

Disappointed Yet Unbetrayed: A New Three-Place Analysis of Trust

Edward Hinchman

Florida State University

Simple observations reveal an obvious difference between betrayed trust and disappointed trust. Say A trusts B to perform some action, ϕ . And say B does ϕ and thereby does not disappoint A's trust. Could B somehow nonetheless betray A's trust? If A's trust in B to ϕ amounts to something more than his merely relying on B to ϕ , then it is easy to see how B might betray A's trust even though she does not disappoint it – since she does, at least, ϕ . Perhaps B ϕ s only because someone – perhaps A himself – coerces her into ϕ ing. Or perhaps B ϕ s with no memory of A's trust and with a firm disposition not to ϕ were she to remember it. In each case, B betrays A's trust in her to ϕ , though B does ϕ and thereby does not disappoint A's trust. Such cases lead many – as we'll see – to assume that trust builds on mere reliance, that the fundamental normative structure of trust is 'A trusts B to ϕ ,' and that the task for defining trust, by contrast with mere reliance, lies in explaining how the necessary condition imposed by this structure falls short of sufficiency – as revealed when trust is undisappointed yet betrayed. I aim to question that assumption – that trust builds on mere reliance – by investigating a more complex case. What if the relation between disappointment and betrayal works the other way round? What if, instead of being undisappointed but betrayed, A's trust is disappointed yet unbetrayed?

To set up the more complex case, let's ask why it seemed that trust builds on mere reliance. We were gripped, it seems, by two more specific assumptions: (i) that there are two conditions in play – one of *fidelity* (B does not omit to do what A trusts her to do), another of

concern (B shows the right attitude toward A) – and (ii) that we can answer our initial question about betrayal by seeing how the condition of concern builds on the condition of fidelity. When B disappoints A's trust in her to ϕ , she thereby betrays that trust, we assumed, because B cannot show appropriate concern for A if she omits to do what A is trusting her to do. I argue that this assumption is false, not because B can show concern without fidelity but because there are contexts in which B cannot show fidelity without concern. On some occasions for trust – I focus on interpersonal trust in a promise and intrapersonal trust in an intention – the condition of fidelity builds on the condition of concern: nothing that fails to show appropriate concern could count as fidelity to the normative understanding that informs the trust.

That possibility explains the more complex possibility that we'll explore: trust disappointed yet unbetrays. Here B does not ϕ , disappointing A's trust in her to ϕ . Is A's trust thereby betrayed? To take the question seriously, we must work from a suitably sharp contrast between the concepts in play of disappointment and betrayal. Toward that end, I stipulate that someone trusted to ϕ who does not ϕ thereby disappoints that trust, even if (perhaps) she does not betray it, and that someone trusted to ϕ who does ϕ does not disappoint that trust, even if she does betray it. This concept of disappointment is not, as such, psychological or moral. It is a concept whose application is governed entirely by the single question: did you perform the action that you were trusted to perform? Given that the concepts are distinct, my question is whether our understanding of betrayed trust builds on this understanding of disappointed trust, in a way that parallels our initial assumption that trust builds on mere reliance. I argue that it does not.

Since there are two contrasts in play – fidelity versus concern, disappointment versus betrayal – we'll naturally wonder how they compare. Note first that the question of fidelity

arises only when trust manifests a normative understanding that responds to an assurance: when you trust B to ϕ though B has not promised to ϕ , there is, strictly speaking, nothing for B to show fidelity to. (I discuss this issue in Section VII.) I argue that one who has promised to ϕ can meet the fidelity condition, not omitting to do what her promisee trusts her to do, even when that trust is disappointed, since she fails to ϕ . It follows that not omitting to do what your promisee trusts you to do is not the same as doing what you have promised to do. Colloquially, what your promisee trusts you to do is to ‘keep’ your promise. I argue that keeping your promise – remaining ‘true’ to it – does not reduce to doing what you’ve promised to do. Promissory trust is disappointed yet unbetraysed when the promisor remains true to her promise despite not doing what she has promised to do. My approach aims to explain how that is possible. I pursue the parallel with intention to help develop that explanation.

We need a new analysis of trust because the possibility of trust disappointed yet unbetraysed undermines the core rationale for the traditional three-place analysis, ‘A trusts B to ϕ .’ I argue that trust most fundamentally has this different normative structure: ‘A ϕ s through trust in B.’ We do, of course, trust people to do things. But its more fundamental normative structure articulates how trust puts us in touch with reasons to do or plan to do other things on the basis of the trusted’s worthiness of that trust – whether to act on trusted advice, to believe on trusted testimony, or to plan through trusting a promise or intention. I call this an Assurance View of trust because it treats trusting as accepting an invitation to trust – in effect, an assurance that the other (perhaps your own earlier self) is relevantly trustworthy. We need an emphasis on assurance to explain how trust can be disappointed yet unbetraysed – as, for example, when a promisor remains faithful to her promise despite, in unexpected circumstances, failing to do what she has promised to do. And an emphasis on assurance explains key instances of trust

undisappointed yet betrayed – as, for example, when following through on your intention betrays the self-trust informing that commitment. We'll discuss both kinds of case in detail.

The commitments – promises and intentions – on which we'll focus mark a contrast between faithful and rigid execution, between fidelity to A's normative expectation and a rigid adherence to it that violates the understanding that gives normative content to A's trust. Even when A does not accept an explicit invitation to trust – I discuss such cases in Section VII – A's trust rests on implicit assumptions about how B's doing what A trusts B to do serves A's planning needs or interests. When B does it without appropriate regard for those needs or interests, A's trust may prove betrayed though undisappointed. And when B refrains from doing it from concern that doing it violates those needs or interests, A's trust may prove disappointed yet unbetrays. When A goes through trust in B in these cases, A relies on B to provide a planning reason to ϕ . That reliance ensures that A's trust is not a mere extension of reliance on B to act.¹

I. Toward a new three-place analysis of trust

Let me begin by presenting my approach systematically within the broad dialectic that informs it. In this section, I explain how it provides an alternative to standard approaches to the nature of trust. In the next section, I begin my argument by explaining the core issue driving my alternative. These stage-setting discussions are, perhaps unfortunately, rather abstract and schematic. Having clarified what I think and roughly why I think it, I develop my argument with

¹ For a different account of trust that also emphasizes commitment, see Hawley 2014. In other respects, however, my approach differs substantially from Hawley's. E.g. Hawley endorses the traditional three-place analysis, does not distinguish fidelity from 'rigid' execution, and does not consider the sort of parallel between the interpersonal and the intrapersonal that informs my approach.

rich examples in Section III. (Readers who need examples to motivate their intuitions could skip directly to Section III and only later return to the overviews presented in the first two sections.)

My approach engages two philosophical debates about trust, deriving from two distinct questions about the nature of trust. The first asks how to define trust. Does trusting B to ϕ involve anything more than relying on B to ϕ ? Over the past three decades, debate on this question has revolved around three positions. Reductionism answers no: trusting B is just believing B relevantly reliable.² The Affective Attitude View answers yes; trust manifests felt optimism about B's goodwill.³ The Reactive Attitude View also answers yes, though for a different reason: the expectation informing trust is not merely predictive but normative, backed by a disposition to resent B for not ϕ ing (among other possible reactive attitudes).⁴ I argue for a fourth view, the Assurance View, by pairing the definitional question with a second question.

This second question addresses the normative structure of trust. Does trust most fundamentally embody a two-place or a three-place relation? On the standard three-place model, 'A trusts B to ϕ ' is most fundamental.⁵ On the two-place model, 'A trusts B to ϕ ' is less fundamental than 'A trusts B': what is fundamental is the relation between A and B.⁶ In the recent literature, this second debate comprises challenges to the standard three-place model meant to motivate a shift to the two-place model, but my challenge pushes in a different direction toward a new three-place model, on which 'A trusts B to ϕ ' is less fundamental than 'A ϕ s through trust in B.' Though we trust people to do things, trust more fundamentally lies in doing something through trust in a person: believing through trust in someone's testimony, acting

² See e.g. Hardin 2002, Chapter 3; Nickel 2007, Section 6; and Rose 2011, Chapter 9.

³ See e.g. Baier 1994, Chapters 6-9; Jones 1996; and McGeer 2008 (on the role of hope in trust).

⁴ See e.g. Holton 1994; Jones 2004; Walker 2006, Chapter 3; Hieronymi 2008; and Helm 2015.

⁵ Shared by approaches as diverse as Baier 1994, Holton 1994, Hardin 2002, and Hawley 2014.

⁶ See e.g. Lahno 2001; Faulkner 2015; and Domenicucci and Holton 2017.

through trust in someone's advice, and engaging in mutual activity through trust in someone's invitation to share an intention – along with the cases of promising and intending that I highlight. Though I agree with proponents of the standard three-place model that trust in a person most fundamentally makes reference to an action, I agree with proponents of the two-place model that the question whether to perform that action makes fundamental reference to the trust relation.

