse. My opponent does not display untrustworthiness in her open campaigning against me. And if she turns out to be more helpful than I had expected, I need not feel remorse about my previous attitude of nonreliance.

Now, I have not explored all the options here, nor represented the full complexity of Hardin's and Jones's positions. Nevertheless, I have shown that neither account of trust can handle distrust easily, and this for reasons which generalise to other motives-based accounts. Both Hardin and Jones take care to distinguish genuine trust from mere reliance. But each considers decisions about distrust only in situations where either genuine trust or genuine distrust is appropriate, where a person's behaviour demonstrates either trustworthiness or untrustworthiness (i.e. distrust-worthiness). This narrow focus means that trust, distrust, and indecision seem to exhaust the options, leading Hardin and Jones to think of distrust as a kind of decisive lack of trust (Hardin 2002, 90; Jones 1996, 17; see also McLeod 2002, 34).

Instead, we should ask about the preconditions for trust-or-distrust: what is it about the excess-lunch-bringer, the champagne-non-buyers, and indeed inanimate objects which mean that they are not suitable recipients of either trust or distrust in the relevant respects? The primary reason that trust is not appropriate in these cases is that neither trust or distrust is appropriate.

5 Other Accounts of Trust Are Incomplete

Motives-based accounts of trust do not provide the materials for explaining the difference between distrust and nonreliance. But motives-based accounts do not exhaust the field. In this section I will discuss Richard Holton's account of trust in his (1994), and Karen Jones's later (2004) account, arguing that whilst these accounts can be adapted to address distrust, they are unexplanatory in important ways.

Holton invokes the Strawsonian participant stance:

I think that the difference between trust and [mere] reliance is that trust involves something like a participant stance towards the person you are trusting...trusting someone is one way of treating them as a person. But if this is right, it shows how important it is that we do not treat the participant stance as an all or nothing affair. Even when you do trust a person, you need not trust them in every way.... You can trust a person to do some things without trusting them to do others. (1994, 4)

Taking the participant stance towards someone does seem to be a necessary condition for trust—indeed, like others in this literature, I earlier identified trust via its connections with reactive attitudes such as resentment and the sense of betrayal. Holton does not explicitly mention distrust in his article, but taking the participant stance seems also to be a necessary condition for distrust: we do not distrust faulty machines. So there is clearly something right about this approach.

However, like Hardin (2002) and Jones (1996), Holton is misled by his lack of explicit attention to distrust. He seems to suggest that where our trust in someone is limited, then so too is the extent to which we adopt the participant stance to that person. But where we distrust, rather than trust, someone in a particular respect, this marks no diminution in our tendency to hold reactive attitudes towards that person. Indeed, attitudes such as resentment are to the fore in situations of distrust.

Moreover, interpersonal respect for others can require us neither to trust nor to distrust them in a given respect. In some situations, either of trust and distrust would be an imposition. Trust involves anticipation of action, so it's clear why someone might prefer not to be trusted to do something she would prefer not to do. But that does not mean she wants to be distrusted in that respect.

Even trusting someone to do something she is happy to do can be an imposition. Suppose that my colleagues do after all plan to buy me champagne. Still, they do not invite or welcome my trust in this respect; instead, they want to give me a treat, not merely to act as trustworthiness requires, and certainly not to risk betraying me if they forget to buy the champagne. Such situations are neglected by most discussions of trust: after all, my colleagues are happy for me to know that they have adopted my interests amongst their own, and that they bear me goodwill, yet they don't want champagne-buying to become an issue of trust between us.

So it goes in even the most intimate, trusting relationships. Suppose that I cook dinner for my partner each evening and he comes to rely on this. Even if I enjoy cooking, I do not want my partner to make this a matter of trust. That is, I do not want to risk betraying him in even a minor way if I don't cook one evening, and nor do I want that to count against my trustworthiness. We aspire to a completely trusting relationship—we would like to avoid even the slightest distrust—but we do not aspire to turn all our interactions into issues of trust.

Holton correctly identifies the participant stance as a necessary element of trust, and adopting this stance is necessary for distrust too. But relying upon someone to whom you take a participant stance does not always mean trusting that person: some interactions lie outside the realm of trust and distrust. Likewise, deciding not

to rely upon someone to whom you take a participant stance—deciding to buy your own champagne—need not mean distrusting that person. And adopting the participant stance can sometimes require us not to turn every interaction into a matter of trust and distrust. But such nontrust interactions are still within the scope of the participant stance: it's appropriate for my partner to express his gratitude for my cooking, even though he should not convert his reliance into trust.

Holton says that he is not offering a reductive analysis of trust (1994, 5), so we should not demand of him a reductive account of distrust. And he has identified a necessary condition for both trust and distrust. Nevertheless it is clear that this condition needs significant supplementation if we are to understand when trust-or-distrust in a given respect is appropriate, and when mere reliance or nonreliance is the appropriate interpersonal attitude.

