



When Monitoring Facilitates Trust

Emma C. Gordon¹

Accepted: 14 March 2022 / Published online: 7 May 2022
© The Author(s) 2022, corrected publication 2022

Abstract

It is often taken for granted that monitoring stands in some kind of tension with trusting (e.g., Hieronymi 2008; Wanderer and Townsend 2013; Nguyen forthcoming; McMyler 2011, Castelfranchi and Falcone 2000; Frey 1993; Dasgupta 1988, Litzky et al. 2006) — especially three-place trust (i.e., A trusts B to X), but sometimes also two-place trust (i.e., A trusts B, see, e.g., Baier 1986). Using a case study involving relationship breakdown, repair, and formation, I will argue there are some ways in which monitoring can be conducive to two-place trust, and to instances of three-place trust that are likely to be repeated over time—especially when previously established two-place trust has broken down. The result, I hope, is not any kind of abandoning of the important idea that monitoring can undermine trust, but an appreciation of where the conflict between monitoring and trust *doesn't* lie – one from which future work will hopefully be better positioned to illuminate where exactly the conflict is.

Keywords Applied Ethics · Monitoring · Psychotherapy · Trust · Romantic Relationships · Couples Counselling

1 Monitoring as Undermining Trust

We refer to trust frequently in navigating our relationships with others, and we are generally invested in whether—or to what extent—it is deserved. Everyday concerns around trust include trying to assess which news sources we can trust for accurate reporting, whether we can trust people in our professional lives to carry out certain tasks, and whether we can trust our loved ones to act in a way that respects commitments to us. And clarifying the nature of trust, what it means to trust well and what it means to be trustworthy are important philo-

✉ Emma C. Gordon
Emma.Gordon@glasgow.ac.uk

¹ University of Glasgow, Room 2.07, 69 Oakfield Avenue, G12 8LP Glasgow, UK

sophical issues for both ethicists¹ as well as social epistemologists working on testimonial knowledge.

A distinction between two forms of trust—two-place trust and three-place trust (Horsburgh 1960)²—will be relevant to the present discussion. Sometimes, we trust a particular person in a general way—we are not thinking about any task they might do for us, but rather about (roughly) our general sense that they will honour commitments they make to us. This is *two-place* trust—i.e., schematically: A trusts B.³ In other cases of trust, we’re trusting a person with *something* in particular—in such cases, A trusts B *to* X. For example, we might trust our colleague to submit their essay marks in time for the course deadline, or we might trust a friend to look after our pet cat while we’re out of town. In the epistemic case, we might trust someone to tell us the truth.

Of course, these two forms of trust often exist in tandem. When people earn our two-place trust, we subsequently rely on them (all things equal, moreso than others) for three-place trust tasks. And, likewise, on the basis of a successful track-record of three-place trust (e.g., suppose a trustee about whom we have no general view has proved trustworthy in carrying out a series of three-place tasks), individuals who don’t already have it can earn our two-place trust.

One live research question asks whether and on what basis we might regard one type of trust (two- or three-place trust) as more *fundamental* than the other, though for the present purposes I’ll remain neutral on this point.⁴ Rather, the focus will be on an assumption in the philosophy of trust that cuts across this distinction, and is so taken to apply to both forms. This is the assumption, to a first approximation, that *monitoring* destroys, or is in some sense incompatible with, trusting.

One early notable articulation of this idea is due to Baier (1986, p.260), who says that “Trust is a fragile plant [...] which may not endure inspection of its roots, even when they were, before inspection, quite healthy.” In other words, even if—say—we have a good, consistent personal or professional relationship with someone, and we have experienced multiple instances in which our trust appeared warranted, we can damage or destroy that trust by taking part in monitoring of that person. Meanwhile, Faulkner (2011, italics mine) holds that “too much *reflection on the trust relation*, perhaps in conjunction with making attempts to minimize risks that trust will be betrayed, can undermine trust.⁵ Broadly similarly, Keren (2014; 2020) observes how reasons to trust “oppose actions like those of carefully monitoring the behavior of the trustee or weighing the available evidence that this person is trustworthy” (see McLeod, 2020, § 4).

¹ Along with foundational discussion of the nature of trust in ethics (Baier 1986; Jones 2004; Alfano (2016); Coleman 1990; D’Cruz 2018), there is also a substantial literature on trust in business ethics, where the value of trust and trust-building is a central concern (e.g., Solomon and Flores 2003; Das and Teng 2004; Ashraf, Bohnet, and Piankov 2006; Berg, Dickhaut, and McCabe 1995).

² See Hardin (1992) and Holton (1994) for further discussion of the distinction between two-place and three-place trust.

³ For a recent overview of work on two- versus three-place trust, see Carter and Simion (2021).

⁴ For some representative articulations of the idea that three-place trust is more fundamental, see, e.g., Holton (1994); Jones (1996); Faulkner (2007); Hieronymi (2008); Hawley (2014). For criticism, see Faulkner (2015) (who takes the attitude of trust to be basic) and Domenicucci and Holton (2017) who take two-place trust to be more fundamental.

