

4

On the Risks of Resting Assured An Assurance Theory of Trust

Edward S. Hinchman

What do you risk when you trust? On some influential accounts, trust distinctively risks not mere disappointment but betrayal.¹ What exactly is it for trust to be betrayed? If the trusted simply fails to do what you've trusted her to do, that looks like an occasion for mere disappointment. Though there may be something in your relations with the person that warrants your feeling betrayed, betrayal does not appear to lie in the performative lapse as such. Where else may it lie? I'll argue that the risk of betrayal lies at a deeper level: in the risk that her action or inaction—whether disappointing or not—will manifest a failure to engage your needs in the way that you're trusting her to engage them. Though trust includes reliance, and reliance aims not to be disappointed, I'll argue that the aim distinctive of trust is more complex and is compatible with disappointment. Disappointed trust may thus fall short of betrayed trust, and betrayed trust may not derive from disappointed trust. It is betrayed trust, not disappointed trust, that violates the mutual understanding at the core of interpersonal trust. And it is betrayed self-trust, not disappointed self-trust, that violates the self-understanding at the core of intrapersonal trust. To understand this normative structure at the core of both social and personal agency, we must understand how each form of trust distinctively risks not mere disappointment but betrayal.

What is it to risk betrayal of your trust? Annette Baier argues that the risk is moral:

The assurance typically given (implicitly or explicitly) by the person who invites our trust, unlike that typically given in that peculiar case of assurance, a promise or contract, 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. (Baier 1994: 137)

Baier thus posits a contrast between an invitation to trust and a promise. I reject this contrast, but I more fundamentally reject her moral emphasis. Here I side with those

¹ See, for example, Baier (1994: Chapters 6–9); Holton (1994); Jones (2004); Walker (2006: Chapter 3); Hieronymi (2008); McGeer (2008); McMyler (2011: Chapter 4); and Hawley (2014).

who have argued against moralizing trust.² But these critics also reject Baier's emphasis on betrayed as opposed to disappointed trust. The critics link their rejection of what they regard as moralism with a claim that the risk of betrayal adds nothing, as such, to the risk of disappointment, thereby omitting the interpersonal element that Baier marks with the concept of an assurance. Baier is on the right track in emphasizing betrayal, I'll argue, but the assurance at the core of an invitation to trust targets the trusting's rationality, not the trusting's welfare or any other distinctively moral status. I do not emphasize rationality to the exclusion of morality: I'm going to argue that a promissory assurance gives rise to a promissory obligation insofar as it targets the promisee's rationality, and a promissory obligation is paradigmatically moral. I claim merely that the rational obligation is more fundamental. Though it does not follow from how trust risks betrayal that trust is a moral relation, it does follow from how trust risks betrayal that trust is a rational relation.

Since it is easier to understand how the obligation is thus fundamentally rational in the context of our self-relations, I'll explain the rational obligation at the core of interpersonal trust by developing an analogy with intrapersonal trust. Both species of trust, interpersonal and intrapersonal, differ from mere reliance insofar as they underwrite distinctive rational relations. The analogy marks how betrayal reflects the form of understanding at the core of any trust relation. Interpersonal trust underwrites a form of interpersonal reasoning that rests on a non-evidential mechanism whereby one person can make a reason available to another—'non-evidential' because the reason is not grounded in evidence of reliability. Intrapersonal trust underwrites the medium wherein an individual person maintains enkratic rational coherence.

In each case, trust is betrayed, rather than merely disappointed, when one party—whether a person or an aspect of one's own self—violates the shared understanding that generates norms governing the trust relation. Interpersonal trust is betrayed when the trusted violates the shared understanding of how the trust matters to the trusting, where how the trust matters reflects how the trust provides or fails to provide the trusting with reasons. And intrapersonal trust is betrayed when the trusted—an aspect of one's own self—warrants akratic self-mistrust, thereby violating one's own understanding of the requirements of enkratic coherence. In speaking of a rational relation, I don't claim that reasons and rational requirements amount to the same thing. I'll merely use each to cast explanatory light on the other, treating rationality as a broad category that includes both (if you want, disjunctively). The understanding at the core of both species of assurance, conceived as shared between parties to the trust relation, informs that relation in this key respect: when the understanding is violated the trust is betrayed.