Karen Jones (2004), in a shift from her earlier view, also connects trust and the reactive attitudes but in a more modulated fashion:

[Three-place t]rust is accepted vulnerability to another person's power over something one cares about, where (1) the truster foregoes searching (at the time) for ways to reduce such vulnerability, and (2) the truster maintains normative expectations of the one-trusted that they not use that power to harm what is entrusted. (2004, 6)

I will take it that the notions of accepted vulnerability plus foregoing the attempt to reduce such vulnerability capture roughly the notion of reliance. One might have something like this attitude to an inanimate object like a car, for example. I do not underestimate the differences between Jones's notion, characterised in terms of power, care, vulnerability and harm, and the thinner characterisation of reliance I have adopted from Holton. But I do not have space to explore these differences here, and will focus instead on normative expectations, which must do the work of distinguishing trust from mere reliance.

Normative expectations, for Jones, are "multistranded dispositions, including dispositions to evaluative judgement and to reactive attitudes" (2004, 17, note 8): when you trust someone, you are liable to feel resentful if she lets you down through ill will or laziness, and whilst you might not feel resentful if she lets you down by accident, you may still think that an apology is warranted. For Jones, this multistrandedness is a key advantage of her account over Holton's, which she reads as overly-focused on liability to resentment.

If this is trust, what might distrust be? Let's understand nonreliance as a refusal to accept vulnerability, and/or a continuing attempt to reduce such vulnerability. One might have this attitude to a machine one takes to be unreliable. What more is needed for distrust? Not just the absence of normative expectations, since such expectations are absent from our attitudes towards inanimate objects.

Jones distinguishes normative expectations from predictive expectations: we can normatively expect something of someone without predicting that she will in fact do what we expect of her. Though Jones doesn't say so, it is then plausible that the normative expectations involved in distrust are exactly the normative expectations which would otherwise be involved in trusting that person in that respect. So

distrust is nonreliance plus a tendency to resentment, a tendency to judge the distrustee negatively, or tendency to think that an apology is warranted: distrust is something like disappointed trust, though perhaps not preceded by an episode of trust.

Because Jones pins normative expectations to specific tasks (or, rather, to specific cared-for things), she can accommodate the fact that respect for others, even in very intimate relationships, can require us to stick with reliance-or-nonreliance rather than trust-or-distrust in certain respects. I am happy for my partner to predict that I will cook dinner tonight, but I do not want him to develop normative expectations, to be poised to resent my not cooking.

Thus Jones's account of trust in terms of normative expectations can be extended to provide an account of distrust which is plausible so far as it goes. But it does not go far enough. Flawed though they are, the motives-based accounts of trust do provide an account of what it is that the truster attributes to the trustee (e.g. goodwill). This permits investigation of the correctness conditions for trust: what is it about someone which makes it appropriate to trust her? The Holton and later Jones accounts, in contrast, tell us more about the truster's attitudes than they do about the features of the trustee to which those attitudes are directed.

Such accounts enable us to characterise the attitude of trust, and, by extension, of distrust. But we need more. In particular, we need a story about when trust, distrust or neither is objectively appropriate—what is the worldly situation to which (dis)trust is an appropriate response? When is it appropriate to have (dis)trust-related normative expectations of someone? This is not just a question of practical self-interest or mental hygiene: we owe it to others to get this right, since wrongful distrust and even wrongful trust can harm others (Fricker 2007, Jones 2002, Marsh 2011). We also need to understand the virtue of trustworthiness and the vice of untrustworthiness, as they are distinguished from reliability and unreliability.

To do all this, we need a basis for our judgements about reliability: how, if at all, can we predict what others will do? (A closely related question is key to the epistemology of testimony.) But we also need a basis for our judgements about when it is appropriate to trust-or-distrust, not merely to rely-or-not-rely. Many of the relevant norms apply only when we enter the realm of appropriate trust-or-distrust.

I propose that it is appropriate to trust or distrust someone to do something only if that person has an explicit or implicit commitment to doing it. In the next section, I explicate this proposal, then in following sections I illustrate its explanatory power. However my overarching concern is to establish that the following is an essential question: what are the features of other people to which either trust or distrust is an appropriate response? I answer this question with reference to commitment, but other answers are available.

6 The Commitment Account

Recall the situation in which you reliably bring too much lunch to work, because you are a bad judge of quantities, and I get to eat the excess. My attitude to you

in this situation is one of reliance, but not trust, and your reliability in this respect shows nothing about whether you are trustworthy. This is not a matter of trust—or distrust—because you have made no commitment to provide me with lunch. But if we adapt the case so as to suggest commitment, it starts to look more like a matter of trust. Suppose we enjoy eating together regularly, you describe your plans for the next day, I say how much I'm looking forward to it, and so on. To the extent that this involves a commitment on your part, it seems reasonable for me to feel betrayed and expect apologies if one day you fail to bring lunch and I go hungry.