This new position on the second debate yields an equally new position on the first. I vindicate my new three-place model of trust by vindicating an affirmative alternative to the Affective Attitude and Reactive Attitude Views which more effectively counters the argument for Reductionism. The Assurance View thus yields a new account of the normative structure of trust and a new explanation of how trust differs from mere reliance. Trust differs from mere reliance through how it gives access to a reason to rely on the trusted. In mere reliance, your reason to rely on B is exogenous to the reliance relation: these reasons are typically grounded in B's track record of relevant reliability. When you trust B, by contrast, you take yourself to have a reason to rely on B that is endogenous to this interpersonal relation: you take yourself to have a reason grounded not merely in B's reliability but in B's responsiveness to your relevant needs.

There we have an outline of my position. Let me now offer an outline of my argument. I defend my three-place model by agreeing with defenders of the two-place model up to a point. A simple dialectic generates the two-place view. If trust does not reduce to mere reliance, how do they differ? Two-placers explain how trust differs from mere reliance by viewing three-place trust as informed by a deeper two-place trust relation. I too aim to explain how trust differs from mere reliance by viewing the three-place trust relation 'A trusts B to ϕ ' as informed by deeper trust. But I do not regard the deeper trust as two-place trust. This deeper trust, which explains how trust differs from mere reliance, has a three-place structure: 'A ϕ s through trust in B.'

What exactly is this three-place relation? Here's one thing I do not mean – though my shorthand formula may suggest it. Say A trusts B to do some particular thing. Must that trust lead A to perform some *action*, ϕ , through trust in B? Obviously not. Say you trust B to make plans for a picnic; there need be no action that you perform, or even plan to perform, in trusting B to make those plans. When I say that A ϕ s through trust in B, I do not mean that A takes some positive step to express his trust in B. A may merely rest assured that B will act. In our schema, 'A ϕ s through trust in B,' ϕ ing may merely amount to resting assured. What's important is that A regards himself as having a reason to adopt that stance: a reason to rest assured through trust in B. In our example, you presume that you have a reason to take B's picnic planning for granted – a reason to rest assured that B will plan competently – without seeing any reason to perform any further action to express that trust in B. On my Assurance View, that's part of the point of trust: to reap this rational reward of trust – the right to leave positive action up to the trusted – while undergoing trust's distinctive risk of betrayal. My thesis is that we can explain how trust differs from mere reliance not only by focusing on the risk of betrayal but also by viewing trust and trustworthiness as a source of reasons irreducible to mere reliance or reliability.

Why prefer my three-place model to the two-place? The key contrast between my three-place model and a two-place model derives from the contrast between rationality and morality. On a two-place model, given its explanatory aims, the trust relation is broadly moral. In trusting B, A brings himself into a moral relation with B either (along the lines of an Affective Attitude View) by making himself morally vulnerable to B's will or (along the lines of a Reactive Attitude View) by holding B morally accountable for her will. On my three-place model, the trust relation need not be moral, and in Section VI we'll consider one explanation why: B's betrayal of A's trust need express nothing like ill will toward A, or make B an appropriate target

for reactive attitudes such as resentment. In the cases on which we'll ultimately focus, B betrays A's trust through misinterpreting what the promissory agreement between them requires of her – an error that need not be moral or call out for reactive-attitudinal response. In φing through trust in B, A treats B's trustworthiness primarily as a source of reasons. Reductionists are right to oppose the Affective and Reactive Attitude Views: trust is not moral *as such*. But I say that as an anti-reductionist: what distinguishes trust from mere reliance is the distinctive way that trust gives reasons, in this key dimension of interpersonal (and, I later argue, intrapersonal) rationality. Assuming B worthy of A's trust, the reason B makes available is A's *reward* in trust. To receive that reward, A must make himself vulnerable to betrayal, the distinctive *risk* in trust.

II. On the risks and rewards of trust

I now begin my argument by motivating my emphasis on reason-giving rather than morality. This section, like the last, is rather abstract. I begin working from examples in Section III.

Any anti-reductive view of trust must explain how trust risks betrayal, not merely disappointment. It is clear enough how trust risks betrayal on the two established anti-reductive views. On the Affective Attitude View, B betrays A's trust if she fails to vindicate A's optimism about her goodwill toward him. On the Reactive Attitudes View, B betrays A's trust if she fails to live up to the normative expectation that informs his trust, a normative expectation backed by appropriate reactive attitudes toward B. My alternative approach begins from the hypothesis that what distinguishes trust from mere reliance lies not in either party's attitude toward the other but in the point of trusting someone – in what trust does for the one trusting. The hypothesis leads me to focus on B's act of assurance, and on what trusting B insofar as B assures A does for A.

Though trust need not respond to an assurance (more on that in Section VII), the normative structure of trust is clearest when A responds to B's invitation to trust. I argue that the normative structure of the trust relation takes this form: B represents herself as giving A a reason in offering her assurance, and A responds by coming to regard himself as acquiring that reason. If betrayal is what A risks in trusting, this reason is A's reward in trusting – if B proves appropriately trustworthy. B betrays A's trust if she fails to provide the reason – if she is not thus trustworthy.

My approach engages a background issue about the role of assurance in a trust relation. In her defense of the Affective Attitude View, Annette Baier characterizes that role as follows:

The assurance typically given (implicitly or explicitly) by the person who invites our trust... is not assurance of some very specific action or set of actions, but assurance simply that the trusting's welfare is, and will one day be seen to have been, in good hands." (1994, 137)

Baier develops a contrast on this point between an invitation to trust and a promise, which she calls "that peculiar case of assurance." I disagree that promises are peculiar instances of trust, but I more fundamentally disagree with her moral emphasis. Baier is on the right track, but the assurance at the core of an invitation to trust targets the trusting's rationality – his responsiveness to reasons – not the trusting's welfare. When we grasp how assurance works we'll grasp why a promise is not a peculiar case of assurance: in a key respect, it is the purest case.

The difference between my approach and Baier's arises from my emphasis on rationality and reason-giving, which I intend as a corrective to Baier's moral emphasis. In arguing that the risks of trust track the rewards of trust, I make two claims. First, it does not follow from how

trust risks betrayal that trust is a moral relation. Second, it does follow from how trust risks betrayal that trust is a rational relation. Trust is a rational relation, I argue, because to invite trust, as opposed to mere reliance, is to represent yourself as a source of planning reasons, both interpersonally for a promisee to do things that depend on your promise and intrapersonally for your later self to do things that depend on your intention. Betrayal can derive from a rational norm, because when you invite A's trust, by offering a propositional assurance, you represent yourself as taking responsibility for A's status as rational in letting himself be guided by your assurance. We thereby work our way up from rational considerations to moral ones. The obligation not to betray another's trust manifests a kind of normative power: you undertake this obligation when you give your assurance. On my approach, we best understand this normative power from the inside out: just as you undertake an obligation to be worthy of your own trust when you invite your own trust by forming a judgment or intention, so you undertake an obligation to be worthy of A's trust when you invite A's trust by giving A 'your word.'

On my Assurance View of trust, the distinction between trust and mere reliance mirrors Paul Grice's distinction between natural and non-natural meaning.⁷ Grice drew that distinction by contrasting an evidential mechanism – “Those clouds mean rain,” “Those spots mean measles” – with a mechanism that works through recognition of the speaker's intentions. In non-natural meaning, Grice argued, a speaker intends to give her addressee a reason to produce a certain response grounded not in evidence of her reliability or in any other evidential basis but specifically in the addressee's recognition of her intention to give him this reason.

⁷ Grice 1957.

What my approach inherits from Grice's is an emphasis on the distinctive way in which speech acts aim to give reasons: not through an evidential mechanism but through a structure of mutual recognition and understanding. To observe that such assurances aim to give a reason *simply* through the addressee's recognition of that aim is to observe that the speaker invites the addressee's trust. Such an assurance is an invitation to trust in this respect: B invites A to regard himself as having a reason to act or believe grounded, in part, in how B undertakes an obligation in issuing the invitation. I call such an invitation to trust a *propositional assurance*. Generally speaking, reliance is mere reliance unless thus invited by a propositional assurance; thus invited, however, it grounds an obligation, on B's side, and reasons, on A's side. Because the speaker's obligation derives from her claim to relevant reliability, the reasons are grounded in her status as relevantly reliable (not, as we'll see, in the trust relation itself). While both 'natural' and 'non-natural' reason-giving depend on B's reliability, there is a crucial difference in how they depend on reliability, and a correlative difference in what counts as relevant reliability.