⁵ Note that while Faulkner’s characterization above expresses acknowledgement of a tension between knowledge and trust, his use of ‘can’ here leaves it compatible with various ways in which monitoring might be consistent with trusting. As such,

To make this idea more concrete, consider Wanderer and Townsend's (2013) babysitter case, in which a parent watches their babysitter via a remote video feed. A natural reaction to this scenario is to deny that the parents trust the babysitter in this case. And indeed, we might think that a plausible explanation for *why* the parents would watch the babysitter using remote video feed is precisely that they *don't* trust that person. We can easily think of many other similar cases where someone seemingly fails to trust because of acts of monitoring, which can manifest in various ways. Imagine someone who purports to trust their child to stop using the computer at midnight but installs an app on that computer to transmit a log usage times to the parent's phone, for example. Alternatively, picture a tutor who watches their student do a practice test to make sure they don't cheat, or a person who double-checks all of their accountant's calculations, or who follows their child to a party to check whether they are drinking alcohol (after purporting to trust them to not do so).

While it is uncontroversially at least *prima facie* intuitive that monitoring somehow undermines trust, at least across a broad range of cases, what is less clear is why exactly this is so. One explanation, due to Hardin (1992) is that when we trust, we put ourselves in a position to be betrayed, a position we would then fail to occupy while actively monitoring in such a way as to eliminate (or significantly militate against) this vulnerability. As Hardin puts it: "virtually all writers on trust agree that part of trusting is becoming "subject to the risk that the other will abuse the power of discretion" (1992, 507). If, as the thought goes, we take steps to protect ourselves from the potential to be betrayed, we are doing something that undermines trust by eliminating (or significantly reducing) the kind of vulnerability we constitutively subject ourselves to by trusting.

Another kind of explanation is found in the literature on therapeutic trust, or trust that aims at trust-building (Horsburgh 1960).⁶ As the thought goes, when attempting to promote trustworthy behaviour in a trustee (say, a teenager to whom you are lending your car), making yourself vulnerable to the teenager by trusting them with the car helps to increase trustworthiness in the relevant trustee. However, in so far as monitoring then counters such vulnerability, it would seem that monitoring not only undermines trust already present, but also potentially wrecks attempts to build trust initially.

A third such characterisation is due to Castelfranchi and Falcone (2000), who take the idea that monitoring undermines trust to be motivated by a more underlying idea that (i) trust is conceptually opposed to *control*; and that (ii) monitoring essentially involves exercising a form of control.⁷

Fourth, recent work by C. Thi Nguyen (forthcoming) holds that trust is an *unquestioning attitude* that one takes up only if one refrains from 'monitoring, challenging, checking, and questioning'. Putting this all together, what we find in the above literature is (i) a range of theses that capture in different ways an alleged tension between trusting and monitoring;

⁶ In cases of therapeutic trust, which aim principally at trust-building, the trustor needn't have any expectation at all the trustee will prove trustworthy. For example, one's trust of a disloyal person with a small task might be 'therapeutic trust' if the central aim in trusting is not the task at hand, per se, but to lead the individual to undertake commitments that would then make them more trustworthy in the future. Given that what is described as 'therapeutic trust' does *not* involve (as trust is taken to involve) any expectation or belief or even minimal optimism that the trustee will do as entrusted, it is contested whether this kind of attitude constitutes *genuine* trust (for discussion, see Hieronymi 2008). For the present purposes, I am setting aside such cases in order to focus on the relationship between paradigmatic trust and monitoring.

⁷ Note that a slightly different formulation of the idea is found in McLeod (2021), who takes a precondition of monitoring, which is reflecting on the trust relationship, to be in tension with trusting.

and (ii) rationales for *why* trust stands in tension with monitoring which cluster around the idea that monitoring contributes to eliminating something – namely, vulnerability – that seems integral to trusting.

However, a complexity of this debate is that there are really two kinds of ‘tension theses’ that get run together under the slogan that monitoring can ‘undermine’ trust. One thesis, suggested by remarks like Nguyen’s, Baier’s, Wanderer and Townsend’s and Castelfranchi and Falcone (albeit not by weaker characterisations, such as Faulkner’s), holds the monitoring/trusting relationship to be one of *constitutive tension*: such that one is *by monitoring* thereby not trusting.⁸ Alternatively, part of what it is to trust is to refrain from some level of monitoring.⁹ Call this the *strong tension thesis*:

STRONG TENSION THESIS: Trusting and monitoring are in constitutive tension with one another such that, when a trustor A monitors a trustee B with respect to task X (at time T), A *thereby* is not trusting B with X at T.

A slightly weaker tension thesis makes no commitment when it comes to constitutive tension but upholds a just a kind of negative contribution, or ‘erosive effect’ monitoring has on trusting. This more moderate view is suggested by the kind of characterisations (e.g., Hardin 1992; Horsburgh 1960; see also McLeod 2021) of ‘monitoring/trusting’ tension that indicate the latter’s effect on the former is *at least* an undermining or erosive one, even if not one of constitutive tension. Call this the *moderate tension thesis*:

MODERATE TENSION THESIS: Trusting and monitoring are in tension with one another in the sense that monitoring has an *erosive* effect on actual and possible trust relationships.

Note that Hardin’s, Horsburgh’s and McLeod’s contention that trusting someone with something, X, essentially involves subjecting oneself to vulnerability with respect to S’s taking care of X as entrusted, offers at least *prima facie* support for the moderate tension thesis, given that monitoring aims at limiting trust-relevant vulnerability. Likewise, views such as Nguyen’s and Castelfranchi and Falcone’s are also committed to the moderate tension thesis in virtue of holding views on which monitoring leads one to enter in to certain kinds of states (a state of control over the trustee, for Castelfranchi and Falcone’s, and one of questioning, for Nguyen), the persistence of which prevents the development (or persistence) of trust.