Recognised lack of commitment also explains our judgement about the colleagues who do not buy me champagne. They are unreliable in this respect, but it would be unreasonable of me therefore to distrust them, or to consider them untrustworthy in any respect. They have not offered to buy me champagne, and there's no social convention that they do so. They have incurred no commitment to buy me champagne and their failure to do so is not a failure of trustworthiness.

Here is my account of trust and distrust:

To trust someone to do something is to believe that she has a commitment to doing it, and to rely upon her to meet that commitment. To distrust someone to do something is to believe that she has a commitment to doing it, and yet not rely upon her to meet that commitment.

In later sections I will put this account of trust and distrust to work, illustrating its power and exploring its consequences. But the central notion of commitment needs immediate clarification; I will also discuss an apparent counterexample to the account.

6.1 Commitment

In one sense, having a commitment to do something involves having a determined intention to do it; this is the sense in which we can admire someone's commitment to a project. But this is not the relevant sense for the commitment account of trust and distrust. After all, this sort of psychological commitment is often exactly what's missing when distrust is appropriate.

In the relevant sense, one can be committed to doing something one has no intention of doing: if I've promised to come to your birthday party, but I now decide I can't be bothered, I still have a commitment in the relevant sense, even though I have no intention of fulfilling it. Promising is one clear way of acquiring the relevant sort of commitment, though explicit promising is not the only route to commitment.

My account of trust and distrust depends upon this notion, but I do not offer a fleshed-out theory of commitment. Can my account of trust and distrust be gerrymandered to fit just any old judgement about cases? No. Although there are borderline cases, there are also clear cases of commitment, and of noncommitment, and these are, respectively, cases in which either trust or distrust is appropriate, and cases in which neither trust or distrust is appropriate.

Aren't there obvious, immediate counterexamples to the claim that trust and distrust presuppose commitment of this kind? We often trust people to do things which we know they have not explicitly promised to do. I trust my friends not to steal the cutlery when they come to dinner, and, at least in some areas, I trust strangers to let me walk unhindered. To make the account plausible I must use a very broad notion of commitment: commitments can be implicit or explicit, weighty or trivial, conferred by roles and external circumstances, default or acquired, welcome or unwelcome. In particular I will take it that mutual expectation and convention give rise to commitment unless we take steps to disown these. And in section 8 I will argue that assertion involves commitment, so that trusting someone's word falls within the scope of the account.

Should I instead place obligation rather than commitment at the heart of trust and distrust? Commitments typically give us obligations; and perhaps I have stretched the notion of commitment far enough to ensure that obligations always give us commitments. Perhaps trusting someone to do something is a matter of thinking her obliged to do it, and relying upon her to fulfil her obligation, whilst distrusting involves taking someone to have an obligation, yet not relying upon her to fulfil it.

Hollis (1998, 11) and Nickel (2007) link trust to obligation, whilst for McLeod (2002) trust involves the expectation that the trustee will act out of moral integrity. None of these explore either the distrust/nonreliance distinction or the trust/distrust parallels, but we might on this basis develop an obligation account of both trust and distrust. Such an account would, however, wash out the distinctiveness of trust, distrust, trustworthiness and untrustworthiness: for example, the virtue of being trustworthy is not the very general virtue of meeting one's obligations. Relatedly, being trustworthy does not involve meeting all legitimate normative expectations, only those legitimate normative expectations which are distinctively associated with (dis)trust. I will try to vindicate these claims in the sections which follow, and will revisit the obligation account in the final section of the paper.

A final (for now) question about commitment: when I trust-or-distrust someone, must I believe that she has a commitment *to me*? Or is it enough that I believe that she has a commitment to someone or other?

I will take it that trust-or-distrust requires only that we think the person has a commitment to someone or other. For example, suppose your daughter's friend promises to her (not to you) that she will stay to the end of the party and give your daughter a lift home. Suppose you rely upon the friend to keep this promise: you drink several glasses of wine, making it impossible for you to safely drive and fetch your daughter yourself. I will take it that you trust your daughter's friend to keep her promise to your daughter. It seems that Holton (1994) and Jones (2004) likewise permit trust outside of direct relationships: you are liable to resent your daughter's friend if she proves unreliable, though it is debatable whether she then owes an apology to you, or only to your daughter.

In contrast, motives-based accounts typically, though not inevitably, presuppose a direct relationship: on such accounts, when I trust someone, I take her to bear me goodwill, to respond to my dependence, or to consider my interests.

The commitment account could in principle be restricted so that genuine trust-or-distrust is only available when we think someone has made a commitment to us. We might then say that you judge your daughter's friend to be trustworthy, that you believe it's appropriate for your daughter to trust her, but that strictly speaking you can neither trust nor distrust her yourself. I will avoid this restriction, though I will return to the issue when I discuss betrayal, in section 7.1.