The chapter has the following structure. In section 1, I'll develop examples that vindicate the possibility of betrayed though undisappointed trust. In section 2, I'll show what that possibility reveals about the rational relations invited by an interpersonal

² See, for example, Hardin (2002: Chapter 3); Nickel (2007, Section 6); and Rose (2011, Chapter 9).

assurance. In section 3, I'll develop examples that vindicate the possibility of unbetrayed though disappointed trust. In section 4, I'll show what that reveals about the rational relations invited by the intrapersonal assurance at the core of an intention. In section 5, I'll generalize the conclusion of section 1: the self-trust that informs intrapersonal assurances can be betrayed without being disappointed. In each kind of case, whether interpersonal or intrapersonal, the key to understanding the nature of trust lies in understanding how trust makes you vulnerable to the risk of betrayal.

1. Trust Undisappointed yet Betrayed

Though betrayal may bring its own disappointments, the concept of betrayed though undisappointed trust is perfectly coherent. Let me illustrate with some simple cases—cases that do not, I'll stipulate, involve any explicit or implicit assurance. These cases will mark a useful contrast with the more complex cases that we'll consider in later sections.

To get us started, consider this case:

Lovely Leaf. Andrew and Bernice are out hiking, and Andrew finds a leaf that strikes his fancy. Nursing it in his palm, Andrew asks Bernice to keep the leaf in her backpack, since Andrew's pack is full. 'Why?' she asks. 'No particular reason,' he replies. 'I merely like it.' Bernice accepts the leaf and places it in her pack. Though Bernice makes no explicit or implicit promise (let's stipulate), Andrew nonetheless trusts her to take good care of the leaf.

Will Bernice betray Andrew, or betray his trust, if she fails to take adequate care of the leaf? The obvious answer is that it depends on how she fails. If she underestimates the leaf's fragility and tucks it a bit too tightly against the tent poles, that would reflect on her competence and thus on her worthiness of Andrew's reliance, but it is difficult to see how that would amount to any betrayal of Andrew. How might we get a betrayal? An obvious betrayal emerges if we simply imagine Bernice's incompetence willed:

Lonely Leaf. As in *Lovely Leaf*, except that Bernice places the leaf on the ground next to her pack, instead of in her pack, confident that Andrew will not notice that she has left it behind and will soon forget about it.

Bernice thereby betrays Andrew's trust, since she doesn't even try to do what he is trusting her to do. But what if she not only tries but succeeds in doing it? Could Bernice nonetheless betray Andrew's trust? Consider this case:

Less Lovely Leaf. As in *Lovely Leaf*, except that when Bernice places the leaf in her pack, she thinks to herself 'What an idiot—I'll shut him up by putting the leaf in here with my stuff where it probably won't come to any harm.'

One might think that this could not amount to betrayed trust as long as Andrew is not aware of Bernice's contemptuous attitude. I'm not sure that's right: why can't her

betrayal simply be unknown to him? Let's nonetheless set that issue aside by considering a realistic case in which Andrew is aware of Bernice's attitude:

Unlovely Leaf. As in *Lovely Leaf*, except that Andrew and Bernice are now hiking back to their campsite in fading daylight. When he asks her to keep the leaf, Bernice's 'Why?' reveals impatience. 'And why,' she adds, 'are you occupying yourself with such trivia when we're trying not to get lost in the dark?' 'I like it,' he replies. Rolling her eyes in disgust, Bernice accepts the leaf and places it in her pack with sarcastic gestures of play-acted 'carefulness'.

Could Andrew think to himself 'She's contemptuous of my needs, but I nonetheless trust her to keep the leaf safe from harm'? He could, if he regards the sarcasm as merely meant for display. But if he regards her as genuinely contemptuous of his needs, or of this particular stated need, then he cannot rationally or reasonably trust her with the leaf—no matter how reliable he may regard her as a custodian of her pack, which he now sees includes this precious cargo. Let me emphasize that this needn't be a question of how Bernice 'really feels' about Andrew: the question is how she is disposed to act, not the true state of her feelings. If he regards her as disposed to act without appropriate concern for his needs, then, while he can rationally rely on her to keep the leaf from harm, he cannot rationally trust her with the leaf.

