

Negotiating Domains of Trust

This is a PREPRINT - Please cite published version

Forthcoming in *Philosophical Psychology*

Elizabeth Stewart

Abstract

When trust is broken, how should we determine who is at fault? Previous discussions of broken trust typically attribute the fault to trusters who place trust foolishly or trustees who act in an untrustworthy manner. These discussions take for granted the ability of the truster and trustee to communicate and understand the boundaries of what is being entrusted, that is, the domain of trust. However, the boundaries of entrusted domains are not always clear to either party which can result in broken trust despite the best efforts of both truster and trustee. In this paper, I argue that determining who to blame when trust is broken is a messy affair in which disagreements over fault regularly arise. I introduce three features of trust domains that take center stage in negotiations regarding who is at fault when trust is broken: scope, rigidity and ordering.

Keywords: Trust, Trustworthiness, Disagreement

Word Count: 10,230

Introduction

When we tell someone that we trust them, we generally do not mean that we trust them *with everything*. Rather, we mean that we trust them regarding some particular domain of interest. Formally, we can express the trust relationship as a three-place relation: A trusts B with respect to domain C (Baier, 1986; Jones, 1996; D’Cruz, 2018; Hawley, 2014; Hawley, 2019; D’Cruz, 2020; Kelp and Simion, 2022)¹. When I tell the hairdresser that I trust them, I mean that I trust them to make my hair presentable, not that I trust them to perform dental surgery on me or commit themselves to spend the rest of their lives with me. Likewise, I trust the dentist with my teeth and my partner with sharing in my life journey. The domain specificity of trust is nothing new in the literature (Baier, 1986; Jones, 1996; Hardin, 2002; Hardin, 2006; Nguyen, 2019), yet little is said about what a domain is or what determines the boundaries of a given domain. Are the boundaries determined by societal norms? Individual ideals? Platonic forms?

Why does it matter how we think about domains? Recently, several philosophers have drawn attention to the challenges that trusters and trustees face in establishing and maintaining trust. I believe that many of these challenges have to do with misunderstanding the boundaries of trust domains. For example, Karen Jones points out that when “trust assigns discretionary powers to the one-trusted, then there can be legitimate dispute as to whether or not that trust has been fulfilled. What might, at first look, appear a failure in responsiveness to the fact of someone’s dependency might, on closer inspection, be revealed as a creative way of meeting it” (Jones, 2017, p. 104). That is, the trustee misunderstands the extent to which their discretionary actions fall within the scope of the domain of trust.

Furthermore, Katherine Hawley describes how sometimes when we accept another’s trust, there is “an implicit commitment to take on further, more specific commitments in the

¹Trust is also often understood as a two-place relation wherein A trusts B. It is an open question how two-place trust relates to three-place trust, although recent work has made progress on addressing it ((Domenicucci and Holton, 2017; Faulkner, 2017; Kelp and Simion, 2022)). In this paper, I bracket this issue and focus on trust as a three-place relation.

future” (Hawley, 2019, p. 77). Hawley calls these commitments-to-take-on-commitments “meta-commitments”. These meta-commitments may “vary in strength and explicitness, and disagreements about these can be the source of deep problems in relationships, around mismatch of expectations. If you think that a friend needs a very good reason not to accept an invitation to take a vacation together, and I think that friends are free to choose whether or not to vacation together, then that will be a source of difficulty for our friendship when I turn down your invitation to join you in Blackpool” (Hawley, 2019, p. 77). In other words, there is some ambiguity in whether vacationing together is included in the domain of friendship.

The exercising of one’s discretionary powers and the ambiguity in identifying the specific commitments that follow from meta-commitments both pose domain-related challenges to establishing and maintaining trust. A closer inspection, then, of domains of trust helps clarify why these challenges arise and how they might be dealt with. In this paper, I explore how disagreements and misunderstandings about the boundaries of trust domains generate problems within trust relationships. These disagreements and misunderstandings can lead to broken trust, sometimes despite the best efforts of both truster and trustee. When this occurs, it is not always clear who is to blame for the broken trust.

This paper has several aims. I first wish to define trust domains and distinguish them from related concepts. Secondly, I differentiate three kinds of broken trust that arise from disagreements and misunderstandings regarding trust domains. Lastly I describe three features of trust domains that can generate these disagreements or misunderstandings and discuss how these features complicate the process of attributing blame when trust is broken.

1 Trust Domains

As noted above, the domain specificity of trust is frequently mentioned in accounts of trust as a three-place relation. For example, Karen Jones’s (1996) account of trust states that

trust is “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” (Jones, 1996, 4). Jones elaborates on the domain specificity of trust in the following:

This is not to say that the optimism itself is qualified and instead of being unreserved optimism is a qualified or restricted optimism. What is qualified is not the optimism itself, but the domain over which it extends. So, for example, the optimism we have about the goodwill and competence of strangers does not extend very far. We expect their goodwill to extend to not harming us as we go about our business and their competence to consist in an understanding of the norms for interaction between strangers. For a man to run up at full speed behind a woman on an ill-lit street is to display a lack of such competence, and, even if he was simply out for a late night run and meant no harm, he has given the woman reason to distrust him (Jones, 1996, 7).

Jones is here focused on understanding the requirements of competency, which are restricted to certain domains. We expect strangers to be competent with respect to the domain of basic social norms. Jones thus seems to suggest that trust domains include norms of interacting, although it is unclear whether she thinks that domains always include these. Indeed, Jones does not offer much explanation of how she understands the concept of domains, despite the central role that domains play in her later accounts of trust and trustworthiness. She defines trust and trustworthiness as:

Trust—“A trusts B in domain of interaction D if and only if, A has an attitude of optimism that B’s competence and responsiveness to the fact that she is being counted on will extend to cover that domain.” (Jones, 2019, p. 99)

Three-place trustworthiness—“B is trustworthy with respect to A in domain of interaction D, if and only if she is *competent* with respect to that domain, and

she *would* take the fact that A is counting on her, *were* A to do so in this domain, to be a *compelling* reason for acting as counted on.” (Jones, 2012, p. 70-71)

In what follows, I elaborate on what a trust domain might be and why a clearer concept of a domain can help make sense of how trust relationships can go awry. My hope is that this account of trust domains will complement existing work, such as Jones’ and Hawley’s, that view trust and trustworthiness as three-place relations.

As noted above, there are a range of different ways we might conceptualize trust domains including as individual ideals and social norms. If we think of domains as determined by social norms, then a domain is a kind of social construction that, at least in part, is determined by factors external to both truster and trustee. Viewing domains as social constructions has some advantages. For example, they could function as guidelines for both truster and trustee, indicating when trust is placed poorly and when trust has been betrayed. I have trusted poorly if I trust someone with something that lies outside of the social norms associated with that relationship. I trust my dentist poorly when I trust them to cut my hair instead of check my teeth for cavities. Similarly, I am an untrustworthy dentist if I can cut your hair, but not check your teeth for cavities. Understanding domains as social constructions, however, obscures some important dynamics of trust relationships. We feel betrayed, for example, not when someone acts in a way that is incompatible with social norms but when someone acts in a way incompatible with the expectations that we, as individuals, have of them. The expectations that matter to a trust relationship, I believe, are those held by the person doing the trusting, rather than the society to which they belong. Thus, I will treat trust domains as individualized phenomena, rather than social norms.