We can understand the distinctive element in each species of propositional assurance – whether testimony, advice or a promise – by contrasting how it makes reasons available with the evidential mechanism that makes reasons available through mere assertion. Some common cases rest on this evidential mechanism: A treats B's assertion as reason-giving through his assessment of evidence that B is relevantly reliable. But the equally common cases that my approach highlights rest on a quasi-Gricean mechanism: A treats B's speech act as reason-giving through his trusting receptivity to B's influence, as governed by his exercise of a counterfactual sensitivity to evidence that B is *not* relevantly reliable. The contrast illustrates a key distinction between two forms of uptake. Sometimes you rely on someone because you have positive evidence that the person is relevantly reliable, but other times you lack such evidence. When

you lack evidence that someone is relevantly reliable, you can let your reliance on her be guided by your sensitivity to evidence that she is not relevantly reliable: if you get such evidence, you'll cease to rely on her; and (counterfactually) if you had such evidence you would not have relied on her. This second species of reliance is trust, and the species of reliability at issue is trustworthiness – worthiness of reliance not governed by positive evidence of relevant reliability.

Framing the contrast between forms of uptake now from the speaker's perspective, you give an assurance that invites your addressee to rely on *you* as a source of reasons, not merely on the reason-giving force of evidence of your reliability, and you present yourself as worthy of precisely that species of reliance. What grounds the reason is what would vindicate the presumption at the core of your invitation: your presumption to do justice to your addressee's needs in respects relevant to the shared understanding at the core of the invited trust relation. Mere reliance can be disappointed, but interpersonal trust can also be betrayed, I'm arguing, because betrayal taps into the normative role played by this shared understanding.

Here, then, is how the risks and rewards of trust fit together into a single normative structure. If you uphold your end of the trust relation by being relevantly trustworthy, your addressee gets a reason. If your addressee upholds his end of the trust relation, by trusting you in appropriate ways as determined by the understanding that you invite him to share in trusting you, you count as undertaking an obligation to be thus trustworthy. A trust relation can be betrayed because this shared understanding can be betrayed. In the next section, I treat two forms such a shared understanding can take: the implicit agreement at the normative core of a promise, and the ongoing understanding of what you're up to at the normative core of an intention. Just as a trustworthy promise makes available reasons for the promisee to do things that depend on the promisor's remaining true to the promise, so your trustworthy intention makes available reasons

to do things that depend on your following through on the intention. The understanding that is betrayed when trust is betrayed plays this normative role because it plays a more fundamental normative role in interpersonal and intrapersonal reason-giving.

III. The crux: trust disappointed yet unbetrayed

Let me now illustrate my approach with a detailed case – a case in which trust is disappointed yet unbetrayed. On the three-place model that I reject, a trust relation most fundamentally takes this form: A trusts B to ϕ . Within that normative structure, B's failure to ϕ suffices for her to count as betraying A's trust. I reject that model because I deny this claim of sufficiency. How, then, might B thus disappoint A's trust without betraying it? As I'll use the case to show, B does not betray A's trust when disappointing it manifests appropriate concern for A's planning needs.

One complexity is that I do not see how there can be such a case without a propositional assurance. Another is that, among interpersonal propositional assurances, it is plausible that only promissory assurances permit this possibility, since it appears to depend on the temporal articulation of a promise. I'll explain these complexities as we proceed, but let's first consider a case of promissory trust disappointed yet unbetrayed.

I'll use up-to-date examples in later sections, but let's first work from a case modeled on Hume's famous farmers whose corn ripens on different days (1978, 520).

Three Farmers. Ben owns a large farm and needs help in harvesting his crops next month. To get that help, he has promised to harvest this month the crops of two friends, Albert and Andrew, who own smaller farms, given that these friends

are traveling for urgent personal reasons and the crops they've planted need to be harvested this month rather than next. Albert and Andrew are depending on Ben; each is planning in concrete ways (rescheduling other commitments, committing to rent necessary equipment for those dates, etc.) that depend on Ben's remaining faithful to his promise. But as Ben makes preparations to harvest these crops, a swarm of locusts descends on his and his nearest neighbor's farm, preventing him from acting on these promises, since he is too busy fending off the locusts.

Should he apologize? Well, of course he should – to Andrew. As it happens, Ben's closest neighbor is Albert, whose crops are equally threatened by these locusts, and given this background it does not appear to make sense for Ben to apologize to Albert. The threat to Ben's crops is also a threat to Albert's crops – though not to Andrew's crops, since Andrew lives farther away. Ben *could* harvest Albert's crops, but let's stipulate the assessment that harvesting them under these conditions would damage them more, in ways that matter, than leaving them unharvested and instead fending off the locusts. However we imagine the details, this is key: not harvesting better serves not just Albert's interests but his *planning* interests; if Ben does not fend off the locusts, Albert will have to hire others to do so and perhaps to return from his urgent travels, thereby thwarting many planning needs, including the very needs that informed Ben's promise. It therefore does not make sense for Ben to apologize to Albert, and it equally would be wrong for Albert to hold Ben's non-performance against him when he decides whether to help Ben harvest his own crops next month.

What grounds the intuitions about accountability and apology is our grasp on what is at stake in the promissory agreement: Albert trusts Ben to do justice to Albert's ongoing needs in respects relevant to the point of Albert's reliance on Ben's promise, which is determined not by Albert's expectation the Ben will rigidly execute the promise but the implicit agreement that informs it. Our intuitions about accountability and apology derive from this difference between Ben's promises. Ben owes Albert an apology because he violates that promissory agreement, though with an excuse that makes the violation forgivable. But Ben need not offer Albert any excuse or ask for his forgiveness. All he owes Albert is an explanation that makes it clear how fidelity to his promise does not require acting in these unexpected circumstances. I'll say more about this distinctive element in the case in Section IV, using less canonical but more realistic examples.⁸

This distinctive element in the case focuses the crux of my argument. When A trusts B to keep her promise to ϕ , A need not engage in contingency planning for fear that B will ϕ *rigidly* – that is, without being guided by their shared grasp of the point of the promise. The planning reasons that B's promise gives A derive from B's aretaic concern to do justice to the promise, to remain faithful to it. In gaining access to these reasons, A does not trust B simply to ϕ . Again, if A trusts B simply to ϕ , then A will have to plan for the contingency that B ϕ s rigidly, without regard to their promissory agreement – an absurdity illustrated if Albert had to plan for the off-chance that Ben would harvest the crops amidst the locust swarms, thereby damaging them more than if he had left them be. As we'll see in Sections IV and VI, a promise thereby resembles an intention. Neither, as such, rationalizes contingency planning for counter-normative rigidity.

⁸ I discuss cases with this structure more fully in Hinchman 2017.

We can now see how in promissory trust, at least, the condition of fidelity serves the condition of concern. Here again are the two conditions, formulated generally:

Condition of fidelity: B does not omit to do what A is trusting her to do.

Condition of concern: B shows the right attitude toward A.

In testimony and advice, these conditions appear to apply separately, the second building on the first. A trusts B's testimony to give him the truth, but it betrays that trust if B does so through a lucky guess, or in some other way that shows no concern for his epistemic needs – including his need to meet the epistemic standard that applies in his context of inquiry. A trusts B's advice to tell him what he has reason to do, but it betrays that trust if self-absorbed B does so because she happens at the moment to have practical needs that resemble A's, with no responsiveness to A's needs conceived as such.⁹ But, as Three Farmers illustrates, B as promisor does what A as promisee is trusting her to do *by* thereby manifesting appropriate responsiveness to a subset of A's needs, conceived as such: those that inform the promissory agreement. B keeps the promise, thereby vindicating A's trust, not simply by doing what she promised to do but by remaining faithful to this promissory agreement. Unlike a testifier or advisor, a promisor cannot meet the condition of fidelity without also, and thereby, meeting the condition of concern. Disappointed trust need not be betrayed trust because disappointing your promisee's expectations, by failing to do what you have promised to do, need not amount to infidelity to your promise.

⁹ I defend the claims in this sentence and the last in Hinchman 2005a and 2014. In the latter, I argue that testimony resembles promising on the present point, but I do not want to presuppose that argument here.

We must therefore be careful how we describe the simpler case from which we began – trust undisappointed yet betrayed – when the trust in question is promissory trust. In such a case, A trusts B to keep her promise to φ , B does not disappoint A's trust (as we're using the term) because she does φ , yet B betrays A's trust because she does not φ with the right attitude toward A. In failing to meet the condition of concern, B also fails to meet the condition of fidelity. Trust can be undisappointed and yet betrayed when it is trust in testimony or advice, or when there is no propositional assurance. A testifier or advisor who speaks truthfully but without appropriate concern for your other relevant needs may count as relevantly faithful – as we might say, 'to the truth' – but as nonetheless betraying your trust. And someone whom you trust to φ , with no promissory agreement that she φ , may faithfully do what you're trusting her to do but betray your trust through her attitudes toward you. In promissory trust, however, what you trust the promisor to do is to remain faithful to her promise by being appropriately responsive to needs of yours – planning needs – that the promise itself makes a condition of her fidelity. If she is not thus responsive, she fails both conditions, not merely the condition of concern.