I do not claim that one couldn't get away with describing Andrew as 'trusting' Bernice with the leaf. We sometimes say 'trust' when we merely mean 'rely'—as when we wonder whether to 'trust' this toaster with a slice of bread. I claim merely there is a distinction worth drawing here, and that it makes good sense to draw it as the distinction between trust and mere reliance. If we challenge Andrew when he says that he 'trusts' Bernice with the leaf, it would make good sense for him to retreat to the claim that he is relying on her. It would be irrational for him to trust her, *over and above* relying on her. Though the simple cases that we're considering lack an assurance, I'll argue in section 2 that we need the concept of an assurance to understand the normative point of this distinction.

Beyond mere reliance, trust appears to involve optimism about the trusted's responsiveness to a subset of one's needs—not optimism merely that the trusted will do what one trusts her to do. But why should that be so?³ Is trust a moral relation, part of which involves appropriately just recognition or acknowledgement of the trusting's vulnerability to harm? It seems not. Andrew need not be worried about any moral dimension of Bernice's failure to recognize or acknowledge his needs. As I'll explain in section 3, he may be worried about a dimension of her performance: that she may *fail* to disappoint his expectations of performance if such disappointment is what it takes, given an unforeseen shift in his needs, for her not to betray his trust. In general, one may rationally fail to trust where one doesn't expect the trust to be disappointed because the trusted

³ Following Jones (2004), I don't think this appearance actually reveals a necessary condition on trust. But the appearance is plausible (for theories of trust that emphasize it, see Baier (1994) and Jones (1996)), and a theory of trust should explain that plausibility.

may need to disappoint one's trust in order not to betray it. We're now considering undisappointed yet betrayed trust, but in section 3 we'll turn to unbetrayed yet disappointed trust. This complexity will reveal how a trust relation can amount to a rational relation that unfolds through time, thus paralleling a core dimension of intrapersonal trust. Before we explore that complexity, we need to see how interpersonal trust relations more generally amount to rational relations. As I'll now argue, they do so by serving as a medium whereby the trusted can make reasons available to the trusting.

2. Trust as the Medium of Interpersonal Reason-Giving

On an assurance view of trust, as I conceive it, the distinction between trust and other forms of reliance mirrors Paul Grice's distinction between natural and non-natural meaning. Grice drew that distinction by contrasting an evidentiary 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 (though she may well expect or hope that there is such evidence) or in any other evidentiary basis (though she may well believe that such evidence is available) but specifically in the addressee's recognition of her intention to give him this reason. I'm not sure that Grice's approach can yield a plausible account of meaning, but I believe it provides the orienting insight for a powerful theory of the dimension of human sociality that we mark with the word 'trust'. 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 evidentiary mechanism but through a structure of mutual recognition and understanding. What the assurance theory adds to this Gricean approach is the further claim that such cases reveal how trust differs from mere reliance.

I focus not on speech acts in general but on propositional assurances: paradigmatically, testifying, advising, and promising. Observing, with Grice, that such assurances aim to give a reason simply through the addressee's recognition of that aim amounts to observing that the speaker is inviting the addressee's trust. Such an assurance is an invitation to trust in this respect: S invites A to regard himself as having a reason to act or believe grounded, in part, in how S undertakes an obligation in issuing the invitation. In such a case, reliance is mere reliance unless it is thus invited by a propositional assurance; thus invited, it can then serve as a basis of obligation, on the side of the speaker, and practical or epistemic reasons, on the side of the addressee. The reasons in question are grounded not in the trust relation itself but in the speaker's status as relevantly reliable. Adapting Grice's terminology, we may say that while both 'natural' and 'non-natural' reason-giving depend on the speaker's reliability, there is a crucial difference in how they depend on reliability. In light of this difference, there is also a further difference in what counts as relevant reliability. I'll explain these differences in turn, thereby explaining what it is for an assurance to 'invite' the addressee's trust.