More precisely, I take trust domains to be sets of expected behaviors or attitudes held by individual trusters. Expectations come in many varieties. Some expectations are normative and express what we think should be done, but not necessarily what we think will actually be done. Other expectations are just the opposite, describing what we think will happen, but not necessarily what ought to happen. It is an open question which kind of expectation

is relevant to trust (Jones, 2012; Nguyen, 2019). While I think that what follows applies regardless of which variety of expectation one focuses on, many of the accounts of trust that I will reference view trust as involving normative expectations and so I will focus on normative expectations as well.

Our expectations determine the boundaries of what behaviors or attitudes are included in the trust domain. For example, when I consider the domain of dentistry, I expect behaviors like diagnosing dental problems and filling cavities and I expect attitudes like general goodwill towards myself. We can represent such trust domains formally as sets:

$$Dentistry_{trust} = \{diagnosing\ dental\ problems,\ filling\ cavities,\ general\ goodwill\}$$

Note, however, that each item in the set can be broken down into another set of expectations. For example, the expected behavior of filling cavities includes expected behaviors like administering anesthesia, drilling out rotten areas, and replacing the drilled out area with the appropriate substance. Thus, trust domains can consist of *nested sets* of expectations:

$$Dentistry_{trust} = \{ \{diagnosing\ dental\ problems\}, \{administering\ anesthesia,\ drilling\ out\ rotten\ areas,\ filling\ area\ with\ appropriate\ substance\}, \{general\ goodwill\} \}$$

These domains are determined by the expectations of individual trusters. Thus, trust domains can differ among different people. A and B might hold different expectations regarding the domain of dentistry, in which case their trust domains diverge accordingly. In the above examples, I haven't qualified this. Trust domains would be better characterized as *domains given the individuals whose expectations comprise the domain*:

$$Dentistry|A_{trust} = \{\{diagnosing dental problems\}, \{administering anesthesia, \\ drilling out rotten areas, filling area with appropriate substance\}, \\ \{general goodwill\}\}$$

Importantly, what a truster expects may differ from what a trustee *believes* that a truster expects. It is important to recognize that a trustee’s perceptions of a trust domain can, and often do, come apart from the trust domain itself. Thus I will call a trustee’s perception of a trust domain a “perceived domain”. A perceived domain is the set of behaviors and attitudes that a trustee believes is expected of them in a trust relationship. Ideally, the trustee holds accurate beliefs about what the truster expects of them and thus the perceived domain should be identical to, or at least similar to, the trust domain. We can represent the ideal case where a dentist, B, accurately perceives a patient, A’s, expectations as follows:

$$B(Dentistry|A)_{per} = \{\{diagnosing dental problems\}, \{administering anesthesia, \\ drilling out rotten areas, filling area with appropriate substance\}, \\ \{general goodwill\}\}$$

A perceived domain is different still from the set of behaviors and actions that a trustee commits to doing when they encounter another person’s trust, which I call the “response domain”. Whenever someone recognizes that trust is extended to them, they face some decision about how to respond. Should the trust be accepted or rejected? If accepted, what sorts of attitudes or behaviors will be undertaken in order to meet the trust? I am not here talking about what the trustee ought to do, but the things that they intend or plan to do.²

²Importantly, when I speak of what trustees commit to doing in response to trust, I am not using “commitment” in the sense that Katherine Hawley uses in her (2014) account of trust (Hawley, 2014). Instead, I am using it in the broader psychological sense of intending or planning to do something.

Thus, when the trustee doesn't plan to do anything in response to the trust, their response domain remains empty. This may be appropriate when the trust domain is inappropriate. For example, when I get an email from a scammer asking if they can trust me to send them money so that they can start a new business that will most assuredly make both of us rich, I have good reason to do nothing in response.

However, in accepting extended trust, a trustee ought to have some idea of what they will do to meet that trust. The attitudes or behaviors that are included in the response domain, however, may differ from those in the trust domain. This is not always a problem and, indeed, is often beneficial. When a patient goes to the dentist, they may simply expect the dentist to diagnose any dental problems. The average patient, however, probably isn't aware of the domain of actions that are nested within the fairly coarse-grained expectation of "diagnosing dental problems". That is, they are most likely unaware of what dental problems to look for or how to do so effectively. The average dentist, however, hopefully is aware of what actions should be nested within this coarse-grained expectation and is also probably aware that the patient is ignorant of these. Thus, what the patient expects, what the dentist thinks the patient expects and what the dentist intends to do in response to the expectations all come apart. While a response domain may be different from a trust domain, a positive feature in cases like dentistry, it should still ideally be compatible with the trust domain. As I am not a dentist, I will refrain from attempting to spell out the range of actions that dentists should intend to do when accepting the trust of patients. It is enough to recognize that the range of actions the dentist undertakes when accepting trust will be far more fine-grained than the average non-dentist's set of expectations. Thus, the response domain for dentist, B, who has accepted patient, A's, trust, will look generally like the following:

$$B(\text{Dentistry}|A)_{res} = \{B's \text{ fine - grained attitudes and behaviors towards } A \\ \text{with respect to dentistry}\}$$

This understanding of trust, perceived, and response domains leaves open the possibility that some individuals might have expectations or actions in their domain that seem very weird or inappropriate to others. For example, it is possible on my account of trust domains for someone to expect their dentist to cut their hair, even though most people would agree that cutting hair should not be included in the domain of dentistry. Similarly, a dentist might intend to cut their patient's hair alongside checking their teeth for cavities. Is the potential weirdness of domains a problem for my account? I do not think so. People are often unpredictable in their beliefs of what is normal or expected. Indeed, some people hold seemingly bizarre expectations of others. However, when an individual's expectations or commitments violate social norms, we have good reason to think that they have trusted poorly or are culpable for failing to meet the trust of others. Thus, while domains can vary widely and include weird or bizarre expectations, attitudes or actions, we need not think that all domains are appropriate. Given the social norms regarding dentistry, most people would agree that the individual who expects their dentist to cut their hair is misguided, to say the least. Likewise, if my dentist tries to cut my hair, I would be deeply confused and a little alarmed. Thus, while social norms do not restrict the boundaries of trust domains on my account, they do play an important role in determining whether a given individual's trust domain or commitment domain is sensible. Of course, social norms are not always a good guide for what we ought to expect or do. Sometimes upsetting social norms is the morally appropriate thing to do. In such cases, we might expect there to be some disagreement or confusion over the sensibility or appropriateness of such trust or response domains. I will explore these dynamics later in the paper.