6.2 *The Confidence Trickster*

According to Holton (1994, 65), "the confidence trickster might rely on your goodwill without trusting you". This is raised in objection to Baier's goodwill account of trust; does it also threaten the commitment account? You foolishly promise to send a \$10,000 down-payment to release your million-dollar 'inheritance', and the trickster relies upon you to fulfil this commitment, acting under the supposition that you will do so. If Holton is right, the trickster nevertheless does not trust you, in apparent conflict with the commitment account.

Could I insist that the trickster does trust you? This would demand an explanation of why Holton and others find it obvious that he does not. The trickster is a Bad Man and untrustworthy himself, he does not accord you the respect which normally accompanies trust; perhaps even he respects you less than he respects people who are untrustworthy in this domain. And perhaps all this muddles our intuitions about whether the trickster trusts you to send the money. (Baier might try this response to Holton.)

But I can avoid the 'counterexample' without needing to deny Holton's intuition. Although you don't realise it you do not have a genuine commitment, the trickster recognises this, and this is why he does not trust you. You have issued a 'promise' under a misapprehension, indeed to the very person who has deceived you about your situation, and such 'promises' do not bind. This explains why the confidence trickster does not distrust you either: he recognises that you are not an appropriate target of either trust or distrust in respect of this transaction. Thus the commitment account is not undermined by Holton's confidence trickster.

7 **Being Trusted, Being Distrusted**

I now turn to questions about (un)trustworthiness and the appropriateness of trust-or-distrust, showing how the notion of commitment can help us understand these issues. My main concern is to argue that an effective theory of trust, distrust, trustworthiness and untrustworthiness must explain what makes it reasonable to have (dis)trust-relevant normative expectations of others. As I show, invoking commitment is one good way of doing this.

7.1 *What's Wrong With Violating Trust?*

If someone relies upon a rope bridge to hold her weight, but the bridge is unreliable, the practical consequences might be very serious, even unto death. If someone relies upon Kant's regular morning walk as a reminder to take her medication, but Kant

is unreliable in this respect, this might also lead to the ultimate bad consequence. Relying on the unreliable can be dangerous.

But trusting the untrustworthy can generate harms which go beyond those caused by relying on the unreliable. One sort of harm is emotional damage caused by discovering that you have trusted someone who turned out not to deserve your trust. This sort of damage can have long-ranging and complex consequences. (Jones (2004) usefully writes of “basal (in)security” in this context, whilst Cass R. Sunstein (2007) discusses our aversion to “betrayal risk”, though without strictly marking the trust-reliance distinction.) Important though such harms are, I will set them aside here in order to discuss the kind of distinctive harm done when trust is violated, even if the violation is never discovered and so does not cause emotional damage.

It does seem that there is something intrinsically problematic about betrayals or violations of trust, something in addition to both the practical consequences of unreliability, and the emotional consequences of discovering such violations: the wrong of lying, cheating and promise-breaking goes beyond its contingent practical harm. I will argue, however, that being trusted is neither a necessary nor a sufficient condition for the possibility of such betrayal.

Trust is not necessary, because we can be betrayed by those we distrust, and indeed distrust can include expectation of betrayal. You can know that you will be betrayed: Jesus knew that Judas would betray him. Yet trust is impossible in such a situation, so betrayal does not require trust. Relatedly, if I lie to you and you are not misled, the practical damage may be mitigated, yet I harm you nonetheless, and not just through any emotional damage I may cause. Finally, if you trust me and I prove untrustworthy, my behaviour does not become acceptable once your trust ceases.

In such cases, there may be practical consequences of unreliability, and/or emotional consequences of the discovery of untrustworthiness. But, on the assumption that there is a further, intrinsic wrongness in violating trust, there is likewise a further, intrinsic wrongness in being untrustworthy even when you are not trusted.

Conversely, we do not always do wrong when we knowingly disappoint those who trust us. Distinguish two kinds of mistake in trusting: trusting someone who merits distrust, and trusting someone in a respect which merits only reliance or nonreliance. The first of these involves recognising that someone is a suitable target for our normative expectations, but misjudging their reliability: this kind of mistake results in betrayal. But the second is a mistake about whether someone is a suitable target for normative expectations in this respect: if such a person proves unreliable, we may feel betrayed, but we have not been betrayed, and the trustee may rightly complain of our moralising presumption.

The commitment account can explain all this: we do wrong when we fail to fulfil a commitment, absent mitigating circumstances. (Investigating the nature of this wrong would stretch the present article beyond breaking point, but a natural thought is to look to the literature on promises and promissory obligations for help.) This is independent of whether we are in fact expected to fulfil the commitment, i.e. whether we are trusted or distrusted in this respect. Being trusted is not necessary. Moreover, if someone mistakenly thinks I have a commitment, and I do not act as

that supposed commitment would require, then I have not in fact behaved badly. Letting down someone who trusts is not sufficient for morally-problematic violation of trust.