How could reliability ground a reason non-evidentially? First consider how evidence of reliability could ground a reason, as in this new case:

Guide's Assertion. Aaron knows that Bonnie is a trained wilderness guide, but he also knows that she despises him and would never help him find the campsite where necessary food and shelter await them. Fortunately, he doesn't have to ask her: he spies Bonnie ahead on the trail below him, and training his binoculars on her mouth he lip-reads her as muttering under her breath that the campsite is to the north, via the rightmost fork of the trail. Believing her reliable about the location of the campsite, he comes to believe what she thereby asserts. And he is correct in both beliefs: she is reliable, and her assertion is true.

Here, reliance on another's assertion does not amount to trust. How might we get an instance of trust? Consider this emendation:

Guide's Testimony. As in *Guide's Assertion*, except that Bonnie now addresses Aaron, looking him in the eye and telling him that the trail is to the north, and Aaron trustingly believes her.

The same contrast emerges when we shift from testimony about the campsite's location to advice about how to reach it. First, the mere reliance case:

Guide's Practical Assertion. As in *Guide's Assertion*, except that Aaron lip-reads an assertion not about the campsite's location but about where Aaron ought to go: 'That lost-looking guy ought to keep heading north.'

And now the trust case:

Guide's Advice. As in *Guide's Testimony*, except that Bonnie makes the assertion in *Guide's Practical Assertion*, now addressed to Aaron.

Further modifications yield a parallel contrast for promising. First, a statement of intention that does not amount to a promise:

Guide's Assertion of Intention. As in *Guide's Assertion*, except that Aaron lip-reads an assertion of intention: 'I've got to keep heading north, in order to get back to the campsite before dark.'

And now the promise:

Guide's Promise. As in *Guide's Testimony*, except that instead of testifying to the whereabouts of the campsite, Bonnie promises to take Aaron there.

In each of these three pairs of case, we have a contrast between the assurance and a mere assertion. In each pair, Aaron can get reasons from the assurance and also from the assertion. He can get an epistemic reason to believe that the campsite is to the north in both *Guide's Assertion* and *Guide's Testimony*. He can get a practical reason to go north in both *Guide's Practical Assertion* and *Guide's Advice*. And he can get planning

reasons—reasons to do or to plan to do things premised on Bonnie’s leading him to the campsite—in both *Guide’s Assertion of Intention* and *Guide’s Promise*. That Aaron can get these reasons does not distinguish the cases. What distinguishes them is *how* he gets the reasons.

We can understand what’s distinctive of each species of assurance—testimony, advice, promise—by contrasting the mechanism whereby it makes reasons available with the evidential mechanism whereby reasons are made available in the mere *Assertion* cases. One concise way to draw the contrast is to observe that each *Assertion* case can be quasi-Gettierized: in each, it is possible for Bonnie to be relevantly reliable, and for her assertion to be true, though Aaron’s reliance on her does not track this reliability and therefore does not track the truth of the belief that he acquires by this mechanism. Here is one way that could happen:

Quasi-Gettierized Assertion. As in *Guide’s Assertion*, with two new bits: (i) Aaron believes Bonnie reliable because he believes that wilderness guides are trained to ‘sniff out’ campsites by the scent of their campfire, and (ii) as it happens Bonnie has a sinus infection, though she nonetheless remains reliable. She cannot ‘sniff out’ anything in her condition, but she does reliably judge the direction of the campsite by visual acuity, drawing on a skill that she developed in her former career as a fire-spotter for the forest service. Aaron has no idea that she was ever in the forest service, or that she has such visual skills. It is thus mere luck that his belief that she is reliable turns out to be true. This luck ensures that he fails to derive from her assertion any actual reason—as opposed to the mere appearance of a reason—to believe what she asserts.