In addition to helping us determine the reasonableness of others' trust or response domains, social norms also help trustees to accurately perceive others' trust domains. Dentists can be reasonably confident that their patients expect them to check for cavities rather than cut their hair because this is what is socially normal for dentists to do. Thus, while trust, perceived, and response domains are relativized to particular individuals, social norms play

several important roles in shaping and evaluating trust relationships. While it might be useful to introduce a fourth kind of domain to capture the social norms associated with a given sphere of activity, such as dentistry, I have resisted doing so here because I am here interested in the dynamics of individual trust relationships, which, ultimately, depend on the truster's and trustee's particular (potentially bizarre) expectations and actions.

I believe this account of trust domains is compatible with domains as they are described in most accounts of three-place trust. It is, for example, compatible with Jones' stating that we trust others to be competent with respect to social norms, because our individual expectations are often shaped by the society and culture to which we belong. Indeed, this shared understanding is important for accurately perceiving one another's trust domains. However, the expectations that bound trust domains also arise from a variety of sources beyond shared norms, including our knowledge, our personalities, prior experiences, and the like (Potter, 2002; Scheman, 2020). Additionally, these expectations may be distorted due to prejudice or bias, leading us to distrust when we ought to trust and vice versa (Medina, 2020). In practice, then, I think it is actually often quite difficult to pick up on what others expect of us and even when we do, we may find their expectations inappropriate. When this happens, trust is easily broken. In the following section I will elaborate on how trust-breaking occurs.

2 Types of Trust-Breaking

A trust relationship involves a truster forming some set of expectations of another, a trustee who attempts to understand these expectations and who commits to a set of actions in response. A trustee breaks a truster's trust when their actions or attitudes violate the expectations in the trust domain. There are three different ways that this can occur: 1) they may inaccurately perceive the trust domain, 2) they may accurately perceive the trust domain, but commit to actions that are outside its limits, or 3) they may accurately perceive

the trust domain, intend to fulfil those expectations, but violate a hidden expectation.

I will use an example from Annette Baier’s (1986) paper to demonstrate the differences between these types of trust-breaking. In the example, a babysitter is entrusted with the care of a child. Now, the babysitter has some discretion in how to go about meeting this trust. That is, they have some discretion over what actions to include in their response domain. In Baier’s example, the babysitter thinks the nursery would be improved with a coat of purple paint and sets to work accordingly. In taking on this task, Baier says that the babysitter has not acted in a trustworthy manner. She argues that the babysitter is untrustworthy because they have obviously exercised their discretionary powers poorly. However, I will argue that it isn’t immediately clear what, precisely, has gone wrong in this trust relationship or whether we should consider the babysitter untrustworthy for how they’ve exercised discretion in responding to the trust. I am proposing that there are three distinct ways in which the babysitter might have violated the parents’ trust and in each type of trust-breaking, it is not obvious that the babysitter is at fault for the broken trust.

2.1 Direct Misunderstanding Mismatch

Firstly, it is possible that the babysitter simply did not accurately perceive what the parents expected of him. Perhaps he misheard their instructions. Or, perhaps every previous babysitting gig involved repainting nurseries and so he assumed this one did as well. Either way, the trust domain differs from the perceived domain:

$$Childcare|Parents_{trust} = \{feed\ child,\ change\ diapers,\ put\ child\ down\ for\ nap,\ age\ appropriate\ play\}$$

$$Babysitter(Childcare|Parents)_{per} = \{feed\ child,\ change\ diapers,\ put\ child\ down\ for\ nap,\ age\ appropriate\ play,\ upgrade\ bedroom\}$$

In this case, the babysitter simply misunderstood what expectations were included in the trust domain. That is, there is a mismatch between the perceived domain and the trust domain. I call this first kind of trust-breaking a *direct misunderstanding mismatch*. When these mismatches occur, trust is broken as trustees act on their misperceptions of what is entrusted.

Broken trust from a direct misunderstanding mismatch may be the fault of either the trustee or truster. We might attribute fault to the trustee when they agree to do something that they actually don't know how to do and thus fail to meet the trust. If someone commits to babysitting, but they don't actually know anything about children, then it is unlikely that they will accurately perceive what the parents expect of them. Even in cases where the trustee is familiar with the entrusted task, they may still be at fault if they fail to take the appropriate measures to listen to and understand the particular truster's expectations. If the parents left written instructions asking the babysitter not to paint the nursery, but the babysitter failed to read the instructions because he felt confident that he already knew what was expected, then the parents would have good reason to blame the babysitter for their broken trust.

More problematically, epistemic injustice can cause direct misunderstanding mismatches. Trustees can demonstrate what Miranda Fricker terms "testimonial injustice" when they do not assign an appropriate level of credibility to the words of the truster (Fricker, 2007). This may happen if they wrongly make assumptions about the trust domain based on unfair biases or stereotypes, rather than on what the truster has actually said to them. Suppose, for example, that a woman trusts her male colleague to take care of his part of a work project and to not interfere with her part. That is, she expects him to let her do her share of the work. If her colleague, due to sexist stereotypes, fails to believe her when she says that she's got her part covered and thus sets about doing everything himself, he has misunderstood her, acted in violation of her expectations, and thus broken her trust. It is unclear how often cases of broken trust due to testimonial injustice actually take place as those who hold such

stereotypes often don't bother to listen to the truster in the first place.³ When it so happens that trustees don't listen to or even try to understand trusters' expectations and instead rely on unfair biases or stereotypes to guide their actions, they are similarly likely to break trust due to direct misunderstanding mismatches. In both cases, the trustee is at fault for the broken trust.

Sometimes, however, direct misunderstanding mismatches may be the fault of the truster, rather than the trustee. We might blame the truster for broken trust due to a direct misunderstanding when they fail to communicate their expectations clearly. Suppose the babysitter thinks he is supposed to arrive at 7:30pm because this is the time he has arrived for previous babysitting gigs for this family, but in this case, the parents expected him to arrive at 5:00pm. If the parents never communicated the change in expected arrival time to the babysitter, the babysitter can hardly be blamed for arriving late.

However, it is not always obvious who is to blame when a direct misunderstanding mismatch occurs. The trustee may argue that the limits of the trust domain were not communicated clearly. Likewise, the truster may argue that the trustee failed to take appropriate measures to listen and understand what was communicated about the limits of the trust domain.