IV. The crucial parallel between promises and intentions

When we see this difference between promissory and these other forms of trust, our next question is why the former should have this feature. The answer lies in grasping how a promise resembles an intention. If we try to view an intention as a promise made 'to yourself,' we'll be tempted to view the trust in an intention as forward-looking: you trust your later self to remain faithful to your intention. That overlooks how an intrapersonal assurance – an invitation to trust – informs an intention. If we view an intention as an intrapersonal trust relation that unfolds

through time between distinct selves, earlier and later, which self invites trust and which accepts the invitation? The logic of the ‘invitation’ metaphor gives this answer: the earlier self issues the invitation and the later self, if all goes well, accepts it – trusting the earlier self, not to ϕ (since that’s up to the later self), but to have been worthy of trust in forming an intention to ϕ . Though this backward-looking trust relation conflicts with the notion that you ‘promised yourself you’d ϕ ,’ an intention nonetheless resembles a promise insofar as a trustworthy intention, like a trustworthy promise, serves as a source of planning reasons – of reasons to do things that you would not have reason to do if the intention or promise were not relevantly worthy of your trust.

There are thus two points of resemblance between a promise and an intention. In a promise and an intention alike, one party offers the other an assurance of performance – an invitation to trust that the commitment thereby undertaken will be faithfully executed. And in each, the party who receives an invitation that is indeed worthy of trust thereby also receives reasons to act or plan on the assumption that the commitment will be faithfully executed. The reasons are typically not reasons that support doing these other things in particular but reasons not to avoid doing them given that there are independent reasons to do them. Your promise to help thus gives me reason to plan on moving that heavy couch tomorrow – though the consideration that you would help did not contribute to my decision to move that couch. And my intention to clear my schedule for that hour gives me a reason to persist in my plan to move that couch – though I did not decide to move the couch because I had that gap in my schedule.

We can see the parallel more clearly by reminding ourselves how a promise may simply replace an intention. A would normally manage the picnic himself, let’s imagine, but on the present occasion he has other obligations. So B steps in and promises A that she will manage it.

Had A managed the picnic, his intention to grill meat would have given him a reason to buy meat – not, of course, a conclusive reason but a pro tanto reason inasmuch as it would make no sense to buy meat that no one will grill – a lacuna filled by his intention to grill. Since A cannot manage the picnic, B’s promise to grill the meat fills that gap. And so on for the other intended or promised performances involved in managing the picnic: each gives A planning reasons provided that the commitment in question is worthy of A’s trust.

What is it for such a commitment to be worthy of A’s trust? Here the parallel deepens: in each case, the commitment crucially serves an understanding shared with A. In the interpersonal case, it is the understanding that A shares with B about the point of the promise: for example, that the promise serves the needs of picnickers that are not served by attempting to picnic in an unexpected thunderstorm. In the intrapersonal case, it is A’s ongoing understanding of what he is up to in planning the picnic, which likewise includes the caveat that it makes no sense to persist in the plan in an unexpected thunderstorm. And so we may ask with equal rhetorical force across the two cases: would it constitute a failure to remain faithful to the promise or intention – to ‘follow through’ on it in a way controlled by the understanding that gives the promise or intention its normative content – if the agent in question failed to grill in an unexpected thunderstorm? Obviously not. Obviously, an agent who did grill in an unexpected thunderstorm (under an umbrella to keep the coals hot, but with soggy buns and tastelessly waterlogged meat) – doing what she or he promised or intended rigidly with no acknowledgment that circumstances have changed in ways relevant to the shared understanding at the core of the promise or intention – would show that she or he failed to share that understanding. By grilling in an unexpected thunderstorm, B no more keeps her promise to grill – in the sense that goes with remaining faithful to it – than A simply follows through on his intention to grill.

Imagine now that A holds B to her promise by insisting that, given her promise, she has a promissory obligation simply to grill – whether in a downpour, or in an earthquake, or whatever the circumstances. Contrary to what is sometimes called an ‘authority view’ of promissory obligation,¹⁰ A’s refusal to let B off the promissory ‘hook,’ in such unexpected circumstances, appears to have no bearing, just as such, on actually keeping B on that hook. What keeps B on the hook are the mutually understood terms of the promissory agreement – which I’m assuming do not include grilling in a thunderstorm, much less in a dangerous natural disaster. The terms of the agreement may include a provision that gives A some authority to interpret how the agreement applies, but even when B thus cedes to A authority to demand performance, A cannot exercise the authority however he wants. There must be some mutually understood point to the promise, defined in terms of how keeping the promise would serve A’s interests or needs. The parallel with intention clarifies why the agreement plays this normative role.

This feature of the normative relation is completely obvious for intention: when you form an intention to ϕ at t , you do so with a set of expectations about what the world will be like at t , in respects relevant to your intention, and an implicit understanding of which of these expectations are relevant to your follow through on the intention in this respect: if a relevant expectation is falsified, you should not simply follow through on the intention. In our picnicking example, A expects it not to rain but also expects that no one will show up wearing a bowtie. A has not formulated the latter expectation explicitly in his consciousness, but his betting dispositions, etc., show that he has it, and that it is just as strong as his expectation that it will not rain. If someone does show up in a bowtie, his understanding of the point of intending to grill

¹⁰ See, for example, Darwall 2006, Chapter 8, and 2011; Owens 2006 and 2012, Parts II and III; Shiffrin 2008 and 2011; and Watson 2009.

ensures that the falsification of his expectation tends not at all to show that he ought not to follow through on that intention. Does the difference lie in his consciousness of the expectation? We can easily think of expectations not present to A's consciousness that would nonetheless be relevant to A's follow through on this intention to grill. For example, A expects that a war will not break out among rival armies on the picnic grounds, though he has given the matter no conscious thought. If rival armies do go to war there, he will immediately see that this is relevant to his intention to grill and will not simply follow through on that intention.

These observations clarify how the parallel with promising informs our objection to the authority view. Does B have a promissory obligation to do what she has promised to do unless and until A positively lets B off this 'hook'? If B has promised some picnic guest A to grill but war unexpectedly breaks out on the picnic grounds, does A's obstinate insistence that B promised really ensure that B has an obligation to grill – an obligation overridden by other obligations, to be sure, but an obligation nonetheless? By hypothesis, this battlefield scenario formed not even an implicit part of their promissory agreement and in fact runs counter to that agreement, just as – in the intrapersonal parallel – it falsifies relevant expectations informing A's intention. How could B, by virtue of promising with that understanding of the point of the promise, be *on* this hook – imagine, if you need an even more extreme case, a deadly tornado bearing down on them, or warmongering Martians landing on the grounds with a tornado-like descent. If one replies that B ought not to grill merely because her obligation to grill is overridden by an independent obligation not to remain in harm's way, I revert to the simpler case in which grilling is not dangerous but pointless – because it does not serve the mutually understood point of the promissory agreement.¹¹ B does not need A's permission to manifest her

¹¹ Again, I'm assuming a stereotypical case in which there is relevant agreement that the

understanding that grilling in such unexpected circumstances would fail to keep her promise to grill for everyone's pleasure on what they all expect will be a Sunday afternoon perfect for grilling, with no forecast of grill-undermining rain or wind, not to mention bellicose Martians.

That parallel between promise and intention yields this formula: a promisor aims to do for the promisee what the promisee would have aimed to do for himself had he formed an intention to do it. The formula does not, of course, mean that promisees are always or even usually in position to intend to do what the promisor has promised to do. What it means is this: just as an intention loses its intrapersonal normative force, requiring redeliberation before rational follow-through, when a relevant expectation informing the intention is falsified, so a promise loses its normative force, no longer requiring performance of the promised act, when a relevant expectation informing the promissory agreement is falsified. What gives promises and intentions their parallel normative structures is the distinctive temporality of their commitments. When you form an intention, you aim to do justice to your ongoing needs in respects relevant to your understanding of the point of the intention. If your needs change because your circumstances unexpectedly change, your rational obligation to do what you intend to do may merely lapse, rather than being outweighed. If your promisee's needs likewise change through changing circumstances, then your obligation to do what you promised may likewise lapse.