Since Aaron relies on Bonnie’s assertion in order to form a belief, the case amounts to a Gettier case of the classic sort: Aaron forms a justified true belief that fails to count as knowledge because his justification fails to be related in the right way to what makes the belief true. If Bonnie had been ‘sniffing out’ the campsite, as Aaron believes she is, then his evidence for the reliability of her ability to ‘sniff out’ destinations would justify his belief in a way that is appropriately related to what makes the belief true. But though she is normally thus reliable—and therefore his belief is justified—what makes his belief true runs through a different causal mechanism, which intuitively undermines any claim he might make to knowledge.⁴

It is irrelevant to the case’s status as a classic Gettier case that Bonnie happens to be reliable in a way that does not engage Aaron’s belief-forming mechanism. But it serves my purposes to distinguish a category of *quasi*-Gettier cases with this structure: A does not have a particular reason that he takes himself to have on the basis of B’s assertion because, though he has good evidence that B is relevantly reliable, and B *is* relevantly

⁴ This appeal to reliability is not the appeal to reliability characteristic of reliabilism. The reliabilist appeals to the reliability of Aaron’s—not Bonnie’s—belief-forming process. In *Quasi-Gettierized Assertion*, Aaron’s belief-forming process is clearly not reliable.

reliable, A's evidence does not track B's reliability. In *Quasi-Gettierized Assertion*, this reason is a reason to believe that the campsite is to the north, but we could produce quasi-Gettierized versions of the two other *Assertion* cases as well, wherein Aaron would (i) fail to have a reason to act in accordance with Bonnie's practical assertion that he ought to head north, and (ii) fail to have reasons to do or plan to do things premised on the assumption that Bonnie is leading him to the campsite. As we originally imagined the cases, we assumed that Aaron's evidential relation to Bonnie's reliability gives him these reasons. But when we quasi-Gettierize the cases, we see that he fails to have the reasons: she is reliable in a way that vindicates his reliance on her—he is relying on her to guide him to the campsite, by reliably asserting the truth about its location, or about what he ought to do, or about the state of her own intentions—but not in a way that would give him the reason that he takes himself to have.

It is distinctive of the *Assurance* cases, by contrast, that they cannot be thus quasi-Gettierized. When Bonnie turns to address Aaron—assuring him that the campsite is north, or that he ought to head north, or that she will lead him there—we lose grip on the idea that Bonnie could be reliable in a way that would vindicate Aaron's reliance on her, and yet not in a way that would provide the reason that he presumes she gives him. Of course, he could fail to have that reason, but only because she is not reliable in a way that would vindicate his reliance on her. That possibility is present in quasi-Gettierized cases as well: Bonnie may simply prove an unreliable guide to the campsite. When we shift from mere assertion to addressed assurance, we're not talking about a breakdown in her status as a reliable guide but in her status as a reliable interlocutor. The form of address yields the crucial difference that Aaron is now relying on her not merely to lead him to the campsite but to pull her weight in an interlocutory relation. Because of this difference in the nature of Aaron's reliance, the reason that he gets from Bonnie if she is reliable must not be the same sort of reason as he got in the original *Assertion* cases. Of course, in each case Bonnie's assurance includes an assertion, so it may be that Aaron gets both sorts of reason. But they would nonetheless remain different sorts of reason.

What other sort of reason could he get? I have characterized propositional assurances as 'invitations' addressed to your interlocutor—'invitations' that he or she should trust or rely on you in specific ways. It's time to cash in that metaphor. When you give a propositional assurance, you invite your addressee to rely on you as a source of reasons, and you present yourself as worthy of precisely that species of reliance. How would you be thus worthy? What grounds the reason, I'll argue, is your undertaking an obligation to do justice to your addressee's needs in respects relevant to the understanding at the core of the trust relation that you invite. Interpersonal trust can be betrayed, rather than merely disappointed, because of the normative role played by this shared understanding. When your addressee accepts your invitation to trust, the now-shared understanding at the core of this trust relation comes with important normative consequences: if you uphold your end of the trust relation by being relevantly trustworthy, your addressee gets a reason (to believe through your testimony, to act through your advice, or to plan through your promise); 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 (we'll inquire into what such appropriateness comes to), you count as undertaking an obligation to be thus trustworthy. One can betray a trust relation because one can betray the shared understanding at its core—a normative structure with attendant obligations and entitlements.