2.2 Direct Conflict Mismatch

Secondly, it is possible in the babysitting case that the perceived domain is identical to the trust domain, but the response domain is incompatible with the trust domain. Suppose that the babysitter recognizes that the parents do not expect or want him to renovate the bedroom, but he decides to paint the nursery anyway because he thinks that it is in everyone's

³Many thanks to an anonymous reviewer for making this point.

best interest to do so:

$$\text{Childcare|Parents}_{trust} = \{\text{feed child, change diapers, put child down for nap,} \\ \text{age appropriate play}\}$$

$$\text{Babysitter}(\text{Childcare|Parents})_{per} = \{\text{feed child, change diapers, put child down for nap,} \\ \text{age appropriate play}\}$$

$$\text{Babysitter}(\text{Childcare|Parents})_{res} = \{\text{feed child, change diapers, put child down for nap,} \\ \text{age appropriate play, upgrade bedroom}\}$$

I call this kind of problem a *direct conflict mismatch* because trust is broken due to a direct conflict between the expectations in the trust domain and the actions in the response domain. Direct conflict mismatches are, I believe, the most commonly discussed type of broken trust. The standard case of broken trust is one in which the trustee accepts another's trust, but does not follow through. That is, the trustee fails to do what is necessary to meet the expectations of the truster. There are a range of reasons why trustees fail to meet the trust placed in them. I will just mention several here. The most straightforward reason is that trustees sometimes lie, assuring trusters that they will do something that they have no intention of doing. Alternatively, trustees may be honest in accepting trust, but may fail for other reasons. For example, Hawley discusses cases where trustees agree to do something, but they are not actually competent to do it and cases in which trustees over-commit themselves such that they cannot not meet all their commitments (Hawley, 2019).

However, not all cases of direct conflict mismatches are due to a moral failing on the part of the trustee. Sometimes, the failure occurs for reasons beyond the trustee's control. For example, if the babysitter fails to arrive on time because they got a flat tire, this is

not a moral failing. Katherine Dormandy suggests that when a trustee fails to meet the expectations despite trying their best within reason, this results in disappointment, whereas when the trustee fails to meet expectations because they did not try their best, this also results in betrayal (Dormandy, 2020).

Additionally, in some direct conflict mismatches, the trustee may refuse to meet truster's expectations because they believe those expectations are misplaced. That is, they disagree about what is appropriate to expect in a particular domain. Hawley's notion of "meta-commitments" may help further clarify how these mismatches occur. Recall that for Hawley a "meta-commitment" is a commitment that involves "an implicit commitment to take on further, more specific commitments in the future" (Hawley, 2019, p. 77). She provides the example, cited earlier, about two individuals who hold different views about whether vacationing together is an appropriate expectation to have of friends. While both parties may be committed to being a trustworthy friend, they have different notions of what kinds of things trustworthy friends do. That is, there is a direct conflict between the expectations of one friend and the response of the other. Direct conflict mismatches can thus arise when individuals have different ideas about what commitments follow from meta-commitments.

Just as in the case of direct misunderstanding mismatches, trustees can wrongly break trust when they inappropriately presume to know better than the truster. For example, the babysitter may think that while the parents don't expect a purple nursery, they *should* expect it. Some of these cases may look quite similar to those discussed in the section on testimonial injustice, with the key difference that in these cases the trustees do listen to the trusters and understand what is expected of them, they simply think that the expectations are misplaced or misguided. In such cases, trustees may act with the assumption that in the end, the trusters will see how misguided they were and be grateful that their expectations weren't met. Having one's trust broken in this way can be hurtful on multiple levels. Not only are expectations not met, but one's autonomy is also undermined in the process.

Yet sometimes trustees have legitimate reasons to respond to trust in ways that do

not align with the expectations in the trust domain. Trust can, for example, be exploitative (Dormandy, 2020). While trusters are obviously vulnerable to the actions of trustees, trustees are also often vulnerable to those who trust them (Baier, 1986; Dormandy, 2020). In such cases, trusters may place unfair expectations on trustees. When burdened with an unfair expectation, trustees have good reason to refuse if they are able. For example, if an employer entrusts an employee with a task over and beyond what they are paid to do, the employee has good reason not to do it. Resisting such unfair expectations is often key to enacting important social change, such as ensuring reasonable working hours and time off in employment settings.

Trust can also be genuinely misguided. Suppose a patient has chronic pain and trusts their doctor to make it stop. If this is not medically possible, however, the doctor should not attempt to meet this expectation. What then, should the doctor do? Ideally, the doctor will be able to explain why the expectations are not feasible and will be able to work with the patient so that they can develop a more reasonable set of expectations. Sometimes, however, this process may fail and the doctor may have to intentionally violate the patient's unrealistic expectations. I will explore such cases in further detail in section 3.

2.3 Indirect Misunderstanding Mismatch

Direct conflicts are not the only source of mismatches between the response domain and the trust domain. Recall that response domains can be, and often are rightly, finer-grained than the expectations that bound trust domains. Suppose that the parents entrust the babysitter with the general well-being of their child. The babysitter takes this task very seriously. Having read that purple rooms help foster a child's cognitive development and noting that the child's room is plain white, he sets about to remedy the situation with a few coats of purple paint. After all, the parents have entrusted this child's well-being to him and surely fostering a child's cognitive development is part of promoting general well-being. In this

case, the following has occurred:

$$\begin{aligned} \text{Childcare|Parents}_{\text{trust}} = \{ & \textit{feed child, change diapers, put child down for nap,} \\ & \textit{promote general well - being} \} \end{aligned}$$

$$\begin{aligned} \text{Babysitter}(\text{Childcare|Parents})_{\text{per}} = \{ & \textit{feed child, change diapers, put child down for nap,} \\ & \textit{promote general well - being} \} \end{aligned}$$

$$\begin{aligned} \text{Babysitter}(\text{Childcare|Parents})_{\text{res}} = \{ & \textit{feed child, change diapers, put child down for nap,} \\ & \textit{age appropriate play, bedroom upgrades} \} \end{aligned}$$

It is less clear how trust was broken in this example. The babysitter did, after all, understand that they were being entrusted with promoting the child’s general well-being and did, indeed, promote the child’s well-being (assuming that purple nurseries do promote cognitive development). However, the babysitter’s actions with respect to promoting the child’s general well-being were still incompatible, but incompatible with what? The response was compatible with every expectation in the trust domain. Thus, the mismatch is not between the response domain and the trust domain. Instead it is between the response domain and some other expectation that was not previously apparent. Indeed, the parents had probably never even considered whether their trust domain should include or exclude renovations. Yet, once faced with a purple bedroom, it was clear to them that this was a violation of an expectation implicit in the trust domain but perhaps not recognized by either party. I call these *hidden expectations* because neither party is aware of them when the trust relationship begins. Thus, despite doing his best to be trustworthy, the babysitter still broke the parent’s trust because he violated a hidden expectation. When a response domain contains more fine-grained actions than the expectations in the trust domain, trusters can find themselves

faced with unexpected responses their trust. In this case, the response was incompatible with hidden expectations. I call such problems *indirect misunderstanding mismatches*.

However, not all unexpected behavior is incompatible with hidden expectations. If the babysitter washes all the dirty dishes, this may not have been expected, but the parents may find it acceptable as it does not violate any hidden expectations. The actions may actually set a precedent for expectations of future babysitters. Baier argues that

When we are trusted, we are relied upon to realize *what* it is for whose care we have some discretionary responsibility, and normal people can pick up the cues that indicate the limits of what is entrusted. (Baier, 1986, 236)

This is similar to Jones’ statement that competency includes an “understanding of the norms for interaction”. If Baier and Jones are correct, normal people should recognize that washing the dishes, while not explicitly within the trust domain, is not incompatible with its boundaries or with any hidden expectations, while re-decorating the nursery is obviously incompatible. However, picking up these cues is not always straightforward.