Granted, in some circumstances we acquire commitments simply by allowing others to continue to rely upon us, or by allowing others to think that we have commitments. And we cannot always minimise commitment and thus minimise the risk of unmet commitments: this fetishisation of honour would in any case make for an unrewarding, lonely life. Nevertheless, we are not entirely at the mercy of others' decisions to trust us.

I argued above that trust-or-distrust does not require a direct relationship: you trust your daughter's friend to keep her promise to your daughter, even though the friend has made no commitment to you. But are you really betrayed when someone fails to meet a commitment to a third party? It seems more plausible to say that the only person betrayed is the person to whom the commitment is made. This weakens, though does not break, the connection between trust and betrayal. The alternative is to rule that genuine trust-or-distrust is available only to those to whom commitments are made.

This uneasy dilemma is not generated by the commitment account *per se*; rather, it arises from a tension between two tempting thoughts about cases like that of your daughter's friend. On the one hand, it is natural to think of your attitude towards your daughter's friend as trust: you are disposed to hold many of the trust-related reactive attitudes towards your daughter's friend, and reasonably so. On the other hand, it is natural to think that it is only your daughter who is betrayed if her friend breaks her promise to her. Any account of trust (and distrust) must make a difficult choice between retaining these two intuitive judgements whilst weakening the trust-betrayal connection, or else retaining the trust-betrayal connection whilst rejecting one of the two intuitions.

Motives-based accounts of trust treat distrust as an afterthought and thus struggle to explain the wrong involved in violating trust. On such accounts, to trust someone to do something is to rely upon her to do it for the right sort of motives. Such trust is disappointed either when the task is not completed, or else when the motives are 'wrong'. This makes unfeasibly heavy demands upon trustees, putting us in moral jeopardy, liable to betray—not just disappoint—people who have unrealistic expectations of us. My colleagues don't care about my interest in drinking champagne, nor do they bear me goodwill in this respect; surely they cannot be criticised for this. But if I foolishly persuade myself otherwise, they will subsequently 'betray' me when they fail to buy champagne.

Moreover understanding trust (and distrust) in terms of normative expectations does not by itself explain when such expectations are appropriate. If I unreasonably develop normative expectations of my colleagues in respect of champagne-buying—whether or not I also predictively expect that they will buy me champagne—then I will wrongly feel betrayed when they do not buy me champagne.

7.2 *Trustworthiness and Untrustworthiness*

Trustworthiness and untrustworthiness can seem more central to moral philosophy than do trust and distrust themselves (Hardin 2002; Potter 2002; Putnam 2000,

136). Trust sometimes has value in its own right, but usually this is conditional on the trustworthiness of the trusted person: Linda Zagzebski argues that the virtue of trust is “a mean between gullibility and suspiciousness” (1996, 160–1), i.e. wise trust (and distrust), not trust *per se*. (In contrast, Thomas Hurka argues that trust is the virtue of “believing beyond one’s evidence in the virtues of particular people, such as one’s friends and family” (2001, 108), but see Hawley (forthcoming).) We owe it to ourselves and to others to place our trust appropriately, where such appropriateness depends at least in part on the trustworthiness of the trustee; this is especially important where we make decisions about trust which affect third parties. Moreover trustworthiness is an admirable trait, whilst untrustworthiness merits condemnation.

Motives-based accounts of trust make it surprisingly difficult to understand what it is to be trustworthy in general. To be trustworthy in some specific respect to some specific person is simple enough: it is to be such that it is appropriate for that person to trust you to perform that task. So for Hardin, to be trustworthy in some specific respect to some person is to have adopted her relevant interests and to act on these; for Jones (1996) it is to bear that person goodwill in that domain, be competent, and respond to her counting upon you.

But these accounts do not generalise: if this is specific trustworthiness, then being generally trustworthy is not a matter of being trustworthy to lots of people in lots of respects, and being absolutely trustworthy is certainly not a matter of being trustworthy to everybody in every respect. If trustworthiness is thought of as a kind of well-meaning helpfulness, it will always be possible for us to subjugate more of our own interests to those of others, demonstrating more goodwill, competence and responsiveness. We will be pulled in different directions by the whims of those around us, unable to be trustworthy to all. Worse, if trustworthiness were a generalised tendency to respond to people’s trust, rather than to their interests or needs, then it would be relatively easy for distrusted people to be trustworthy, since they have little trust to respond to.

The account of trust in terms of normative expectations doesn’t help here. Trustworthiness cannot in general be a matter of responding to others’ normative expectations, because such expectations may be unreasonable, and defying unreasonable expectations is not a sign of untrustworthiness. Trustworthiness involves responding to reasonable or appropriate normative expectations, but this returns us to the question of what makes such expectations appropriate.