Let me schematize this core strand in my argument. To understand the possibility of betrayal, I'll argue, we must understand how the trust relation binds trusting and trusted into something like a contract. My first thesis is that the basis of this contract-like institution lies in the shared understanding of the point of the trust relation—that is, of what the trust relation can do. My second thesis is that the principal thing that an interpersonal trust relation can do is this: provide the trusting with a distinctive source of reasons. In purporting to provide the trusting with such reasons, the trusted undertakes an obligation to be trustworthy in ways that would ground the reasons—that is, to give them their rational force. As we'll see when we get to promising, the obligation may also be moral—but it is more fundamentally rational. (The same applies to obligations of truth-telling in testimony and advice.) I'll explain the structure of this normative relation more precisely in section 3. That will motivate shifting our attention, in section 4, to the parallel structures that inform intrapersonal trust.

3. Trust Disappointed yet Unbetrayed

In section 1, we considered cases of undisappointed yet betrayed trust. Now that we've considered how propositional assurances can provide reasons through the medium of trust, we're ready to consider cases wherein trust is disappointed yet unbetrayed. I'll begin by explaining why we cannot get such cases without assurance. This will reveal why only an account that stresses assurances can explain what it is for trust to be betrayed.

Consider again *Lovely Leaf*, in which Andrew trusts Bernice to care for a leaf that strikes his fancy. We saw how Andrew's trust could be undisappointed yet betrayed. Could his trust be disappointed yet unbetrayed? No. If Bernice disappoints his trust by failing to produce the result—a leaf kept safe—that Andrew is relying on her to produce, she betrays his trust. To say that she betrays his trust is not to say that she intended to betray it or that she betrays it with anything but the best intentions. She may make her best effort not to betray his trust, yet betray it anyway. As we saw, Andrew's trust in Bernice is not mere reliance on her to keep the leaf safe. She may keep the leaf safe yet betray his trust, if she does not manifest appropriate concern for his needs. (Again, she need not feel concern; she need merely be disposed to act as if she is concerned.) So the manifestation of concern constitutes a necessary condition on unbetrayed trust. But it does not constitute a sufficient condition; at least, it does not in simple cases like *Lovely Leaf*. In such simple cases, leaving the trust undisappointed is also a necessary condition on the trust's counting as unbetrayed.

What of testimonial or advisorial cases? Here too we get this structure: unbetrusted trust is not merely undisappointed trust. Here too we need to add a condition of concern: the testifier must manifest appropriate responsiveness to the trusting's context-sensitive epistemic needs, and the advisor must manifest appropriate responsiveness to the advisee's context-sensitive practical needs. But merely manifesting this concern does not suffice for the trust to go unbetrusted. As in *Lovely Leaf*, the trusted must actually produce the result that vindicates the trust—in the testimonial case by asserting the truth, and in the advisorial case by steering the advisee towards what the advisee actually has reason to do. As in *Lovely Leaf*, two conditions are individually necessary and jointly sufficient for unbetrusted trust: (i) the trusted must actually produce the result that the trusting is relying on her to produce, and (ii) the trusted must do so in a way that manifests appropriate responsiveness to a subset of the trusting's needs: specifically, those needs recognition and acknowledgement of which inform the agreement at the core of the trust relation.

Can there be a case in which the first condition drops out? I'll now argue that promissory trust has that structure. Unlike the other forms of trust that we've considered, in promissory trust, the two conditions do not come apart: as we might call them, the condition of execution does not come apart from the condition of concern. What explains this difference, I'll argue, is the distinctive way in which a promise is a propositional assurance: without this distinctive element, the trust relation cannot have this normative structure. In section 4, I'll argue that promises share this structure with the intrapersonal assurance at the core of an intention. What gives promises and intentions their special normative structure is their distinctive relation to time. When you form an intention, you're aiming to do justice to your ongoing needs in respects relevant to your understanding of the point of the intention. If your needs change in certain ways as a result of changing circumstances, your rational obligation to do what you intend to do may merely lapse, rather than being outweighed or overridden. You may thus abandon the intention because your obligation to follow through has lapsed; as we'll see, it would be a mistake to say that the obligation lapses only because you have abandoned the intention. And if your promisee's needs change as a result of changing circumstances—I don't mean just any needs, but specifically those that gave point to the promise—then your obligation to do what you promised lapses, rather than being outweighed or overridden. These are *planning* needs and thus are partly constituted by the promisee's actual plans and dispositions to plan: they give point to the promise insofar as they explain why the promisee is relying on the promise. In this respect, promises and intentions share their normative structure. Though a promissory obligation is also typically a moral obligation, the parallel with intention reveals that it is more fundamentally a rational obligation. We can grasp the nature of these obligations, I'll argue, by understanding how the trust relations at the core of both promises and intentions can be disappointed yet unbetrusted.