It is difficult, when trusted, to know exactly where the boundaries of the trust domain lie or what expectations may be hidden therein. Trusters cannot account for all possible scenarios that the people they trust may face and thus cannot make explicit all the relevant expectations. It is quite reasonable, for example, for parents to have relatively general expectations of babysitters, like “promote my child’s well-being”. This provides the requisite leeway for babysitters to handle unexpected situations. Trustees thus need to exercise discretion when determining how to carry out such coarse-grained expectations. This leaves trusters vulnerable to trustees’ discretionary powers (Baier, 1986; Mackenzie et al., 2014; Jones, 2019). Conversely, trustees cannot account for all possible expectations, hidden or not, that might be placed on them. This puts the trusted person in an awkward position. They are trusted with respect to a certain trust domain, but may not have a clear idea where the boundaries of that domain lie or what commitments may violate a hidden expectation, leading to a direct or indirect misunderstanding mismatch. One person may appreciate hav-

ing their dishes washed. Another may find it offensively patronizing to have someone else clean up their mess without being asked to do so.

If a trustee wishes to be found trustworthy, it is important that they navigate well the boundaries of others' trust domains when exercising their discretionary powers. That is, it is important that they try to align their response domain with others' trust domains, unless inappropriate to do so. Similarly, trusters who do not wish to be disappointed must fix the boundaries of their trust such that trustees can align their response domain accordingly. Both tasks, however, are challenging. Thus, when trust is broken, it may not be obvious who is at fault or what must change in order to restore broken trust. The process of keeping trust and restoring broken trust requires that both parties negotiate the boundaries of their respective domains in order to avoid future mismatches. There are three features of domains that take center stage in these negotiations: the scope of a domain, the ordering of a domain, and the rigidity of a domain. I will focus here on trust domains, rather than perceived domains or response domains. However, perceived and response domains also have these features.

3 Features of Domains

3.1 Domain Scope

Domains of trust vary in scope. As expectations move from general to specific, the trust domain narrows as more fine grained expectations are added. For example, I might go to the hairdresser and ask them to cut my hair so that it is shorter:

$$\textit{Hairdresser} | A_{\textit{trust}} = \{\textit{cut hair shorter}\}$$

This is a fairly *wide* trust domain. It is wide in the sense that there are a wide range of actions that a hairdresser can do that will be compatible with the trust domain. They could cut my hair into a bob, give me a pixie cut or shave my hair off completely. Any of these

actions would do.

There are two different ways that a domain can narrow. One way is that an expectation can become more fine-grained. This is where the idea of nested sets becomes important. Suppose that I do want shorter hair, but I bring in a picture expecting to leave looking exactly like that picture. The resulting domain might look like the following:

$$\text{Hairdresser}|A_{\text{trust}} = \{\{\textit{trim hair into bob}, \textit{add highlights}, \textit{add lowlights}, \dots\}\}$$

In this case, the domain narrows because the expectation “cut hair shorter” has been made more fine-grained and thus the set of actions compatible with the domain is more narrow. Domains can also narrow by adding additional expectations to the domain. Suppose, for example, that I trust hairdressers to not only cut my hair shorter, but to provide a certain kind of experience while doing so:

$$\begin{aligned} \text{Hairdresser}|A_{\text{trust}} = \{ & \{\textit{cut hair shorter}\}, \{\textit{give advice on haircare}\}, \\ & \{\textit{provide positive conversation}\} \dots \} \end{aligned}$$

In this case, the additional expectations unrelated to the actual haircut restrict the range of compatible actions.

Other things being equal, trust is more easily broken when the domain of trust is narrow because there are fewer compatible courses of action. Whenever we trust another, we risk the chance that they will violate our trust. However, we are not only vulnerable to the actions of those we trust, we are also vulnerable to our own expectations. If someone’s expectations are too specific, i.e. their trust domain is too narrow, it may be impossible to satisfy them. When this happens, individuals find themselves continuously disappointed in others and eventually isolated as successive trustees fail to act as expected. Even the best hairstylist may be unable to meet my expectations if the hairstyle I desire is not suited to my hair type. Thus, trust domains ought not be so narrow that it becomes impossible for

others to act within their boundaries.⁴

Yet, trust domains ought not be so wide that all actions are acceptable. When someone has very few expectations or very coarse-grained expectations, trustees may easily take advantage of them. If I don't have precise expectations regarding some domain, I am vulnerable to other people meeting those expectations in ways that I may not understand or approve of (Jones, 2015). In such cases, my expectations have been met and yet I have still been harmed. For example, someone seeking help for chronic pain may simply expect a doctor to help ease their discomfort. However, an unscrupulous physician might prescribe them a highly addictive painkiller, thus meeting the patient's expectations of decreased pain, but, in doing so, also risking their long-term well-being.

In many cases, as a person gains familiarity with a given type of trust relationship, the more balanced their associated trust domain becomes. With respect to medical contexts, Wendy Rogers (2014) notes

...most patients have little understanding of the etiology, pathophysiology, or prognosis of the illnesses from which they suffer, nor do they know which treatments to consider. If patients had this information, they would remain vulnerable to the vicissitudes of embodied existence but would be less vulnerable overall as they would, to a greater or lesser extent, be able to look after their own health interests. (Rogers, 2014, pg.72)

If the patient in the previous example had no medical training, they were doubly vulnerable. Without a clear understanding of what is medically possible, the patient is vulnerable to setting expectations that no doctor is able to meet. Additionally, without a clear understanding of what is medically appropriate, the patient is vulnerable to others meeting their expectations in harmful ways. However, if the patient gains familiarity with their medical

⁴Catherine Elgin describes an analogous phenomenon regarding knowledge. She argues that a belief is "shaky" if a belief's truth conditions and its relation to them must be almost precisely as they are for the belief to be justified (Elgin, 2008). Any small perturbation undermines such beliefs. Similarly, narrow trust domains require such precise courses of action that the trust is easily undermined if a trustee deviates even slightly.

condition and associated treatment options, the patient is in a better position to determine the appropriate boundaries of the trust domain through negotiating between what is desired and what is possible.

Even with familiarity, however, individuals can disagree about the proper scope of trust. Suppose Ann and Dylan are managers working for the same company. Ann has a wide trust domain regarding the employees, expecting employees to get work done, but trusting them to use their discretion in how and when that work happens. Dylan has a relatively narrow trust domain, expecting employees to follow company rules, procedures and norms. In a disciplinary meeting regarding an employee's repeated late arrival to work, Dylan deems the employee untrustworthy. Ann, however disagrees and argues that the problem does not lie with the employee, but with Dylan's unreasonable expectations. Dylan, however, responds that it is irresponsible not to expect employees to arrive on time. In this situation, Ann and Dylan are faced with whether or not to renegotiate the boundaries of their respective domains of trust regarding employees. Whether and how they renegotiate these boundaries determines how they view the trustworthiness of the tardy employee. If Ann realizes that her trust domain is too wide, she may conclude that the employee demonstrated untrustworthiness through taking advantage of her. Alternatively, Dylan may realize his trust domain is too narrow and conclude that the employee is worthy of trust after all.