In this paper, I will focus – given the kind of examples I will be centrally interested in – on the moderate tension thesis as my critical target, though by seeing just *why* the thesis (despite its *prima facie* plausibility) should not be taken as a given will offer us a useful vantage point (which I explore at the end of the paper) for calling into doubt the strong (constitutive) tension thesis. The result will hopefully be a more nuanced picture of the relationship between monitoring and trust—one that accepts their tension in some cases while making room for not only their lack of tension in other cases, but for monitoring to at least in certain circumstance positively facilitate trusting relationships.

Here is the plan for what follows. In § 2, I begin by challenging the moderate tension thesis by considering a simple overdetermination case where monitoring seems to have *no effect whatsoever* on trusting (either in two- or three place cases). The existence of such cases motivates us to think more critically about received wisdom, as it shows at least one

⁸ Specific versions of this strong thesis are found in the literature on trust in business ethics, where authors such as Frey (1993) and Litzky et al. (2006) maintain that monitoring ‘destroys’ trust.

⁹ On one expression of this idea, due to Dasgupta (1988), trusting is constitutively incompatible with monitoring beyond some threshold.

simple structure where monitoring has no erosive effect. §§ 3–4 then goes a step further and show how monitoring might not only *not* erode trust, but that it *can positively help to facilitate it*. § 3 focuses specifically on the monitoring/trusting relationship in the case of two-place trust, by suggesting how in certain circumstances monitoring not only fails to erode two-place trust that is already in place, but further, that it can play an indispensable role the development of trust or the recovery of broken (two-place) trust. § 4 then considers the monitoring/trusting relationship in the three-place case, with a focus on cases of three-place trust that are likely to be repeated over time, and it is shown how in such cases the relationship between trust and monitoring can be a positive rather than erosive one. Finally, § 5 will take stock of key results and, in light of these, critically revisit the initial monitoring/trusting exclusion dogma we began with.

2 When Monitoring is Irrelevant to Trust: Overdetermination Cases

A lesson from the moral responsibility literature is that your conduct's having a certain normative standing needn't depend on your being able to do otherwise. This was a key lesson from Frankfurt (1969): if you *can't help* but to give money to the homeless, because a demon waiting in the wings will intervene and cause you to do it if you don't choose to do it yourself, you might *still be* morally praiseworthy for donating when you choose to do so, and the demon doesn't intervene. In slogan form, you can get a kind of normative credit for doing something even in a circumstance where your acting the way you did is *overdetermined*.

A similar kind of insight – even though the analogy here isn't perfect – applies in the case of monitoring. Suppose that your monitoring someone is practically unavoidable from the outset: suppose, at time T_1 , you trust a very reliable family member F to do something X, which would take F until time T_3 to complete, and – by coincidence – at time T_2 you are in a position where observing F is practically unavoidable.

Now, consider two version of what might take place at T_2 . In the first version of the case, you become suspicious of F (perhaps irrationally so), and these suspicions are manifest in your (albeit, practically unavoidable) observation of F (vis-a-vis X) at T_2 . In the second version of the case, suppose you have no such suspicions, and despite unavoidably observing F (vis-a-vis X) at T_2 .

A strict reading of the moderate tension thesis implies that trust will be eroded in both versions of the described case. However, intuitively, this is not the case; even if monitoring at T_2 has an erosive effect on your trusting F with X in the first version of the case where your monitoring manifests your intention to limit vulnerability, your monitoring so seems to have no such effect at all in the second version of the case, where your monitoring F is no more normatively significant (with respect to the trusting relationship with F vis-à-vis X) than is compelled conduct that is against one's will (viz., as in Frankfurt cases in which one is compelled to act otherwise than one intends). By contrast, your monitoring F in version 1 is normatively significant and plausibly erosive even though that monitoring is overdetermined no less than it is in Version 3.

Two lessons from this case are as follows: the first version of the case offers at least some prima facie support for the moderate tension thesis – after all, it does seem that when your monitoring manifests your intention to mitigate risk of betrayal – that this has an erosive

effect on the trust initially placed. However, the second version of the case seems as though it should lead us to reject the moderate tension thesis *insofar* as we should reject that monitoring the trustee will always have such an erosive effect.

Taken together, these observations suggest we should reject the initial formulation of the moderate tension thesis in favour of a more carefully formulated version of the thesis, as follows:

MODERATE TENSION THESIS (REVISED (MTT-R)): Trusting and monitoring are in tension with one another in the sense that monitoring *that manifests the trustor's intention to mitigate risk* has an erosive effect on actual and possible trust relationships.

This version of the moderate tension thesis neatly diagnoses both versions of our example case. With respect to MTT-R, your monitoring F has an erosive effect on the trust placed in F to X in the first version of the case, but it doesn't in the second (where the monitoring seems normatively irrelevant), which looks like just the right result.

Even so, in what follows, I want to suggest that even MTT-R is too strong, and this is because there are cases (indeed, not fanciful cases, but cases with important practical relevance) that are such that (i) the trustor's monitoring the trustee manifests her intention to mitigate risk; *and* (ii) even so, the monitoring not only *fails to erode* trust but is *positively conducive* to trusting relationships.

§§ 3–4 sets out to establish this point through an examination of relevant empirical literature on the role of monitoring in trust relationships in couples counselling and psychotherapy (as applicable in both cases of (§ 3) two- and (§ 4) three-place trust). Perhaps ironically, the very kinds of close interpersonal cases where monitoring is often taken to erode trust are the very kinds of cases where monitoring can be essential to *repairing* such trust when it has been fractured. § 5 then assesses these results and considers what they might also suggest about the both the moderate tension thesis that has been our working target, but also the strong tension thesis.