We can also frame the parallel in terms of how the commitments give rise to planning reasons. As we've seen, B's promise to ϕ serves A's planning needs in the way that A's hypothetical intention to ϕ would serve his own planning needs. The parallel appears to rest on a deeper analogy: between how promissory trust gives planning reasons and how trust in your own

thunderstorm undermines the point of the picnic. In a less stereotypical case, we could imagine that a storm has no such consequence, as determined by the less stereotypical agreement defining the case.

intending self makes you rationally coherent.¹² In each case, the rational status appears to derive from vindication of the presumption of trustworthiness informing an invitation to trust. In each case, gaining or retaining the rational status rests, not on responsiveness to positive evidence of trustworthiness – evidence that may be hard to obtain – but on responsiveness to evidence of unworthiness of trust. As the promisee withdraws trust upon receiving significant evidence of the promisor’s untrustworthiness, so you redeliberate whether to do what you were intending to do upon receiving significant evidence of your own untrustworthiness in intending. You need not have positive evidence to be guided by evidence. Trust guides you in planning by making available reasons grounded in trustworthiness of which you may lack positive evidence – as long as you also lack evidence that the putative source of the reason is relevantly untrustworthy.¹³

V. How trust, including self-trust, gives access to reasons

At the core of my approach lie the ideas that trust serves as a conduit for giving reasons through assurance, and that this form of reason-giving differs from how a reason is given by evidence of the speaker’s reliability. It’s a large question how an assurance can give reasons, and how that form of reason-giving is distinctive. I’ve had my say about it elsewhere in papers on testimony, advice, and promising.¹⁴ For present purposes, we do best to focus on the intrapersonal case, developing the parallel between promises and intentions sketched in the previous section.

Continuing that argument, this will help us to see why the condition of concern must inform the

¹² I explore this deeper analogy in Hinchman 2017 and more fully in “Commitment as Normative Power,” in preparation.

¹³ While we lack space to pursue the parallel further, we might wonder how we develop this capacity for trust: does the development of capacity for diachronic agency internalize proto-promissory relations?

¹⁴ See Hinchman 2005a, 2005b, 2014, and 2017.

condition of fidelity for promises and intentions: it must do so because such concern lies at the core of reason-giving trustworthiness. Though my argument forces us to confront the objections that I'll now consider, my discussion in this section and the next pulls us into issues that run deeper than we can fully investigate here. I return to our core dialectic in Section VII.

Why emphasize intrapersonal reason-giving? And what could such reason-giving have to do with self-trust? Let's begin with some general observations about why we form intentions. Say it is now t_1 and your deliberative question is whether to ϕ at some later time t_2 . Why might you resolve the question now, at t_1 , by forming an intention to ϕ at t_2 , rather than leaving the matter unresolved until t_2 ? The answer will typically involve your desire to coordinate matters in advance, both because you may now possess better resources, whether better information or increased ability to use it, than you expect to possess at t_2 , and because in forming the intention now you give yourself a rational basis for doing other things that you would not do if you could not count on yourself to follow through on that intention. This rational structure emerges especially clearly for sub-plans within an overarching plan – you build the walls of your house today partly because you intend to add a roof tomorrow – but it also emerges when the actions embody separate plans. Say, in election season, you volunteer to campaign for candidate X. Some of your reasons to do so are the reasons to vote for X that informed your decision to vote for X, but one additional reason derives from your intention to vote for X – assuming the intention trustworthy. Your intention to vote for X does not itself provide a sufficient reason to campaign for X, but it does provide one reason, and that reason may even prove necessary: perhaps if you did not intend to vote for X, it would not in context make rational sense for you to campaign for X – given, say, your ineffectiveness in campaigning with an 'open mind,' your susceptibility to charges of hypocrisy, and the inefficiency of putting off that important decision.

It is thus part of the point of your forming an intention to ϕ that doing so will, you expect, give you reasons to do other things – to do things that you would not regard yourself as having a sufficient reason to do if you did not have this intention to ϕ . I have called these ‘planning reasons,’ since they are reasons, grounded in your presumed trustworthiness, that articulate the normative structure of your planning. Note well that I do not claim that forming an intention to ϕ can later give you a reason specifically *to* ϕ . That claim is more controversial than my anodyne observation about the point of intention, since it appears to yield the result that you can illicitly ‘bootstrap’ yourself into possessing reasons by a sheer act of will.¹⁵ When you lack good reason to do what you intend to do, does the fact that you made one error, in forming the intention, mean that you now have a reason to make another error by following through on it? When we examine the role of trust in agency, however, we see that the bootstrapping problem is more specific than merely willing your way into reasons. Say you’ve formed an intention to ϕ which you treat as giving you a reason to ϕ , and you now worry that this intention may be untrustworthy. If you (re)deliberatively affirm that intention even partly on the basis of that putative reason, you have merely gone round an illicit circle that leaves your worry unaddressed.

Since I’m going to argue that this problem does not arise for reasons to perform actions other than the action that you intend to perform, let’s examine how the problem arises. The problem reveals that even if we assume that you are trustworthy in forming an intention to ϕ , we should nonetheless not regard your trustworthiness as giving you a *deliberative* reason to ϕ . Though I eschew the distinction between ‘objective’ and ‘subjective’ reasons, the bootstrapping problem motivates a distinction between ‘deliberative’ and ‘non-deliberative’ reasons. Imagine

¹⁵ Bratman 1987 23-7, 86-7; Broome 2001. Smith 2016 offers a dissenting perspective.

you forget this intention for a while and then remember it through recognizing your own handwriting in your appointment calendar. You still don't remember the experience of forming the intention, but you treat the handwriting as evidence that you formed it. You now reason as follows: 'I'm reliable on the question whether to ϕ . So the fact that I formed an intention to ϕ gives me reason to believe that I ought to ϕ . So I ought to ϕ .' And you thereby reaffirm your intention to ϕ . There is nothing inherently wrong with this reasoning, and there need be nothing wrong with reaffirming your intention in this way. The rub for our purposes is merely that this is not a manifestation of normal diachronic agency. You don't govern yourself, looking forward, by leaving a trace of your intention where you expect your future self to discover it, and then offload responsibility to your future self to reason in the way just described. If you expect your future self to redeliberate the question whether to ϕ , now in a way informed by its evidence of your reliability in having formed an intention to ϕ , you aren't inviting that future self simply to follow through on the intention. And if you later find yourself in that predicament looking back, your attempt to cope with the deficit deliberately marks a contrast with the normal case in which you don't need to redeliberate because – having earlier deliberated – you can remember what you intend and then, trusting your intending self, simply follow through on the intention.

The bootstrapping metaphor thus codifies a puzzle. Part of the point of forming an intention lies in giving your future self reasons to do things that depend on your having this intention. Planning reasons must count as deliberative reasons, since you deliberate – partly but perhaps decisively – from the consideration that you intend to ϕ when you go on to form a further plan whose rationality depends on your assumption that you will ϕ by following through on this intention. But if your intention to ϕ does not give you a deliberative reason to ϕ , we may wonder, how can it give you a deliberative reason to do things that depend on the intention?

We can resolve the puzzle by noting that the bootstrapping problem does not rule out your having a *nondeliberative* reason to follow through on that very intention. What would ground that reason is not the sheer fact that you intend to ϕ but your trustworthiness in forming and retaining the intention. No bootstrapping problem infects the idea that your trustworthiness grounds a nondeliberative reason to follow through on an intention, since to say that the reason is ‘nondeliberative’ is to say that you cannot weigh it in deliberation and so cannot go round the illicit circle.¹⁶ If, by contrast, you reaffirm your intention to ϕ by deliberating from your presumed trustworthiness in intending to ϕ , you are illicitly bootstrapping, because that presumption is precisely the question posed by this deliberation. But when you deliberately weigh a planning reason to do something else, grounded in your presumed trustworthiness in intending to ϕ , you are not illicitly bootstrapping because you are not addressing any question of your trustworthiness in intending to ϕ . The bootstrapping problem arises only when your trustworthiness is specifically in question. If your trustworthiness in intending is not in question, there is no circle in deliberating from the presumption that it gives you planning reasons.

VI. How trust, including self-trust, is betrayed

What is it for trust, including intrapersonal trust, to be betrayed? Though it may seem like a digression, explaining the possibility of betrayed self-trust lets me reply to a pressing objection to my reliance on the parallel between promising and intending. We have been considering a parallel between reason-giving trustworthiness in promising and reason-giving trustworthiness in intending, a parallel grounded in a deeper parallel between the interpersonal agreement at the

¹⁶ See Hinchman 2003, 2009 and 2010 for full argument on this point.

core of a promise and the intrapersonal understanding of what you're up to at the core of an intention. It is clear enough how a promisor can be untrustworthy, but what exactly is it to be unworthy of your own trust in intending? If you are then, by my account, you do not get any planning reasons from the intention – so it loses a key part of its point as an intention. How should you respond? We now get what looks like a paradox. Insofar as you do have this intention, it seems you should follow through on it. Insofar as the intention is not worthy of your trust, it seems you should mistrustfully not follow through on it. But there is no paradox; we need merely get clear on what it is to betray your own self-trust. Understanding how self-trust is betrayed will put us on the right track to understand what it is to betray trust more generally and thus how to achieve our ultimate theoretical aim of distinguishing trust from mere reliance.