On the commitment account, trustworthiness requires us to ensure that our commitments do not outstrip our actions. This requires judiciousness in acquiring commitments as well as doggedness in fulfilling commitments already acquired, independent of others’ expectations. Trustworthy people must sometimes disappoint up-front by refusing new commitments, rather than violate trust later on: this is the moral ‘power of no’. We can coherently aspire to general trustworthiness, even if none of us ever quite achieves this.

The commitment account of trust and distrust enables us to understand general trustworthiness. What about general trust? I have used a three-place model—A trusts B to do C. But we also speak of trusting one person or distrusting another, without reference to any specific task. We can now understand general trust as

reliance on someone to fulfil whatever commitments she may have, and general distrust as lack of such reliance. Some trust is intermediate between the specific and the completely general: I may trust someone to fulfil her financial commitments, without trusting her to fulfil commitments in her personal life.

So the commitment account of trust and distrust illuminates the relationships between specific trust, specific trustworthiness, general trust and general trustworthiness. If general trustworthiness is a virtue in a rich sense, however, then it must be deeper than a pattern of commitment-meeting behaviour; Potter (2002) develops a virtue account of trustworthiness, whilst McLeod (2006, 8) outlines the range of options here. At the very least, we might expect trustworthy people to be disposed to be motivated by their commitments. I cannot properly engage with the complexities of what it is to act for a reason, but will briefly outline the role of motivation within this picture.

According to the commitment account, you can trust someone to do something without expecting her to be motivated by her commitment. You may trust me to do something because you believe both that I have a commitment to do it, and that I will do it, without believing that I will do it *because* of my commitment. Maybe I am motivated by pure enthusiasm. Would commitment have motivated me in the counterfactual absence of enthusiasm? Perhaps then there would have been no commitment: this needn't indicate that I am untrustworthy.

To be trustworthy in some specific respect, it is enough to behave in accordance with one's commitment, regardless of motive. What about general trustworthiness? In practice, sadly, none of us has independent reasons to do all the things we are committed to doing. So general trustworthiness encompasses more than the 'fair-weather' cases in which we don't need the motivation of commitment: a generally trustworthy person will often meet her commitments simply because they are her commitments, although this is not a requirement of trustworthiness in any specific respect.

8 Trust and Testimony

I have focused so far on trusting or distrusting someone to do something. But we should also consider trusting or distrusting what someone says or writes; Edward Hinchman (2005) and Paul Faulkner (2007, 2011) make explicit connections between trust and testimony, and whilst Richard Moran (2005) does not explicitly discuss trust, his 'assurance' view is related to these concerns. If we can flesh out this connection, then a well-developed account of trust and distrust will improve our understanding of testimony; moreover trustworthiness is a strong candidate for being both an intellectual and a moral virtue (Zagzebski 1996, 158–60). Unlike some other accounts, the commitment account of trust and distrust can explicate trust or distrust in what people say.

Trusting a speaker in some particular instance seems to be a special case of trusting someone to do something: this is no accident of terminology. But what exactly is a trusted speaker trusted to do: trusted to speak the truth, or perhaps to

speak knowledgeably? Sometimes these coincide—if I speak knowledgeably, then I speak truthfully—but not always, for I can unknowingly speak the truth.

Recent work has identified various possible norms of assertion—conditions under which it is proper to make an assertion—and there is disagreement about whether knowledge, truth, belief, or perhaps warranted belief is key (e.g. Williamson (2000), and for a survey Weiner (2007)). I will take it that trusting someone as a speaker is a matter of trusting that person to assert *properly*, whatever proper assertion amounts to.

Assertion can be understood in terms of commitment: commitment to asserting properly, however that placeholder notion is filled out.² Indeed, this gives us another way of framing the question about norms of assertion: what commitment do we make in making an assertion? Then the commitment account of trust and distrust can explain why questions of trust are restricted to assertions, not just any old utterances. We neither trust nor distrust what people say when they are reciting poetry, or sleeptalking: we may wonder about the truth of sleeptalk, but not accuse a sleeptalker of lying or praise her honesty. Recitation, sleeptalking and the like do not involve commitment, so are outside the realm of trust and distrust.

We can sometimes opt for reliability without inviting trust: I am happy for my partner to rely upon my cooking, but not for him to treat this as a matter of trust. But I cannot opt to be a merely reliable asserter who does not invite trust, and I cannot make improper assertions without displaying untrustworthiness. This is because assertion involves commitment: if I want to present a claim without commitment, I must present it as mere speculation or guesswork.

Trusting someone to do something involves both trust in her competence and trust in her willingness to act; part of trustworthiness is the attempt to avoid commitments you are not competent to fulfil. The special case of trusting what someone says involves both trusting her to possess the truth (or knowledge), and trusting her to speak sincerely. People can betray our trust through either error or deceit, and part of trustworthiness is the attempt to avoid making assertions where you lack knowledge (Hursthouse (1999, 11–12) discusses the rich complexity of honesty).