3 When Monitoring is Conducive to Two-Place Trust

It is intuitively plausible and empirically established that infidelity has a negative impact on trust (Bird, Butler, and Fife 2007; Vossler 2016)¹⁰. Consider a standard case in which such infidelity might come to light:

BETRAYAL: Sam and Alice are a monogamous couple who have had reciprocal two-place trust for ten years. Now, suppose that one evening Alice receives a surprising message on her phone from Sam—a message clearly meant for someone else, which refers to arrangements to spend the night together. Sam accidentally sent this message to Alice, and now Alice has evidence of Sam's infidelity. When confronted, Sam admits to having been in a relationship with a work colleague for several months. The ensuing arguments and difficult conversations confront Sam with the reality of losing Alice—while both partners have had some unmet needs, the relationship has been loving and happy in many ways, and Sam wants to fight for its survival. Alice agrees to give the relationship a chance to recover,

¹⁰ Here, "infidelity" refers to a violation of a monogamous couple's agreement to maintain sexual exclusivity. It is not intended to imply that such exclusivity is the only (or even best) way to conduct a healthy relationship.

provided Sam ends the affair and limits contact with the other person to purely professional exchanges.

This is a case in which Alice's two-place trust in Sam is fractured and will likely need to be rebuilt if the relationship is to succeed. This is sometimes called the *restitution phase* of moving past an affair (e.g., Snyder et al., 2007). Once we look more closely at *how* that rebuilding might take place, we can begin to see how monitoring not only doesn't always *erode* whatever limited trust remains post-affair, but actually may in some cases be positively useful in restoring it.

As Glass (2007) puts it, if the relationship is to survive the kind of betrayal that takes place in the case of an affair, then – in the restitution phase – “the antidote is openness, accountability, and honesty. When a partner has been dishonest and deceptive, the only reality that can be trusted is concrete evidence that the affair is over” (2007: 325). One way to acquire that concrete evidence, at least to begin with, is by *monitoring*.¹¹

Marriage and family therapy professors Hertlein et al. (2017) explore the types of surveillance and checking—i.e., monitoring—that often take place in couple relationships after the discovery or disclosure of an affair. As they note (*italics mine*), “A main tenet of infidelity treatment is the reestablishment of trust in the relationship. Part of how trust may be established again is by *demonstrating that the individual participating in the infidelity has ceased such activity*, as well as ceasing any contact with the third party.” Note that the relevant kind of demonstration here is not meant to be mere *testimonial* demonstration (as might be the case when the perpetrator of the affair ‘promises’ or ‘guarantees’ that they have ceased such activity). The breakdown of trust, after all, undercuts the evidential weight of any such promise or guarantee. Rather, the idea is that essential to re-establishing trust post-affair is *demonstration* – e.g., evidence that can be appreciated non-testimonially by the betrayed party, such as by perceptual evidence, of the sort one acquires through monitoring but not through merely receiving someone's word.

Accordingly, relationship therapists typically support some forms of monitoring when an affair is discovered; Snyder et al. (2007, 304) encourage “eliminating secrecy” in an effort to regain trust, but also “taking small gradual risks and tolerating initial discomfort.” And when discussing rebuilding trust, they suggest that worries might “occur less frequently, be less intense, or be put to rest more easily after seeking information” (p. 305). Meanwhile, the betraying partner is urged to “work at being completely open” (p. 306) and understanding the importance of the other party's receiving observational evidence to start out.

What might such monitoring look like in practice? Some examples we find in the psychotherapy literature (e.g., Glass 2007) include free access to the partner's phone and computer, checking in on video from various places, keeping evidence of time spent doing what one has claimed, and potentially attending more events together than would otherwise be the norm. In one case study she discusses, the unfaithful partner invited his partner to sit beside him and read a book when he was on the computer. He also made himself “totally accessible” for a term, holding onto receipts, hotel bills and such. On these sorts of cases, Glass comments: “[Only] being trustworthy can heal the rupture. Willingness to be accountable is essential. In everyday terms that means that unfaithful partners need to answer questions about where they are going, what they are doing, and with whom. Without accountability, there's *no reason to believe their word that the affair is over.*” (p. 325, my italics).

¹¹ Although the kind of case study I am using here primarily involves psychotherapy, see also Walker (2006) for related discussion on *moral* repair after moral wrongdoing.

Of course, this can be challenging for the monitored partner. When exploring some salient cases, Hertlein et al. describe how the couple therapist “normalizes the need to surveil and look for evidence, explaining that trust is built on the establishment of reliability and on the partner’s ability to be predictable” (p. 329). They suggest that where an unfaithful partner reacts negatively to being monitored in the early days of recovery from an affair (i.e., the restitution phase), the therapist should help that person to empathise with the betrayed partner’s experience¹². However, the couple’s therapist will also typically normalize the betraying partner’s *dislike* of this essential surveillance. On this point, Hertlein et al. write that they worked to help this person understand why this surveillance was desired and may even be helpful, so that the betrayed partner “did not have her belief that [she will be pushed way] away confirmed” (p. 330).

Moreover, Hertlein et al. suggest that trust is facilitated “when people act in ways that are predictable. In this way, they become reliable and trustworthy” (p. 319). In essence, then, therapy will often focus on how the betrayed partner might come to repeatedly and consistently acquire evidence that the infidelity has come to an end. At the same time, however, there is a therapeutic aim to address these surveillance activities as time goes on because otherwise it can “[stop] the couple from moving toward their goals and will put them back at square one” (p. 328). Glass (2007, 325) calls the gradual reduction of monitoring “loosening the cord” or “lengthening the string”; to avoid frustration and respect autonomy, it should be the case that accountability, honesty and consistency results in steady reductions in monitoring. Hertlein et al.’s (2017) example cases involved the therapist encouraging an open conversation about “how long [the couple] would be able to tolerate the surveillance” (p. 330), as well as exploring other strategies to improve the relationship (e.g., working on improving communication and spending more quality time together).