How could you 'mistrust' your own intention? At the core of the present objection lies the observation that when you 'mistrust' an intention you make it the case that you no longer have that intention: you reopen your deliberation whether to do the thing in question, which is incompatible with committing yourself to do it in the way of intention.¹⁷ If you cannot simply mistrust your own intention – that is, yourself insofar as you have formed and retain it – how could an intention manifest self-trust? Though you may have worries about an intention that you persist in retaining, just as you may about one of your own beliefs, if you simply 'mistrust' your intention, in the way that you might mistrust another's promise, you thereby cease to retain the intention, and for a reason that applies also to belief: both intention and belief manifest trust in your judgment. While there may be special contexts or respects in which you can trust yourself and yet at the same time count as mistrusting that trust, trust in your judgment does not appear to

¹⁷ For an elaboration of this and related problems, see Kolodny 2005, 528-539. For replies to Kolodny's arguments, see Hinchman 2013, Section III.

admit that possibility. You cannot trust your practical judgment that you ought to ϕ in such a way as to count as intending to ϕ yet at the same time also mistrust that trust in your judgment. But, while you cannot abandon an intention without abandoning the judgment that informs it, it does not follow that the only way to abandon an intention is to abandon the judgment that informs it. You might abandon an intention by mistrusting the judgment that informs it. Our question now is how this works: how might you mistrust a judgment that you nonetheless retain?

What is it to mistrust your own judgment? Let's work within the standard approach that equates your all-things-considered practical judgment that you ought to ϕ with a doxastic judgment that you have conclusive reason to ϕ .¹⁸ The view identifies the specifically practical element in a practical judgment with an element in the content of that judgment: the idea that you have a conclusive practical reason. We might ask what it is to judge that you have a *conclusive* practical reason, but I take that notion for granted.¹⁹ I have noted that you cannot mistrust your own belief, apparently for the very reason that you cannot mistrust your own intention: both intention and belief manifest trust in your judgment, and it does not make sense to say that you cannot mistrust this dimension of your own self-trust. As with intention, to mistrust your own belief that p is no longer to believe that p : to mistrust the belief is to abandon it. Why not say the same of mistrust in your own judgment? How could you retain your judgment that p – say, your judgment that you have conclusive reason to ϕ – while at the same mistrusting that judgment?

Let me now build up to my proposal. When I speak of 'mistrusting' your own judgment, I do not mean mistrusting your faculty of judgment. If 'mistrusting your judgment' could only

¹⁸ For an influential version of this approach, see Scanlon 1998 (25-30) and 2007.

¹⁹ I develop a challenge to this aspect of Scanlon's view of practical judgment in Hinchman 2013, Section VI.

mean mistrusting your faculty of judgment, then when you mistrust your judgment you would be – deliberately speaking – just stuck. You would have to stop deliberating and merely wait for the bout of self-mistrust to subside. But you are not just stuck when you mistrust your judgment; you can mistrust your judgment in this or that precise respect and resume deliberation by trusting your judgment in other respects. We individuate these ‘respects’ with propositions. Self-mistrust thus typically targets a subject matter, which we individuate with propositions, and a subject matter can be so narrow that it coincides with a single proposition. To say that you mistrust your judgment in the respect individuated by the proposition p is to say that you mistrust your judgment on the question whether p . Does it follow that you cannot count as judging that p ? If you were redeliberating whether p , then you would not count as judging that p . But from the fact that you mistrust your judgment on the question whether p , it does not follow that you are redeliberating whether p , since you may be deliberating *whether to redeliberate* whether p . Reconsidering your judgment that p may, in this way, fall short of reconsidering whether p .

What is it, then, to consider – again, not idly but with appropriate engagement – whether to reconsider whether p ? Some ways of considering whether to reconsider whether p *do* amount to abandoning your judgment that p . Considering whether to reconsider whether p on the basis of a worry about the truth-conducive reliability of the doxastic process whereby you formed your judgment that p is the first step, at least, toward abandoning that judgment. Once raised, the worry tends to force open deliberation on the question whether p – perhaps not all at once, but unless you simply let go of the worry that’s where you will wind up. You cannot settle a non-idle worry about your truth-conducive reliability without reconsidering whether p .

But there is another dimension of your reliability that your self-mistrust can target when you mistrust your own judgment: what I’ll call your ‘closure-conducive’ reliability. As I use the

term, ‘closure-conducive’ reliability is very different from truth-conducive reliability; though we aim at or care about the two species of reliability together, a given subject could at once be closure-conducively reliable without being truth-conducively reliable, and vice versa. The question whether a subject is closure-conducively reliable in believing that *p* is the question whether her doxastic deliberation whether *p* is or was informed by an appropriate implicit conception of when her evidence would epistemically suffice for her to close deliberation with a judgment that *p*. If your worry targets your closure-conducive reliability in judging that *p*, by contrast, you can refrain from reconsidering whether *p* – even while non-idly pressing the worry.

To worry about your closure-conducive reliability in judging that *p* is to worry that the doxastic deliberation informing your judgment was not governed by the appropriate epistemic standard: that your deliberation whether *p* was temporally ill-drawn (‘impatient,’ ‘impulsive,’ ‘too hasty,’ ‘rushed’) or insufficiently resourced (‘complacent,’ ‘incurious,’ ‘lazy,’ ‘parochial’). You know you should not take too long to decide, since life is short, but maybe you are being impatient. You know you should not think too hard about the decision, since these mental resources are finite and also needed elsewhere in your life, but maybe you are just being complacent. This is not yet, in either case, to worry that you made a mistake *within* the deliberation. Your worry targets not how you weighed evidence while deliberating but your implicit conception of what was at stake in the deliberation. You need not worry that you have made an error in your reasoning or are less likely than you had assumed you would be to hit on the truth. What you would learn in learning that you are closure-conducively unreliable is that you are likely to have misassessed what is at stake in your doxastic context. You would learn that you are likely to have misassessed how much or generally what quality of evidence would epistemically suffice for you to close deliberation with a judgment.

When we specify that the doxastic judgment in question has practical content, we can diagnose the predicament you confront in a case of practical self-mistrust: you wonder whether following through on your intention to ϕ will amount to betraying your own trust. In following through, you would manifest trust in your intention, but the betrayal would lie in the judgment that informs your intention, your judgment that you have conclusive reason to ϕ . We could pin the self-betrayal on how this judgment invites your trust – though nothing turns on the metaphor. That’s one risk that you run whenever you follow through on an intention: you risk betrayal by your own practical judgment. You can address the risk by being open to self-mistrust. Without yet abandoning your judgment, you can raise the question of your trustworthiness as a worry that you were impatient or complacent in forming it; you thereby worry that your judgment betrays the trust that it invites through representing you as closure-conducively reliable. Combined with relevant truth-conducive reliability, your closure-conducive reliability forms the basis of your worthiness of your own trust as a source of planning reasons. As in the promissory case, responsiveness to betrayal and responsiveness to reasons go hand in hand.

Can we generalize this account beyond betrayed self-trust to betrayed promissory trust? On the account I’m developing, there is room for two explanations of a promisor’s, B’s, betrayal of your trust: (i) B forgets or rejects her promissory obligation; (ii) B misinterprets that obligation, by misinterpreting how the promissory agreement governs what it demands of her. B might misinterpret the agreement by failing to see that it obligates her to ϕ . Perhaps she interprets her promise to grill at the picnic as defeated by unexpectedly bowtied attendees or (more realistically) by an unexpected chill in the air, revealing that she misunderstands how relevant others perceive the bearing of weather on picnics. (In her native south, no one would grill in near-freezing temperature, but here in the north people do so regularly.) Or perhaps she

fails to interpret her promise as defeated by a thunderstorm, or by a tornado warning. Both type-(ii) violations – both her laxly failing to grill in a chill and her rigidly grilling in a downpour – amount to failures in B’s judgment. These failures of judgment betray A’s trust as directly as more stereotypical type-(i) violations, wherein B forgets or rejects her promissory obligation.