Let me introduce my approach to promising with some new cases. Though it would be simpler to use *Guide's Promise* as a template, all of the *Guide* cases in section 2 are

somewhat cartoonish, and the conclusion for which I'm now arguing emerges most clearly in cases that reflect the human complexity of everyday life. If our lives did not contain such human complexity, the concept of a promise might have likewise lacked the complexity that gives it the element I'm emphasizing.

Two Promises. Ben is serving on a search committee with Andrea and Andres, and Ben has promised each colleague to read the remaining dossiers today. Andrea and Andres are depending on the promise; each is planning in concrete ways (for example, drafting paperwork for the dean) that depend on Ben's acting on this promise. But as Ben begins to read the dossiers, he gets a call from the daycare: his child is sick, and since his spouse has firmer commitments today, he'll have to fetch the child and nurse him at home—a full-time job. So he cannot act on his promise. Should he apologize? Well, of course he should—to Andres. As it happens, Andrea is Ben's spouse, and it does not appear to make sense for Ben to apologize to his spouse in these circumstances. Ben's sick child is Andrea's sick child; if Ben doesn't nurse him, Andrea will have to, thereby thwarting many of her planning needs, including the needs that informed Ben's promise.

How might we characterize the difference in Ben's predicaments vis-à-vis his two promisees? Ben's predicament vis-à-vis Andres appears to be one in which violating a promissory obligation is morally permissible in the light of broader considerations (and Ben may even be, in that light, morally required to violate it). Ben's promissory obligation does not simply lapse here. But Ben's predicament vis-à-vis Andrea appears different in just that respect: his promissory obligation appears to be cancelled, not overridden or outweighed, by his obligation to care for his child. The difference seems to lie in how Andrea herself participates in that normative circumstance insofar as she shares custody of the child. And that difference explains the difference in whom Ben owes an apology to. An apology to Andrea would reveal confusion about how that normative circumstance structures his promissory obligation. But there could be no such confusion vis-à-vis Andres. (I'll presently justify these intuitions about apology.)

Does the difference merely reflect Ben's greater intimacy with his spouse? We can easily imagine the cases reversed. Say Ben's non-spousal colleague, with whom Ben is not even friends, undergoes the shift in practical needs, while Ben's spouse's needs remain constant in relevant respects. We could fill in the details as follows:

Role Reversal. Ben is looking after the child of his non-spousal colleague, Andres, while Andres teaches a class. (If that seems to presuppose intimacy with Andres—or a second, competing promise—we could say that Andres's child is old enough to be left unattended in Andres's office, and Ben merely happens to be the only other person on the hall.) As before, Ben has promised Andres to read the dossiers during this time but must suddenly whisk the child off to urgent care instead. When Andres gets Ben's voicemail and calls him back, ought Ben to apologize for having failed to read the dossiers? Of course not. Doing so would show that something has gone

wrong in Ben's relations with this colleague. But now Ben's spousal colleague, Andrea, whom Ben has also promised to read the dossiers, phones to ask if Ben has yet done so. Though Andrea understands Ben's predicament, Ben should nonetheless apologize.

The difference that emerged in *Two Promises* thus does not depend on the question of intimacy.

Does the difference rest on some norm of professional or workplace ethic? We can easily construct parallel cases outside the workplace. Imagine Ben has promised his spouse and his neighbour to plant two trees this afternoon, one in each of the adjoining yards, but a sick child prevents him from executing the promises. To whom does Ben owe an apology? Without unusual background assumptions, we'll say that Ben owes his neighbour an apology but not his spouse. (Make it his neighbour's child to address the objection from spousal intimacy.) No professional or workplace norm explains this difference.