In addition, Warach and Josephs (2021, 76) point out that *without* such loosening of the cord, “betrayed partners may become obsessed with the details of the extra-dyadic involvement and may relentlessly question the unfaithful partner about them.” Hertlein et al. agree, noting the importance of a time limit on this monitoring—“While we want to normalize and develop a time to build trust, we do not want to set up a scenario where the couple permanently lives out the dynamic where one monitors the other. [...] Therapists should set an appropriate time limit for the surveillance” (p. 329).

The takeaway from the above literature on monitoring’s role in the restitution phase following fractured trust, accordingly, seems to be twofold. First, monitoring plays an indispensable role in not merely refraining from further eroding, but in *positively regaining* betrayed trust in a close relationship. Moreover, it is not as though monitoring is *merely* usually present as an expression of the betrayed party’s distrust. Rather, the monitoring *itself* does important work in reestablishing trust. The second main takeaway is that monitoring’s role, even if critical in the restitution phase, also must be managed very carefully, and it should be managed with a kind of ‘time limit’ in mind in order to be most effective. What these takeaways suggest, then, is that – at least in the case of a particular kind of two-place trust that is familiar and widespread – MTT-R should be rejected. When *betrayal* damages trust, monitoring can very plausibly be what is necessary to help regain it over time; in this overlooked respect, monitoring facilitates rather than hinders trust – and *even when* that

¹² See for example Fife et al. (2013, p. 353–355) for the pivotal role that empathy plays in helping the unfaithful partner to engage non-defensively in therapy about the affair.

monitoring itself manifests the trustor's aim at mitigating risk, which is plausibly among the aims of betrayed partner.

I want to now consider two objections the significance of the above diagnoses of these results. One straightforward, though misguided, objection proceeds as follows: "But even if we accept this empirical literature in psychotherapy supports the role of monitoring in regaining trust, it remains that MTT-R seems to get the right result in some (e.g., like Townsend and Wanderer's Nannycam case). After all, when trust has *not* broken down, if you surveil your partner your doing so erodes trust. Thus, the constitutive incompatibility claim stands."

The problem with the above reasoning is that it misconstrues the kind of claim that the empirical literature discussed is useful in pressing back against. MTT-R, recall, represented one restricted way to capture the commonly expressed idea that monitoring is somehow in tension with trusting. And MTT-R represented that tension as one on which the former always has an erosive effect on trusting relationships (even if not being constitutively incompatible with trusting – as the strong tension thesis maintains), at least when the monitoring manifests the trustor's intention to minimize risk of betrayal. We've seen now that *even this* is too strong, given that *some cases* of monitoring (which manifest that intention) straightforwardly facilitate trust rather than erode whatever trust remains after betrayal. However, establishing that MTT-R is too strong (given that it fails to hold as an unrestricted claim) is entirely compatible with maintaining that *some cases* of monitoring that manifest one's intention to minimize risk of betrayal do erode trust.

A second objection to the line advanced in this section challenges the assessment of the significance of the empirical literature directly by maintaining that monitoring isn't *sufficient* in BETRAYAL to rebuild trust (e.g., sometimes monitoring won't result in regained trust) and on this basis, to reject the claimed importance of monitoring to rebuilding trust. Here we should agree completely that monitoring – in isolation – is very unlikely to be sufficient for a couple trying to recover from an affair. For example, Abrahamson et al. (2012) suggest other vital ingredients include changing couple dynamics, understanding the meaning of the experience, forgiveness, and support, e.g., from a therapist. However – and this is an important further clarification – the claim here isn't even that isolation *mixed* with the kinds of contributing factors Abrahamson mention, is *always* sufficient for regaining trust in relationships after it is fractured. Rather, the claim is that we have good evidence from the above literature that the.

positive contribution monitoring makes to rebuilding trust is plausibly indispensable to doing so effectively. In sum, the confusion in the above objection is one of necessity and sufficiency; regardless of whether monitoring plays a sufficient role in strengthening fracture trust, the fact that we have good reason to think it is necessary for doing so is itself enough to stand in tension with MTT-R.

I want to conclude this section by noting that the scope of the above point is actually wider than I've suggested so far, in the importance of monitoring in trust-building – while most evident in cases of fractured trust – isn't limited to such cases. That is, there are other relationship-building cases where monitoring (which manifests the intention to minimize risk of betrayal) can facilitate trust, even when neither of the two agents involved in the case have betrayed the others. Imagine this example, then, where a lack of trust impacts on relationship quality in a different way.

PAINFUL HISTORY: Aiguo and Margot have been dating for a couple of months and are optimistic about their future. Perhaps don't yet have trust at the kind of depth you might have in a longer-term relationship, but Aiguo finds it easy to trust Margot—he experiences her as frank, easy to talk to, and consistent in her commitments. However, Margot doesn't have similar two-place trust in Aiguo—not because of any salient facts about Aiguo himself, but because Margot has never had a relationship in which the other party was trustworthy. Her most recent ex-partner disappeared with a large sum of her money after claiming it was a loan, and the partner before that suddenly ended their engagement with no explanation.