The parallel with intention thus helps us trace some instances of betrayed trust back to a question of judgment: specifically, how judgment is exercised in interpreting what an understanding of the implicit context and orienting content of a commitment normatively requires of you. Our next step is to see that type-(ii) betrayals of trust reveal most clearly how trust differs from mere reliance. We can treat some type-(i) violations as targeting the condition of fidelity without touching the condition of concern, as when B simply forgets or overlooks or non-culpably ignores her promissory obligation. And we can treat other type-(i) violations as targeting the conditions of fidelity and concern in different respects. If B rejects or spurns or rebels against her promissory obligation, that reveals that she has the wrong attitude toward the promisee, A, thereby violating the condition of concern. But her attitude bears a merely causal relation to her violation of the condition of fidelity: she does not violate both conditions with a single attitude or state of her will. If B rejects the promissory obligation, her failure has these two components: she fails to do what she promised, and the causal explanation of that failure lies in her lack of appropriate concern toward A. The first failure is the same as it would have been if B’s failure of performance had manifested mere forgetfulness, and the second failure is the same as it would have been if B had performed but with a contemptuous attitude toward A.

In a type-(ii) violation, by contrast, the two elements do not come apart. Or rather, the only way they come apart is if B does what she has promised to do inadvertently, without any regard to the promise. If we set that sort of case aside, we get the following result. In a type-(i)

case, B can violate the condition of fidelity without violating the condition of concern (e.g. B merely forgets to act), and B can violate the condition of concern without violating the condition of fidelity (e.g. B acts with an inappropriate attitude). But in a type-(ii) case, B satisfies the two conditions together: her fidelity is governed by her concern, and a lack of fidelity already manifests a lack of concern. If B fails to perform in a type-(ii) case, that failure manifests the failure of concern inherent in misinterpreting the promissory agreement. And if B fails to show appropriate concern in a type-(ii) case, that failure manifests a failure of fidelity to the promise – unless, again, she just happens to perform the act for unrelated reasons, a possibility that we’re setting aside for both types of case. We cannot get a case in which B violates the condition of concern, by misinterpreting what the promise requires of her, yet nonetheless satisfies the condition of fidelity, by doing what she promised to do because she promised to do it, or a case in which B violates the condition of fidelity, by failing to do what she promised to do as an expression of her understanding of the promise, yet nonetheless satisfies the condition of concern, by correctly interpreting what the promise requires of her. In a type-(ii) case, but not in a type-(i), violation of each condition goes with violation of the other, or satisfaction with satisfaction. Type-(ii) violations thus reveal more clearly how trust differs from mere reliance: they reveal what links the two conditions, violation of which amounts to trust betrayed.

We can draw one further moral from type-(ii) cases. Focusing on such cases lets us see clearly how the Affective and Reactive Attitude Views of trust misconstrue core instances of trust betrayed. The Affective Attitude View can explain type-(i) cases but not type-(ii) cases, for the simple reason that the betrayal of trust that we find in type-(ii) cases need not manifest any ill will toward the promisee. Some type-(ii) cases will prove entirely non-moral – if, for example, B’s misinterpretation of the promissory agreement was ‘innocent’ because, as the exculpating

catchphrase puts it, ‘merely a misunderstanding.’ When B interprets her promissory agreement with A as requiring her to grill only as long as the temperature does not drop close to freezing – thereby misinterpreting what anyone in A’s position would have taken for granted about its content, she betrays A’s promissory trust, but not in a way that calls for reactive-attitudinal response. At least, we do sometimes encounter cases along these lines: B’s misinterpretation is so natural – she’s from the South, where it is unthinkable to grill at near-freezing temperatures – that A rightly catches himself before giving in to any impulse toward resentment. There is no ill will, contra the Affective Attitude View. And, contra the Reactive Attitude View, betrayal does not warrant reactive-attitudinal response. But since my Assurance View offers a positive account of betrayed – via the parallel with intrapersonal trust – there is, in the end, no need to treat reactive attitudes as the key to explaining how trust differs from mere reliance: what explains betrayal is what the reactive attitude would, when appropriate, respond to. My alternative three-place analysis, with its emphasis on reason-giving through trust, thus explains both how disappointment need not yield betrayal and how betrayal need not be moral.

VII. What then of trust without assurance?

The point of the previous section’s deep dive into intrapersonal trust was to excavate a structure that explains how it and interpersonal trust univocally risk betrayal. Let’s now resume a broader social focus. My approach assigns a role for assurance in mediating trust relations, and that role motivates a novel three-place analysis of trust, on which the fundamental normative structure of a trust relation is not ‘A trusts B to ϕ ’ but ‘A ϕ s (or has a reason to ϕ) through trust in B.’ I have discussed interpersonal cases that appear to vindicate that analysis, but I concede that there are

other cases of trust that do not, since they do not contain anything recognizable as an assurance. How does my approach treat such cases? Must I bifurcate my approach and theorize these assurance-free cases on one of the standard models, whether three-place or two-place?

Before I answer, let me provide a brief review. When we focus on assurances such as promises and intentions, in which the conditions of fidelity and concern do not come neatly apart, we see how the risks of trust track the rewards of trust. A promise or intention aims not merely at rigid execution but at fidelity to an understanding, whether the interpersonal understanding of what is at stake for the promisee at the core of a promise or the intrapersonal understanding of what the agent is up to at the core of an intention. This fidelity to understanding informs the trustworthiness grounding the planning reasons that constitute the rewards of trust – reasons for the promisee to plan or act in a way that depends on the promise, or reasons for the agent to plan or act in a way that depends on his own intention.

We have seen how this framework explains how trust can be disappointed yet unbetrayed, since in unexpected circumstances, fidelity to relevant understanding does not require refraining from disappointing the trust (and broader considerations may actually require disappointing the trust – though this is not strictly necessary for trust disappointed yet unbetrayed). The framework also offers a satisfying explanation of how trust can be undisappointed yet betrayed. An affective-attitudinal approach to promissory trust would emphasize the badness of the promisee's attitude in such a case: she does what she promised to do, but in a way that manifests ill will toward the promisee. But if that amounts to a betrayal of trust, it is not a betrayal of specifically promissory trust. Perhaps that explains why Baier, a proponent of that approach, does not regard a promissory assurance as a paradigmatic invitation to trust. On my assurance-theoretic approach, by contrast, the betrayal goes right to the core of

the promissory relation. When B rigidly executes her promise to A though it does not serve A's needs in respects relevant to the promissory agreement between them, there is a betrayal of A's trust that may have nothing to do with any ill will that B bears A. B may actually have good will toward A yet nonetheless betray A's trust – A's specifically promissory trust – through confusion over what the interpersonal understanding between them requires of her. Here too we see a parallel with intention. If you act on an intention though relevant expectations – expectations that inform your understanding of what you are up to in intending – go false, you may have betrayed your own trust, by betraying this understanding, without bearing yourself ill will or any other problematic attitude – other than your lack of this intention-specific self-concern.

By 'self-concern' here I mean concern to act from this understanding of what you're up to in intending, just as the condition of concern requires that a promisor act from the understanding that she shares with her promisee *qua* parties to the promise. There is an attitude of concern in each case, but it is specific to the normative demands of the promise or the intention. The concern is not an extra element, codifying the role of any further attitude toward the one who offers his trust. And we may say the same of other reason-giving assurances such as testimony and advice. Each brings its own cognate species of concern for the addressee, a species of concern that informs trust by informing the reason-giving relation. Though there is always the possibility that a broader trust relation will be betrayed, betrayal of trust invited by an assurance rests on attitudes that are specific to how that assurance gives reasons.

What then of such broader trust – of trust without assurance? Note first that, like other theorists of trust, I do not aim to explain everything we might call 'trust.' When you say you 'trust' your new toaster (unlike its 'untrustworthy' predecessor) not to burn your toast, you mean merely that you rely on the toaster: there is no conceptual space there to distinguish trust from

mere reliance. Like other theorists of trust, I aim is to explain that distinction, and so my account is silent about uses of 'trust' where the distinction could not apply. The present challenge to my approach is that you can trust someone who has not made an assurance, with trust conceptually distinct from mere reliance on the person. How can an assurance-theoretic approach explain that distinction? How, without an assurance, can trust differ from mere reliance?

We cannot set aside such interpersonal trust, as we might set aside its intrapersonal analogue. How to set aside what at first may look like intrapersonal trust without assurance? Though it is a stretch to appeal to intuitions about self-assurance in intrapersonal practical commitment, there is a kind of case that resembles an intention but without the diachronic complexity that I've argued forms the core of intention. Consider Homer's Odysseus as he confronts the Sirens' song.²⁰ One reason to distinguish Odysseus's strategy for dealing with expected temptation from an intention is that having his hands tied to the ship's mast prevents more than merely giving in to the Sirens' deadly enticement. If, in an unexpected turn of events, a hurricane blows the Sirens away but also threatens the boat, he'll need to escape those bonds to deal with the emergency. It isn't merely that such pre-commitment carries risks but that the risks are not like the risks of inherent in diachronic agency. When you cope with temptation by forming a resolute intention, inviting your later self's trust, you risk forgetfulness (you may simply not remember your intention), weakness (you may be overcome by the temptation), practical unwisdom (perhaps, as in Huck Finn's resolution to betray his friend Jim, the temptations themselves are backed by stronger reasons), and lack of appropriate self-concern (you may misconstrue the bearing of unexpected circumstances on your intention, in the way that I have argued amounts to betraying your own trust). But you do not typically risk the weirdly

²⁰ For provocative treatments of this case, see Elster 1985 and 2000.

mechanical rigidity characteristic of Odysseus's pre-commitment when coping with temptation. Such a strategy may prove helpful in keeping your conduct in line with your reasons, but without an invitation to trust – without what we are conceiving as the intrapersonal analogue of an assurance – it does not involve anything like the self-trust dynamic at the core of an intention.