3.2 Ordering Within A Domain

Trust domains are not just sets of expected behaviors or attitudes. These sets are ordered according to importance. Some types of trust relationship, such as that between a dentist and patient, have obviously important expected behaviors which cannot be reasonably removed from the domain. If A's dentist cannot perform routine dental procedures, then it would be reasonable for A to end the trust relationship. When trust relationships involve professional standards, norms, and systems of accountability, as in the case of dentistry, trustees are provided with guidance on what they ought to expect. Trustees can also simply expect their

dentist to hold the legally required certifications and thus offload the work of determining what is important onto the relevant governing bodies.

Other domains, however, may not have an obvious ordering of expected behaviors. Individuals' trust domains with respect to friendship, for example, often vary widely. A might place greater importance on community ties than on honesty. Thus, B might be dishonest, but if he is deeply embedded in A's community, she may be willing to renegotiate her trust domain to exclude honesty rather than end the trust relationship altogether. If someone else, C, places a very high value on honesty, they may end the friendship even if B is part of their inner circle. To recall Hawley's example, one friend may place a high value on vacationing together, while another friend may not. Depending on how important that expectation is, the friendship may end if the vacationing friend is met with refusal.

Sometimes, the importance of a given expectation isn't clear until that expectation is made salient. The parents in the babysitting case, for example, had not considered whether refraining from redecorating was an important characteristic of babysitters until faced with a babysitter who did not refrain from redecorating. In college, I played a game called "Deal Breaker". In the game, players were asked whether they would break up with their significant other if that other had some unpleasant feature. Most of the time, the unpleasant feature was funny, but harmless. For example, would you break up with your partner if they walked everywhere backwards? Would you break up with your partner if they preceded every sentence with a nasally giggle? In retrospect, the game was kind of weird. However, it points to an important phenomenon. When thinking about what to expect in a partner, I am guessing that most people wouldn't think about whether to include walking backwards or constant nasally giggles. However, after these are brought to our attention, would we be willing to include these features in our expectations? Answers to these questions reveal how important the presence or absence of these features are to the domain of "partner". This is part of the process by which "hidden expectations" are revealed.

Deal breakers, then, are expected behaviors or attitudes that we would be willing to

end a trust relationship over rather than alter our domain of trust to accommodate them. Regarding dentistry, the inability to diagnose or treat dental problems is a deal breaker for me. When people order their expectations differently, they may have different deal breakers for the same domain. For example, the inability of a hairdresser to provide pleasant conversation while cutting my hair is not a deal breaker for me, but it might be for someone else. Similarly, dishonesty may be a dealbreaker with respect to friendship for one person, but not another.

We can disagree, however, over whether a particular expectation should be a deal breaker with respect to a certain kind of trust relationship. Should we trust a dishonest friend? Should we trust dentists who cannot treat dental problems? Should we trust a grumpy hairdresser? Disagreement regarding deal breakers complicates the process of assessing the trustworthiness of another and attempting to be trustworthy. When trust is broken, it may be due to the presence of unfair or morally problematic deal breakers in the trust domain. An employer, for example, should not have as a deal breaker the expectation that their employees will be readily available twenty-four hours a day. This is an exploitative expectation and thus it is unfair to count employees who fail to meet it as untrustworthy. However, other times when trust is broken, it is because the trustee failed to meet a morally important expectation that ought to be a deal breaker. The expectation of not being physically harmed in an intimate relationship, for example, should be a deal breaker. The abusive partner should not be viewed as trustworthy.

When it is unclear whether a given expectation should be a deal breaker, trustees may dispute accusations of untrustworthiness. In the example of the lying friend, B may argue that his dishonesty wasn't *that* big a deal and C has trusted poorly in holding honesty in such high regard. The issue here is not whether B is capable of being a trustworthy friend, but whether being a trustworthy friend should require honesty. If it does, then B is counted untrustworthy, otherwise he is not.

3.3 Domain Rigidity

In navigating the boundaries of trust domains, then, trusters and trustees must negotiate about the relative importance of different expectations. If they can reach agreement, then they can alter the boundaries of their trust domain, or response domain, accordingly. The willingness to alter a domain, however, is a further variable in negotiations. I call the ease with which domain boundaries change “rigidity”.⁵ The scope of a domain has to do with the existing boundaries of a trust domain, while rigidity has to do with how easily those boundaries can change. It is worth noting that the scope, the presence of deal breakers and rigidity are not independent features. Rather, they are conceptually linked to one another. In particular, the presence of deal breakers is tied to the rigidity of a domain. Rigidity is not an all-or-nothing feature, instead it comes in degrees. To the extent that a domain has deal breakers, that domain is rigid. As more deal breakers are added, the rigidity of the domain increases. If a domain does not have any deal breakers, then it cannot be rigid because there are no expectations that the truster would refuse to alter to maintain trust. Additionally, domains with no deal breakers cannot be narrow in scope. Trust domains narrow in scope as more deal breakers are added.

I have said little about response domains thus far as I have intended to focus on trust domains in this paper. However, in order to discuss further how domain rigidity factors into trust negotiations, I must say a little more about response domains. Recall that a response domain is a set of behaviors or attitudes that a trustee intends to take on in response to a truster. Response domains have similar features as trust domains. They may vary in scope, as trustees have more or less fine-grained ideas about what they will or will not do in a particular trust relationship. They may also vary in the ordering of those commitments or intentions. When push comes to shove, they may choose to prioritize one action over another. They also vary in rigidity in that trustees may be willing to adapt to some expectations, but refuse to adapt to others. In response domains, deal breakers are those behaviors or

⁵I am not using the term “rigidity” in the Kripkean sense of rigid designators.

attitudes that trustees refuse to do, even if that means breaking trust.

Ideally, when there is an incompatibility between a trust and response domain, that is, between what a trustee expects and what a truster does or intends to do in response, the truster and trustee work together to renegotiate new, compatible domains in order to maintain the trust relationship. In doing so, each recognizes the other. They see each other as someone who has needs, wants, knowledge, skills, experience, etc. In seeing the other, they are able to adapt their expectations and responses accordingly. For example, suppose A loves to gently tease her friends, but one day she teases her new friend, B, and hurts their feelings. Ideally, they are able to renegotiate their trust and response domains so that their trust and response domains align. This may look like B removing the expectation that friends will not tease them from their trust domain when they realize that the teasing was intended to build intimacy, not make fun. Alternatively, A may choose to cease teasing B, thus altering her response domain to B, when she realizes that B does not enjoy being teased.