In this case, we might also think that Aiguo and Margot could benefit from using some of the same monitoring techniques and agreements present in **BETRAYAL** above, albeit perhaps for less time. However, for present purposes, the salient point is that **PAINFUL HISTORY** is a offers a distinct sort of context (one absent betrayal in the relationship) in which monitoring aimed at minimising potential betrayal-risks can be helpful to a trust relationship rather than harmful. The upshot here is that monitoring can plausibly be useful in facilitating trust not only when one's two-place trust in someone has been damaged, but when one struggles to develop that two-place trust in the first place due to, e.g., factors related to one's previous relationships.¹³

4 When Monitoring is Conducive to Three-Place Trust

Thus far we've been looking at cases that challenge the applicability of MTT-R in connection with two-place trust. Consider now the following line of thought: rather than to give up MTT-R wholesale, we might try to simply vindicate a further restricted version of the thesis that applies only to three-place trust, as follows:

MODERATE TENSION THESIS (REVISED (MTT-R*)): Trusting and monitoring are in tension with one another in the sense that monitoring that manifests the trustor's intention to mitigate risk has an erosive effect on actual and possible *three-place* trust relationships.

Granted, it's less straightforward at least initially to see how monitoring would play anything but an undermining role in cases of three-place trust (remember here the initial nannycam case – a case of three-place trust that seemed obviously wrecked by the parents' monitoring the babysitter via the nannycam). Even so, I think that by simply revisiting the cases from § 3—and one further case—we'll have all we need to make sense of how monitoring is also *sometimes* conducive to three-place trust in a way that is at odds with MTT-R*.

¹³ Although the focus here has been to draw from the psychotherapy literature (rather than merely abstract philosophical thought experiments) in order to highlight at least one strand of empirical evidence that casts doubt on received thinking about monitoring/trusting exclusion, it is worth noting that an analogous kind of argument structure might plausibly hold entirely outwith the context of personal and romantic relationships, and even outside contexts in which there has been any betrayal whatsoever (in the present trust relationship or for that matter in the past). To suggest but one such kind of case – consider a case in a professional context where the stakes are high: **HIGH STAKES:** Francesca has recently joined the CIA after an extensive psychological evaluation, vetting process and series of interviews. There are high stakes attached to her role, and lives could be lost if she doesn't perform well. She is given a set of responsibilities, and at first she is shadowed by an experienced member of the Agency. When it becomes apparent that she is alert, competent and responsible, she is no longer monitored by her experienced colleague. It would seem that in cases where performance (of the kind of activity monitored) has particularly high stakes, even if someone *seems* trustworthy (suppose Francesca's CV is impeccable) the threshold for trusting them might be a lot higher—such that monitoring them is essential to reaching such a higher threshold.

With this in mind, let's first revisit the BETRAYAL case in which Sam and Alice are in the restitution phase of their post-affair relationship. In this situation, it is plausible that for every *individual* case where Alice might extend three-place trust to Sam, if she has engaged in *general monitoring* prior to this case then Alice is more likely to trust in this particular three-place case. To make this idea concrete, just imagine that Sam says "I'm coming straight home as soon as I've finished the day's work" at (1) time t the day after the affair is exposed, and then again later at (2) time $t+1$ when Alice has been monitoring Sam in agreed ways for two months. It is very plausibly going to be more likely, given what the literature indicates in § 3 about two-place trust, that (in this three-place case) Alice (A) will trust Sam (B) to come straight home as soon as the day's work is finished (X) at time $t+2$. Even if we hold fixed that Alice and Sam haven't done *anything* else to work on their relationship, the idea is that three-place trust remains more likely at $t+2$ (even if not as likely as it would be had the couple been doing more to improve their relationship) conditioned on previous general monitoring than otherwise. Monitoring that provides evidence of a person's trustworthiness, then, has an impact on the likelihood of three-place trust in *similar future cases*. In this respect, monitoring (indeed, even which manifests an intention to minimise risk of betrayal) can facilitate three-place trust by establishing conditions that make it more likely, and regardless of whether further monitoring would undermine three-place trust once placed.

Secondly, if we turn to the PAINFUL HISTORY CASE—where Aiguo is deserving of two-place trust but Margot has a difficult time offering it because of untrustworthy ex-partners—we can see something similar happening. In order to generate a three-place case in this context, suppose Aiguo says "I will pay for our trip this weekend." Margot might find it difficult to trust Aiguo to do so at time t prior to have ever engaged in trust-building monitoring of Aiguo—after all (on this particular vignette) past partners have broken their commitments to her many times, and claims to use money responsibly are a particular trigger for Margot's suspicion given that her last partner stole from her. Now, fast forward to Aiguo saying the same thing at $t+1$, when Margot has had time to accumulate evidence—through monitoring—that she can trust him to do such a thing.

Again, here, it looks as though monitoring at one time can facilitate those conditions – and might even be essential to establishing them – *later* whereby one is more inclined to extend three-place trust than otherwise—not just in similar future cases with someone who has betrayed someone (i.e., as in BETRAYAL) but in future cases with someone with no such history.

5 The Strong Tension Thesis

Although the relationship-based cases considered so far most squarely challenge version of the moderate tension thesis, it's worth now considering how the challenge might extend over to the strong tension thesis, which we initially characterised as:

STRONG TENSION THESIS (STT): Trusting and monitoring are in constitutive tension with one another such that, when a trustor A monitors a trustee B with respect to task X (at time T), A *thereby* is not trusting B with X at T.