So we can accommodate trust and distrust in what others say, via the natural move of understanding assertion as commitment. But motives-based accounts of trust do not easily accommodate the important special case of trusting and distrusting what someone says, for a number of reasons. First, such accounts do not adequately characterise distrust. We do not trust what people say in sleep, or in recitation, but nor do we distrust them under such circumstances.

Second, such accounts cannot explain our decisions about whether to trust what others say when they are unconcerned about our needs or desires. For Hardin, to trust someone is to take them to have encapsulated your interests amongst their own. If you ask me a question, you reveal your interest (in both senses) in knowing the answer: if I share your interest in your knowing the answer, then I will tell you the truth if I can. So far so good for Hardin. But suppose you are stuck next to the pompous pub bore, who is forever telling you things without waiting for you to ask, and without caring either way about your interests. It is an open question whether

you should trust what he says, even though it is clear he has not encapsulated your interests amongst his own.

Jones's (1996) account has a similar problem: for her, to trust someone is to be optimistic about their goodwill and competence within a particular domain, and to expect them to be moved by the thought that you are counting on them. If you ask me a question, and I answer, then trusting what I say may well fit this pattern of optimism about responsiveness. But the pub bore bears you no goodwill, and cares not whether you are counting on him, yet the question of whether you should trust what he says remains open. (This difficulty also afflicts Faulkner (2007)).

What does the commitment account tell us about the pub bore? So long as the bore is making assertions, he is making commitments (to speak the truth, or perhaps to speak knowledgeably). Thus you face the question whether or not to regard these assertions as proper, whether or not to act on the presupposition that the bore is fulfilling his commitments. You may suspend judgement, if there is nothing immediately at stake. Nevertheless, the commitment account can make sense of the difference between trusting and distrusting under these circumstances.

I have imagined the pub bore addressing you directly, though not out of any concern for you. Then he is making commitments to you, and you can trust or distrust as you see fit. But what if you are eavesdropping on someone who is not directly addressing you, not making commitments to you? In my view, you can nevertheless trust-or-distrust this speaker, but a more restrictive version of the commitment view would rule this out (see section 6.1).

9 Obligation or Commitment?

The commitment account of trust and distrust has many strengths. It is decisively better than motives-based accounts, and more illuminating than the accounts offered by Holton (1994) and Jones (2004). But, as I previously noted, an obligation account of trust and distrust shares many of its strengths, including the symmetry between trust and distrust, neutrality about the trustee's motives, and the promise of acceptable accounts of both general trustworthiness and the wrongness of violating trust. Since commitments typically induce obligations, the commitment account and the obligation account will give the same verdict about many cases (McLeod 2002, 30). Moreover, cases like the excess-lunch bringer, the non-champagne-buying colleagues and the regularly-cooking partner are all ones in which, so long as there is no commitment, there is no obligation. Thus many of my arguments lend support to both the commitment account and the obligation account. Indeed the main goal of my paper has been achieved if you are by now persuaded that one or other of these must be correct.

That said, I think there are cases in which we have obligations but no commitments, and where neither trust nor distrust is appropriate. If I am right, this tells in favour of the commitment account. Suppose we are in the Wild West. In town, there is an uneasy truce, a semblance of law and order. Out in the desert, as everyone knows, there are no holds barred; you take your life into your hands if you venture there. You and I meet by chance in the desert, and

you see that I am armed (of course). Before we exchange words, you may try to predict whether or not I will let you live. But no matter how confident you are in that prediction, it's not appropriate for you either to trust or to distrust me in this respect. Out here, no one has promised, implicitly or explicitly, to respect life and limb. I have no commitment to let you live, and you should neither trust or distrust me in this respect. If I let you live, I do not demonstrate trustworthiness; if I shoot you, I display many vices, but not that of untrustworthiness. But of course I still have an obligation to let you live, whether or not I recognise this.

Now, such cases are clearly rather unusual, and intuitions may reasonably differ about whether trust and distrust are appropriate in such situations. Moreover, intuitions may reasonably differ about whether commitment has really been cancelled in such cases; this requires us to distinguish between commitment in the psychological sense and commitment of the kind which can persist even when the committed person has no intention of fulfilling that commitment. So such cases are not decisive in favouring the commitment account over the obligation account.

But there is also a more abstract reason to expect some separation between issues of obligation and issues of trust and distrust. Trustworthiness is a relatively thin trait: the requirements which must be satisfied if we are to be trustworthy place only conditional constraints on what we should do, because they do not place constraints on which commitments we should incur. Nevertheless the concept of trustworthiness seems to have at least some substantive content: on the scale between thin and thick, it lies at neither extreme. Likewise for the concept of untrustworthiness. (And honesty and dishonesty seem thicker still.) To characterise someone as trustworthy is to say something fairly specific about her; to aspire to trustworthiness is to aspire to a certain kind of behaviour and attitude; to judge that one's past behaviour was untrustworthy is to make a fairly specific kind of criticism of oneself; we can explain why certain behaviour is wrong by pointing out that it exemplifies untrustworthiness.