During ideally performed negotiations, trusters and trustees must recognize the other, but need not give equal weight to the others' input. For example, when a patient's expectations do not align with what the doctor knows is medically possible, the doctor's medical knowledge and skill should be given greater weight than the patient's unrealistic hopes in the renegotiation process. However, the doctor should still be open to understanding and incorporating the patient's perspective into the treatment plan and the patient should be open to changing their expectations according to the doctor's recommendations. Recognizing the other, then, requires that both the trust and response domains have the potential to change, that is, they must have some degree of flexibility, but need not have an equal degree of flexibility. Sometimes it is appropriate for a trust domain to have more deal breakers than a response domain, or vice versa.

Suppose someone's trust domain regarding hairdressers looks like the following:

$$\text{Hairdresser}|_{A_{\text{trust}}} = \{\{\textit{cut hair shorter}\}, \{\textit{provide advice on haircare}\}, \\ \{\textit{provide positive conversation}\} \dots\}$$

This domain is relatively narrow, in that A has a good idea of what they want in a hairdresser. They don't just expect a new hairdo, they also expect advice and pleasant conversation. However, suppose that A visits a hairdresser who gives them the best haircut they've ever received, but was rather grouchy while giving it. If positive conversation isn't a deal-breaker, then A would be willing to remove this expectation. Thus, for A, hairdressing is a somewhat *flexible* domain. It isn't totally flexible because A presumably still views receiving a shorter haircut as a deal breaker. If the grumpy hairdresser is also bad at giving haircuts, A would not trust them again in the future.

However, let's imagine that after their haircut, A heads to the dentist because of a toothache with the following trust domain:

$$\text{Dentistry}|_{A_{\text{trust}}} = \{\{\textit{diagnosing dental problems}\}, \{\textit{administering anesthesia}, \\ \textit{drilling out rotten areas, filling area with appropriate substance}\}, \\ \{\textit{general goodwill}\}\}$$

The dentist, however, is a terrible dentist. Indeed, the dentist doesn't know anything about teeth and is mean, to boot. They cannot meet a single one of A's expectations. If A is unwilling to remove any of their expectations, then the domain is a relatively *rigid* domain. Of course, A is wise to have a rigid trust domain with respect to dentistry. If B, who has no deal breakers with respect to dentistry, visits this terrible dentist, he is likely to leave worse than he arrived. Not only will he have received poor dental care, but he also will leave with warped expectations about dentists. He may have arrived expecting the dentist to diagnose

and fix his toothache. In renegotiating the trust domain to align with the dentist's response domain, however, B must abandon the expectation that dentists do this kind of work. He thus leaves with a very weird trust domain with respect to dentistry. Indeed, it isn't clear what, exactly, B now thinks that dentists do.

Recall in section 1, I argued that such weird domains are theoretically possible. One way in which trust domains become weird is when trusters have domains that are too flexible. While it is hopefully unlikely for someone to cease to expect their dentists to diagnose their dental problems, there are other kinds of interactions where weirdness is more likely to seep in. Friendships, sexual relationships, and work relationships, for example, all tend to have a greater degree of flexibility than domains like dentistry, which, as noted earlier, has relatively well-defined and accepted deal breakers. Thus, it is more likely for people to have seemingly weird expectations in these kinds of domains. Sometimes, this weirdness is harmless, and sometimes, indeed, makes these relationships uniquely special. If we held very rigid domains with respect to friendship, that is, if we had too many deal breakers, we would be unable to shift and change our expectations and actions in response to the particular needs, vulnerabilities, and personality quirks of our different friends. Thus, it is often important that trust and response domains have some degree of flexibility, so that trust relationships can both be established and maintained.

While flexibility is often important, too much flexibility can result in a domain that is harmful. An abusive partner, for example, can take advantage of an overly flexible trust domain. As noted earlier, intimate relationships should have some deal breakers, like the expectation that one will not be physically harmed by the other. Other deal breakers arguably include the expectation that one's autonomy will be respected, that one's well-being will be promoted, that one will care and be cared for, etc. Over time, however, an individual may cease to view these as deal breakers in an effort to maintain trust in their partner. Additionally, they may be willing to add new expectations that they would have been unwilling to include previously, such as the expectation that surveillance is normal or that the other

should be solely responsible for financial decisions. Over time, this flexible domain begins to look very different from and incompatible with the original domain.

Such harmful trust (and response) domains often emerge when a trust relationship includes a party who is relatively rigid with respect to some domain in which they have some power over the other. When a trustor or trustee has relatively little power, it may be more preferable to adapt to the other as opposed to ending the trust relationship through either failing to meet expectations (when the trustor is in the position of power) or feeling continuously disappointed or betrayed (when the trustee is in the position of power).

In the case of a rigid trustor and flexible trustee, trustors may include deal breakers that are unfair or exploitative. If the trustee is not in a similar position of power, they are faced with the choice to either lose the trust of the person in power and suffer the consequences or to agree to do things they would not otherwise do. The economically vulnerable employee may agree to doing more than is appropriate in order to keep their job. Their relative lack of power in the employer/employee relationship leaves them particularly vulnerable to exploitation. The employer can include ridiculous expectations in their trust domain, knowing that their employee lacks the power to renegotiate the expectations.

When the trustee is in the position of greater power, they may refuse to adapt their responses to the expectations of trustors. In such cases, trustors must either view the person in power as untrustworthy when they fail to meet their expectations or they must bring their trust domain into alignment with what the trustee's response domain. For example, an individual's care for or dependency on an abusive partner, grants that partner some measure of power, allowing them to avoid fairly renegotiating incompatible trust and response domains. They perhaps know that the individual would rather adapt their expectations to the abuser than view them as untrustworthy. In shielding themselves from a life of feeling continuously betrayed, they adopt a trust domain that includes the expectation that the abuser's harmful behavior is normal.

In addition to the material harms that occur through exploitation, there is an additional

harm that may occur when trust or response domains are totally rigid. When domains are totally rigid, there is no room to consider what others think, want, or are willing to do. This, in itself, often constitutes a harm in that it fails to acknowledge the other as a moral equal. This is not to say that recognizing the other is always morally required when negotiating trust relationships. There are non-ideal cases where there is no obligation to recognize the other. For example, as discussed earlier, not all requests for trust need to be accepted or even acknowledged, as in the case of email scammers. It is morally acceptable to have a totally rigid response domain with respect to email scam artists. In most cases, however, there is an obligation to acknowledge the other as a fellow person, worthy of recognition as a moral equal.

What distinguishes cases where there is an obligation to have a flexible domain (such that we may engage in negotiations) from those cases that do not is an interesting question, but one that lies beyond the scope of this paper. The relevant point to make here is that there is room for disagreement about whether a domain is too rigid or too flexible. This disagreement may itself become the focus of negotiation. Negotiations regarding rigidity focus on how willing we ought to be to change our expectations. This is, as noted, related to negotiations regarding deal breakers, but is slightly different. Negotiations about deal breakers often focus on particular expectations. Negotiations about rigidity often focus on domains more generally. For example, if someone is willing to continue seeing the inept dentist despite their failure to provide adequate care, we might encourage them to develop a more rigid domain of trust with respect to dentistry. We might argue that they have set their standards too low, without necessarily spelling out what expectations, specifically, should be added as deal breakers. Similarly, if a domain is too rigid, we may argue that the truster should be more flexible without specifying which deal breakers are unreasonable.