In order to safeguard STT against ‘overdetermination’ cases (of the sort discussed in § 2), let’s make a parallel adjustment to STT which reflects the analogous move from MTT to MTT-R:

STRONG TENSION THESIS (REVISED (STT-R)): Trusting and monitoring are in constitutive tension with one another such that, when a trustor A monitors a trustee B with respect to task X (at time T), *and does so in a way that manifests A’s intention to mitigate risk*, A thereby is not trusting B with X at T.

A first point to note is that cases where monitoring facilitates trust don’t obviously undermine STT-R even when they undermine versions of the moderate tension thesis. After all, monitoring can facilitate trust (in a way incompatible with MTT-R) without being *compresent* with trust.¹⁴ This might, for instance, be the natural way to think of what is going on in cases like BETRAYAL. For example, on one interpretation of what is going on in that case, we have a sequence where (i) Sam’s betrayal entirely undermines Alice’s trust; (ii) Alice monitors Sam (in such a way as to be effective in facilitating the rebuilding of trusting; (iii) Alice begins to refrain from monitoring; and then (iv) Alice’s trusting Sam is regained, albeit, not *compresently* with the monitoring that facilitated its reinstatement.

The above steps (i)-(iv) constitute an interpretation of BETRYAL that count against MTT-R but *not* against STT-R, given that nowhere in the unfolding of (i-iv) is Alice’s trusting and her monitoring *compresent*.

However, the above sequence (i-iv) is but one way to think about how things unfold. Here is an equally if not more plausible alternative narrative: (i*) Sam’s betrayal significantly undermines but does not entirely eradicate Alice’s trust; (ii*) Alice, partly on account of retaining some degree of trust in Sam, monitors Sam rather than severing their relationship, and in the hope that such monitoring will be effective in eventually facilitating the rebuilding of trusting); (iii*) Alice begins over time to refrain from monitoring, by monitoring less often and in a less wide range of circumstances; and then (iv*) upon further refraining from monitoring, Alice’s trusting Sam is strengthened, ultimately to the level prior to the betrayal.¹⁵

The above steps (i*-iv*) count against not only MTT-R, and also (when suitably filled out so as to involve discrete three-place tasks, e.g., Alice’s trusting Sam to come straight home) against STT-R. This is because the transition from (i*-iv*) *does* include monitoring and trusting *compresently*. For example, part of this story will be a case where Alice still trusts Sam to some degree to come straight home (e.g., say, at the rebuilding phase) while monitoring him, indeed, monitoring in a way that manifests an intention to mitigate risk.

What makes the difference, we might ask, between Alice’s monitoring Sam while trusting him to some degree (in this case -- i*-iv*) and cases where one’s monitoring a would-be trustee *really is* in constitutive tension with their trusting them? It can’t be a manifestation in monitoring of an intention to mitigate risk, given that this is assumed in both kinds of cases.

While my limited aim in this section was merely to show that cases that undermine MTT-R also seem to cut some ice against STT-R (as opposed to giving a positive account of what precise features of those cases are distinctive), I will suggest what seems most plausible on this front. Note that when in the course of rebuilding trust with Sam, Alice

¹⁴ Note that a separate issue is whether monitoring can be *compresent* with an absence of distrust. For discussion of this point in the context of political trust, see Lenard (2008).

¹⁵ Thanks to a reviewer at *Ethical Theory and Moral Practice* for suggesting some helpful ways to clarify this alternative narrative.

begins monitoring him, while it is true that her doing so (unlike in, e.g., some overdetermination cases) manifests her intention to mitigate risk, her doing so *also* manifests a further distal intention, namely, to restore trust, that is, to heal a trusting relationship. We don't find this further intention manifest in one's monitoring a trustee in cases (e.g., like the nanycam case) where one's monitoring seems paradigmatically in constitutive tension with their trusting.

6 Taking Stock: Concluding Remarks

The idea that monitoring is in tension with trusting is widely taken for granted in the trust literature, however, often times only in slogan form, and this fact should encourage us to look more closely at the claimed tension with a critical eye.

I've attempted to make some progress on this score. After considering some initial motivation for the idea that monitoring and trusting are in tension with one another, I considered two ways to interpret this claim – what I called the *moderate tension thesis* and the *strong tension thesis*, where the latter submits that the tension is a constitutive one, and the former submits more weakly that the tension, even if not constitutive, is erosive, in the sense that monitoring tends to erode trust.

This more moderate version of the thesis has been the central target of the paper. I first refined this moderate thesis in order to deal with certain kinds of overdetermination cases and it was then shown that even this refined version of the moderate thesis can't be reconciled with the evidence we have about the role monitoring plays in positively establishing or rebuilding broken or otherwise compromised trust, as evidenced from case studies in psychotherapy. The role of monitoring in these cases is, crucially, not merely a *possible* contributor to rebuilding trust, but a critical component of the rebuilding process. Appreciation of this should lead us to part ways with any kind of unrestricted version of a moderate tension thesis.

While challenging the moderate tension thesis was the central aim here, it was shown finally in the last section that some of the psychotherapy cases that should lead us to reject the most plausibly version of the moderate tension thesis (MTT-R*) also put pressure on the strong constitutive thesis.

Ultimately, we saw, even if there are genuinely cases where monitoring has a deleterious effect on trust, it is a mistake to overgeneralise from these paradigmatic cases to general principles about trust's relationship to monitoring that are taken to range over trusting as such. To the extent that monitoring is in tension with trusting, this is in a more limited and nuanced respect than proponents of slogans registering this apparent tension have appreciated.