Though intrapersonal reliance without an assurance falls short of intrapersonal trust, we cannot draw the parallel conclusion for interpersonal reliance and trust. There are several types of interpersonal case to consider, in each of which trust does not respond to an assurance. In a testimonial case, you trust a testifier who has not addressed you, treating her assertion not merely as a reliable guide to the truth but as informed by concern for the context-sensitive epistemic needs of someone in circumstances *like yours*. In an advisorial case, you trust an advisor who has not addressed you, treating her advice not merely as a reliable guide to your reasons but as informed by concern for how someone *like you* would receive it. In a sub-promissory case, your trust in the other party is not informed by an understanding with the diachronic complexity of a promissory agreement: if you merely trust B to ϕ , with no promise, then you are not relying on B to do justice to your ongoing needs in the specific respect in which you do so when B has promised you to ϕ (unless, of course, ϕ ing itself requires such responsiveness).

What do the cases have in common? How might trust without an assurance be betrayed? Without an assurance, B can betray A's trust only under the fiction that B has invited it. In testimonial and advisorial cases, you treat B as concerned with an addressee who resembles you in relevant ways. There are two ways you might do so: you might work from available evidence that this resemblance between you and B's actual addressee does indeed obtain, or (perhaps lacking such evidence) you might treat B's concern as bearing on you more directly. The former option collapses the case into a case of mere reliance, since your normative relation to B now

runs through evidence of B's reliability – both in the dimension of truth-seeking and in the dimension of appropriate concern. To trust B in a way that stands opposed to merely relying on B, you would have to take the latter option, treating B's concern as bearing on you directly. But that is to treat B *as if* she had offered you an assurance, as if she had invited your trust. The most interesting sort of case is institutional: you trust institution B to serve as a source of reciprocal planning reasons, as if you stood to it both as promisee (e.g. relying on a school to educate your children) and as promisor (acknowledging its reliance on your support) – even if there is nothing like an actual promise. Though I lack space to give these observations the theoretical development they deserve, I believe that many forms of social trust have this *as-if* structure.

In each such case, treating the trusted as if he, she, or it had invited your trust can give you planning reasons grounded in his, her, or its status as worthy of that trust. Under this fiction, a failure of trustworthiness amounts to a betrayal of your trust. To say that the betrayal unfolds under the fiction captures its two sides: you feel betrayed by the trusted, but the trusted bears no attitude and performs no act specifically addressed *to you*. Is the trusted's worthiness of your trust, in the more favorable case, therefore also fictive? If it is, must we say the same of the reason it grounds? I do not believe that we should answer either question in the affirmative, but to show why would require an inquiry into the nature of reasons – too much for the present occasion. Setting those more difficult general questions aside, we see illustrated again our guiding thesis, that the fundamental normative structure of trust is 'A ϕ s through trust in B.' The point of such a fiction would lie in how it mediates your access to reasons.

References

- Austin, J. L.
1975 *How to Do Things with Words* (Cambridge: Harvard University Press)
- Baier, Annette
1994 *Moral Prejudices* (Cambridge, Mass: Harvard University Press)
- Bratman, Michael
1987 *Intention, Plans, and Practical Reason* (Harvard University Press)
- Broome, John
2001 "Are Intentions Reasons? And How Should We Cope with Incommensurable Values?" in C. Morris and A. Ripstein (eds), *Practical Rationality and Preference* (Cambridge University Press), 98-120
- Darwall, Stephen
2006 *The Second-Person Standpoint* (Cambridge: Harvard University Press)
2011 "Demystifying Promises," in H. Sheinman (ed.) *Promises and Agreements* (Oxford: Oxford University Press), 255-76
- Domenicucci, Jacopo and Richard Holton
2017 "Trust as a Two-Place Relation," in P. Faulkner and T. Simpson (eds), *The Philosophy of Trust* (Oxford: Oxford University Press), 149-60
- Elster, Jon
1984 *Ulysses and the Sirens* (Cambridge: Cambridge University Press)
2000 *Ulysses Unbound* (Cambridge: Cambridge University Press)
- Faulkner, Paul
2015 "The Attitude of Trust is Basic," *Analysis* 75:3, 424–29
- Grice, Paul
1957 "Meaning," *Philosophical Review*, 66 (1957), 377-88
- Hardin, Russell
2002 *Trust and Trustworthiness* (New York: Russell Sage Foundation)
- Hawley, Katherine
2014 "Trust, Distrust, and Commitment," *Noûs* 48:1, 1-20
- Helm, Bennett
2015 "Trust as a Reactive Attitude", in N. Tognazzini and D. Shoemaker (eds), *Oxford Studies in Agency and Responsibility*, Volume 2 (Oxford: Oxford University Press), 187-215
- Hieronymi, Pamela
2008 "The Reasons of Trust," *Australasian Journal of Philosophy* 86:2, 213-36
- Hinchman, Edward
2003 "Trust and Diachronic Agency," *Noûs* 37:1, 25-51
2005a "Advising as Inviting to Trust," *Canadian Journal of Philosophy* 35, 355-86
2005b "Telling as Inviting to Trust," *Philosophy and Phenomenological Research* 70:3, 562-87
2009 "Receptivity and the Will," *Noûs* 43:3, 395-427
2010 "Conspiracy, Commitment, and the Self," *Ethics* 120:3, 526-56
2013 "Rational Requirements and 'Rational' Akrasia," *Philosophical Studies* 166:3, 529-52
2014 "Assurance and Warrant," *Philosophers' Imprint* 14:17, 1-58

- 2017 "On the Risks of Resting Assured: An Assurance Theory of Trust," in P. Faulkner and T. Simpson (eds), *The Philosophy of Trust* (Oxford University Press, 2017), 51-69
- Holton, Richard
1994 "Deciding to Trust, Coming to Believe," *Australasian Journal of Philosophy* 72:1, 63-76
- Hume, David
1978 *Treatise on Human Nature*, 2nd Edition (Oxford: Oxford University Press)
- Jones, Karen
1996 "Trust as an Affective Attitude," *Ethics* 107:1, 4-25
2004 "Trust and Terror," in P. DesAutels and M. Walker (eds), *Moral Psychology* (Lanham: Rowman & Littlefield), 3-18
- Kolodny, Niko
2005 "Why Be Rational," *Mind* 114:455, 509-63
- Lahno, Bernd
2001 "On the Emotional Character of Trust," *Ethical Theory and Moral Practice* 4:2, 171-89
- McGeer, Victoria
2008 "Trust, Hope, and Empowerment," *Australasian Journal of Philosophy*, 86:2, 237-54
- Nickel, Philip
2007 "Trust and Obligation-Ascription," *Ethical Theory and Moral Practice* 10:3, 309-19
- Owens, David
2006 "A Simple Theory of Promising," *Philosophical Review* 115:1, 51-77
2012 *Shaping the Normative Landscape* (Oxford: Oxford University Press)
- Rose, David
2011 *The Moral Foundation of Economic Behavior* (Oxford: Oxford University Press)
- Scanlon, T. M.
1998 *What We Owe to Each Other* (Cambridge, Mass: Harvard University Press)
2007 "Structural Irrationality," in G. Brennan, R. Goodin, F. Jackson, and M. Smith (eds), *Common Minds* (Oxford: Oxford University Press), 84-103
- Shiffrin, Seana
2008 "Promising, Intimate Relationships, and Conventionalism," *Philosophical Review* 117:4, 481-524
2011 "Immoral, Conflicting, and Redundant Promises," in R. J. Wallace, et al., *Reasons and Recognition* (Oxford: Oxford University Press), 155-78
- Smith, Matthew
2016 "One Dogma of Philosophy of Action," *Philosophical Studies* 73:8, 2249-66
- Walker, Margaret Urban
2006 *Moral Repair* (Cambridge: Cambridge University Press)
- Watson, Gary
2009 "Promises, Reasons, and Normative Powers," in D. Sobel and S. Wall, *Reasons for Action* (Cambridge: Cambridge University Press), 155-78