Indeed, our complaint may not have anything to do with expectations as such, and instead it may simply be that the other's domain is rigid to the degree that we do not feel seen or heard in the trust relationship. All the relevant expectations and responses may be

fine, but the non-rigid party may still feel there is something amiss because they recognize that if there were a problem, they know that the other would not be open to engaging in negotiations to ameliorate that problem. That is, they see that they are not recognized as a moral equal in the trust relationship. The extent to which domains should be flexible to allow such recognition, however, is not clear cut. Thus, one party may claim that the rigidity of their domain is appropriate while the other claims that it is so rigid that they do not feel seen. In such cases, whether the trustee is viewed as untrustworthy depends in part on the rigidity of the trustee's response domain, but also on the truster's trust domain. Both trust and response domains may be inappropriately rigid or inappropriately flexible, such that it may difficult to ascertain who is at fault when trust is broken.

4 Conclusion

In this paper, I have argued that trust breaking is not always due to negligence, incompetence or ill will. Instead, I believe that trust breaking can, and often does, occur despite the genuine efforts of both parties. This can be an especially painful and confusing experience for both the truster and trustee because it isn't always immediately apparent who, if anyone, is to be blamed for the broken trust. Sometimes it may not be important to know who is to blame. Instead, it may be more important to figure out how to avoid future trust-breaking. Identifying the type of trust-breaking that occurs and assessing the associated domains can help clarify who is at fault (if appropriate to determine fault), as well as facilitate negotiations for future trust relationships.

If the trust-breaking was a direct misunderstanding mismatch, we can ask whether trusters communicated their expectations clearly and whether trustees took the appropriate measures to listen well. Did the truster give reasonable guidance as to what they were entrusting? Did trustees seek to understand what was being entrusted to them or did they jump to unjustified conclusions?

If the trust-breaking was a direct conflict mismatch, we must look carefully at the trust domain. In particular, did the trust domain have an appropriate scope and were the deal breakers reasonable? Were the boundaries of their trust domain so narrow that it was impossible for the trustee to act within them? Were the boundaries so wide that the truster was naively susceptible to exploitation? Should they have thought more carefully or done more research about what they ought to expect? Did they place too high a value on something that shouldn't be considered that important? Conversely, trustees must ask themselves whether they should have put forth more effort to care for what was entrusted. Could they have done what was entrusted to them? If they could not because the domain was too narrow, did they do their best to communicate their reservations about this clearly and alert the truster as to what they were willing to commit to? Both parties must ask themselves whether their respective trust and response domains are rigid or flexible such that they are able to appropriately acknowledge the other.

Finally, if the trust-breaking was an indirect misunderstanding mismatch, we may ask whether there were hidden expectations that trusters should have made explicit. Did they do the work of self-searching to know in advance what they expected? Did they communicate this effort clearly to the trustee? Should trustees have known about these hidden expectations even though they weren't made explicit? Should they have asked more questions about the truster's expectations before accepting the trust? In the absence of this greater knowledge of the truster, were their actions reasonable and appropriate?

These questions can help both parties assess whether the broken trust is, in part or in whole, their own fault and thus whether the truster is justified in counting the trustee untrustworthy. Sometimes, however, even after carefully evaluating what happened, who is at fault may still be unclear and both parties left wondering what, if anything, they should have done differently. Sometimes the resulting confusion is enough to end trust relationships entirely. Other times, negotiating about the domains may repair the relationship enough for the truster to trust the trustee again in the future. These negotiations focus on determining

the appropriate scope, ordering of expectations, and rigidity of domains.

References

- Baier, A. (1986). Trust and antitrust. *Ethics*, 96(2):231–260.
- D’Cruz, J. (2018). Trust within limits. *International Journal of Philosophical Studies*, 26(2):240–250.
- D’Cruz, J. (2020). Trust and distrust. In Simon, J., editor, *The Routledge Handbook of Trust and Philosophy*, pages 41–51. Taylor & Francis, New York, NY.
- Domenicucci, J. and Holton, R. (2017). Trust as a two-place relation. In Faulkner, P. and Simpson, T., editors, *The Philosophy of Trust*, pages 149–160. Oxford University Press.
- Dormandy, K. (2020). Exploitative epistemic trust. In Dormandy, K., editor, *Trust in Epistemology*, pages 241–264. Routledge.
- Elgin, C. Z. (2008). Trustworthiness. *Philosophical Papers*, 37(3):371–387.
- Faulkner, P. (2017). The problem of trust. In Faulkner, P. and Simpson, T., editors, *The Philosophy of Trust*, pages 109–128. Oxford University Press.
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, New York.
- Hardin, R. (2002). *Trust and Trustworthiness*. Russell Sage Foundation.
- Hardin, R. (2006). *Trust*. Cambridge: Polity Press.
- Hawley, K. (2014). Trust, distrust, and commitment. *Nous*, 48(1):1–20.
- Hawley, K. (2019). *How to be Trustworthy*. Oxford University Press, Oxford.
- Jones, K. (1996). Trust as an affective attitude. *Ethics*, 107(1).

- Jones, K. (2012). Trustworthiness. *Ethics*, 123(1):61–85.
- Jones, K. (2015). Trust: Philosophical aspects. *International Encyclopedia of the Social & Behavioral Sciences*, 2nd Ed., 24.
- Jones, K. (2017). ‘But I was counting on you’. In Faulkner, P. and Simpson, T., editors, *The Philosophy of Trust*, chapter 6, pages 90–108. Oxford University Press, Oxford.
- Jones, K. (2019). Trust, distrust, and affective looping. *Philosophical Studies*, 176:955–968.
- Kelp, C. and Simion, M. (2022). What is trustworthiness? *Nous*, Forthcoming.
- Mackenzie, C., Rogers, W., and Dodds, S. (2014). Introduction: What is vulnerability and why does it matter for moral theory. In Mackenzie, C., Rogers, W., and Dodds, S., editors, *Vulnerability: New Essays in Ethics and Feminist Philosophy*, pages 1–29. Oxford University Press.
- Medina, J. (2020). Trust and epistemic injustice. In Simon, J., editor, *The Routledge Handbook of Trust and Philosophy*, pages 52–63. Taylor & Francis, New York, NY.
- Nguyen, C. T. (2019). Trust as an unquestioning attitude. *Oxford Studies in Epistemology*.
- Potter, N. N. (2002). *How Can I Be Trusted?* Rowman & Littlefield, Maryland.
- Rogers, W. (2014). Vulnerability and bioethics. In Mackenzie, C., Rogers, W., and Dodds, S., editors, *Vulnerability: New Essays in Ethics and Feminist Philosophy*, chapter 2, pages 60–87. Oxford University Press.
- Scheman, N. (2020). Trust and trustworthiness. In Simon, J., editor, *The Routledge Handbook of Trust and Philosophy*, chapter 2, pages 28–40. Routledge.