Declarations

Conflict of Interest/Competing Interests Statement This work was conducted a part of a Leverhulme-funded project (#RPG-2019-302). The author has no other financial or non-financial interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the

article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrahamson I, Hussain R, Khan A, Margot JS (2012) 'What Helps Couples Rebuild Their Relationship after Infidelity?'. *J Fam Issues* 33(11):1494–1519
- Alfano, M. (2016). 'Friendship and the Structure of Trust' in *From Personality to Virtue* (eds. A. Masala and J. Webber). Oxford: Oxford University Press. pp. 186–206
- Ashraf N, Bohnet I, Pivankov N (2006) 'Decomposing Trust and Trustworthiness'. *Exp Econ* 9(3):193–208
- Baier A (1986) 'Trust and Antitrust'. *Ethics* 96(2):231–260
- Berg J, Dickhaut J, McCabe K (1995) 'Trust, Reciprocity, and Social History'. *Games Econ Behav* 10(1):122–142. <https://doi.org/10.1006/game.1995.1027>
- Bird MH, Mark H, Butler, Fife ST (2007) 'The Process of Couple Healing Following Infidelity: A Qualitative Study'. *J Couple Relatsh Therapy* 6(4):1–25
- Carter JA, Simion M (2021) "The Ethics and Epistemology of Trust", *Internet Encyclopedia of Philosophy*
- Castelfranchi C, Falcone R (2000) "Trust and Control: A Dialectic Link", *Applied Artificial Intelligence* 14(8):799–823
- Coleman JS (1990) 'Relations of Trust'. *Foundations of Social Theory, Cambridge, London*, 91–116
- Das TK, Teng B (2004) 'The Risk-Based View of Trust: A Conceptual Framework'. *Journal of Business and Psychology* 19(1):85–116
- Dasgupta P (1988) "Trust as a Commodity", in ed. D. Gambetta, *Trust: Making and Breaking Cooperative Relations*. New York: Basil Blackwell, pp. 49–72
- D'Cruz J (2018) 'Trust within Limits'. *Int J Philosophical Stud* 26(2):240–250
- Domenicucci J, Holton R (2017) 'Trust as a Two-Place Relation' in *The Philosophy of Trust* In: Faulkner P and Simpson T (eds.). Oxford: Oxford University Press. 149–60
- Faulkner P (2007) 'A Genealogy of Trust'. *Episteme: A Journal of Social Epistemology* 4(3):305–321
- Faulkner P (2011) *Knowledge on Trust*. Oxford: Oxford University Press
- Faulkner P (2015) 'The Attitude of Trust Is Basic'. *Analysis* 75 (3): 424–29
- Fife ST, Gerald R, Weeks, Jessica Stellberg-Filbert (2013) 'Facilitating Forgiveness in the Treatment of Infidelity: An Interpersonal Model'. *J Family Therapy* 35(4):343–367
- Frankfurt HG (1969) 'Alternate Possibilities and Moral Responsibility'. *J Philos* 66(23):829–839
- Frey B (1993) 'Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty'. *Economic Inquiry* 31(4):663–670
- Glass S (2007) Not "Just Friends": Rebuilding Trust and Recovering Your Sanity after Infidelity. Simon and Schuster
- Hawley K (2014) 'Trust, Distrust and Commitment'. *Noûs* 48(1):1–20
- Hertlein KM, Dulle C, Cloud R, Leon D, Jenna Chang (2017) 'Does Absence of Evidence Mean Evidence of Absence? Managing the Issue of Partner Surveillance in Infidelity Treatment'. *Sex Relatsh Therapy* 32(3–4):323–333
- Hieronymi P (2008) 'The Reasons of Trust'. *Australasian J Philos* 86(2):213–236
- Holton R (1994) 'Deciding to Trust, Coming to Believe'. *Australasian J Philos* 72(1):63–76
- Horsburgh HJN (1960) 'The Ethics of Trust'. *Philosophical Q* 10(41):343–354
- Jones K (1996) 'Trust as an Affective Attitude'. *Ethics* 107(1):4–25
- Jones K (2004) 'Trust and Terror'. In *Moral Psychology: Feminist Ethics and Social Theory*, edited by Peggy DesAutels and Margaret Urban Walker, 3–18. Lanham, MA: Rowman & Littlefield
- Lenard P (2008) 'Trust Your Compatriots, but Count Your Change: The Roles of Trust Mistrust and Distrust in Democracy', *Political Studies* 56 (2)
- Litzky BE, Kimberly A, Eddleston, Kidder DL (2006) 'The Good, the Bad and the Misguided: How Managers Inadvertently Encourage Deviant Behaviors'. *Academy of management perspectives*, 20(1):91–103
- McMyler B (2011) *Testimony, Trust, and Authority*. OUP USA
- Nguyen CT (forthcoming). "Trust as an Unquestioning Attitude." *Oxford Studies in Epistemology*
- Snyder DK, Baucom D, Gordon KC, Gottman JM, Halford WK (2007) *Getting Past the Affair*. Guilford Press
- Solomon RC, Flores F (2003) *Building Trust: In Business, Politics, Relationships, and Life*. Oxford University Press

- Vossler A (2016) 'Internet Infidelity 10 Years on: A Critical Review of the Literature'. *Family J* 24(4):359–366
- Walker MU (2006) *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. Cambridge: CUP
- Wanderer J, Townsend L (2013) 'Is It Rational to Trust?'. *Philos Compass* 8(1):1–14
- Warach B, and Lawrence Josephs (2021) 'The Aftershocks of Infidelity: A Review of Infidelity-Based Attachment Trauma'. *Sex Relatsh Therapy* 36(1):68–90

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.