But the concept of generally doing as one is obliged to do is about as thin as can be. To characterise someone in these terms is give her a positive moral evaluation, but not along any particular dimension; there are various ways in which we can fail to meet our obligations, not just through untrustworthiness; saying that certain behaviour is obligation-violating is an unilluminating 'explanation' of why it is wrong. These differences between obligation-meeting and trustworthiness suggest a conceptual distance between obligation and trust, some reason to look for a candidate account of trust and distrust which does not rest entirely on notions of obligation. And the commitment account fits the bill nicely.

Notes

¹ Lots of people have helped me with this, including Antony Duff (who suggested a crucial example), several anonymous referees, and audiences at Trinity College Dublin, the Mangoletsi lectures in Leeds, the Epistemic Inclusiveness conference in Copenhagen, the Crossing Borders conference in Vienna, and, of course, St Andrews.

² This is much weaker than the claim that assertion also involves commitment to either vindicate or withdraw once challenged (e.g. MacFarlane 2005). Moreover, my commitment view of assertion does not entail Moran's (2005) view that assurance, a commitment-like notion, is of great epistemic significance in testimony.

References

- Baier, Annette (1986). "Trust and Antitrust" *Ethics*, 96: 231–60.
- Faulkner, Paul (2007). "On Telling and Trusting," *Mind* 116: 875–902.
- (2011). *Knowledge on Trust*, Oxford: Oxford University Press.
- Fricker, Miranda (2007). *Epistemic Injustice*, Oxford: Oxford University Press.
- Hardin, Russell (2002). *Trust and Trustworthiness*, New York: Russell Sage Foundation.
- Hawley, Katherine (forthcoming). "Partiality and Prejudice in Trusting," *Synthese* (published 'Online First' in May 2012, forthcoming in print).
- Hieronymi, Pamela (2008). "The Reasons of Trust," *Australasian Journal of Philosophy*, 86.2: 213–236.
- Hinchman, Edward (2005). "Telling as Inviting to Trust," *Philosophy and Phenomenological Research*, 70.3: 562–97
- Hollis, Martin (1998). *Trust Within Reason*, Cambridge: Cambridge University Press.
- Holton, Richard (1994). "Deciding to Trust, Coming to Believe," *Australasian Journal of Philosophy* 72: 63–76.
- Hurka, Thomas (2001). *Virtue, Vice and Value*, New York: Oxford University Press.
- Hursthouse, Rosalind (1999). *On Virtue Ethics*, Oxford: Oxford University Press.
- Jones, Karen (1996). "Trust as an Affective Attitude," *Ethics* 107: 4–25.
- (2002). "The Politics of Credibility," in *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, edited by Louise Antony and Charlotte Witt, Boulder CO: Westview Press.
- (2004). "Trust and Terror," in *Moral Psychology*, edited by Peggy DesAutels and Margaret Urban Walker, Lanham MA: Rowman and Littlefield.
- MacFarlane, John (2005). "Making Sense of Relative Truth," *Proceedings of the Aristotelian Society* 105: 321–39.
- Marsh, Gerald (2011). "Trust, Testimony and Prejudice in the Credibility Economy," *Hypatia, a Journal of Feminist Philosophy* 26.2: 280–93.
- McGeer, Victoria (2008). "Trust, Hope and Empowerment," *Australasian Journal of Philosophy*, 86.2: 237–54.
- McLeod, Carolyn (2002). *Self-Trust and Reproductive Autonomy*, Cambridge MA: MIT Press.
- (2006). "Trust," *Stanford Encyclopedia of Philosophy*.
- Moran, Richard (2005). "Getting Told and Being Believed" *Philosophers' Imprint*, 5.5.
- Nickel, Philip J. (2007). "Trust and Obligation-Ascription," *Journal of Ethical Theory and Moral Practice*, 10: 309–19.
- O'Neill, Onora (2002). *Autonomy and Trust in Bioethics*, Cambridge: Cambridge University Press.
- Pettit, Philip (1995). "The Cunning of Trust," *Philosophy and Public Affairs* 24.3: 202–25.
- Potter, Nancy Nyquist (2002). *How Can I Be Trusted?*, Lanham: Rowan and Littlefield.
- Putnam, Robert D. (2000). *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon and Schuster.
- Sunstein, Cass R. (2007). "Moral Heuristics and Risk," in Tim Lewens (ed.) *Risk: Philosophical Perspectives*, Abingdon: Routledge.
- Weiner, Matt (2007). "Norms of Assertion," *Philosophy Compass* 2: 187–95.
- Williamson, Timothy (2000). *Knowledge and its Limits*, Oxford: Oxford University Press.
- Zagzebski, Linda (1996). *Virtues of the Mind*, Cambridge: Cambridge University Press.