At a methodological level, one might object that I have framed my argument in terms of intuitions about when one person owes another an apology. What is the basis of these intuitions? Their basis lies, I'll argue, in the normative relation between promisor and promisee, and specifically in the nature of the authority that the promisee exercises in holding the promisor to the promise. I'll first say what I believe that authority amounts to. Then, in section 4, I'll use the analogy between promising and intending to explain the basis of the authority.

What grounds the intuition that Ben does not owe Andrea an apology when his crucial care for their child prevents him from executing his collegial promise to read those dossiers? The question asks us to consider the nature of Andrea's promissory right to demand performance from him. In a typical case with this structure, the promisee will let this promisor off the promissory 'hook', if not in prospect then in retrospect; even if through distraction or forgetfulness she doesn't, it will be clear all along that if she had attended to the circumstances she would have done so. But imagine that as it happens Andrea doesn't let Ben off the hook, and moreover that she wouldn't even if fully aware of the circumstances. Must it be implausible to think that she nonetheless ought to let him off the hook? Must it be implausible to think that she ought to let him off the hook by the terms of that very promissory agreement? There is no general reason why this must be implausible. If professional anxiety gets the better of Andrea and she holds her spouse to his collegial promise—overlooking that they share custody of this child and that the failure to attend to the child's needs would have a direct impact on her planning agency—then it seems she is making a mistake. It is possible that she is not making a mistake, but only if she really doesn't care about the child enough to have adopted plans and policies characteristic of a care-giving commitment. What grounds the intuition that Ben doesn't owe Andrea an apology for having failed to execute the promise is that, assuming that she does care about their child in the ways typical of parenthood and that her spouse's failure to execute the promise reflects his success in tracking her parental

planning needs, she simply does not have any promissory right or authority to demand performance from him. Ben doesn't owe her an apology because she has no right to one.

What grounds these intuitions about apology, then, is our grasp of what is at stake in the promissory agreements between Ben and Andrea and between Ben and Andres. Compare that understanding with the understanding that informs the non-promissory trust relation between Bernice and Andrew in *Lovely Leaf*. Part of what it is for Bernice not to have promised in *Lovely Leaf* is for it not to be the case that her obligation has the diachronic complexity that we see illustrated in the promissory trust relation between Ben and Andrea in *Two Promises*. This is a substantial claim about what a promise adds to a sub-promissory trust relation. There is, I claim, a distinctively diachronic dimension to the norm informing a promissory trust relation. Andrew trusts Bernice to do justice to those among his needs that inform the point of his relying on her at the time at which he gets her to acknowledge how he is relying on her. Andrea trusts Ben to do justice to her ongoing needs in respects relevant to the point of his promise. This reference to the 'point' of the relation is in each case reference to the implicit or explicit agreement between them. In a sub-promissory agreement, the point of the relation lacks the diachronic complexity characteristic of the point of a promissory agreement. As I'll now explain, this difference derives from how these two types of trust relation admit of betrayal.

4. Trust as the Medium of Intrapersonal Rational Coherence

How could a promise serve as a source of planning reasons? I have characterized the promissory assurance as an invitation to trust, but an invitation to trust does not by itself give rise to the reasons that it represents itself as making available, nor does the trust itself.⁵ What gives rise to a reason is the trusted's worthiness of this trust. In the testimonial and advisorial cases, this includes reliability in not disappointing the trust, in addition to whatever further condition might codify how the trusted will not betray the trust. But in the promissory case, as we've seen, it is a mistake to posit an anti-disappointment condition alongside the anti-betrayal condition. One might worry that this disanalogy undermines my claim that promissory trustworthiness can serve as a source of planning reasons. If trustworthiness does not include reliability simply in doing what one has promised to do, how could it serve as the basis of the promisee's planning reason?

We can assuage the worry by articulating an analogy between the interpersonal reason-givingness of a promise and the intrapersonal reason-givingness of an intention, which shares the just-noted feature of promising that distinguishes it from testimony or advice as a source of reasons. Though no developmental claim is crucial to my assurance view of trust, we can vividly pose the analogy by asking how we learn to treat

⁵ On whether trust can itself ground reasons, see Faulkner (2011). For objections to the idea, see Hinchman (2012). For replies to those objections, see Faulkner (2012).

a promise as a source of planning reasons. It is not an implausible hypothesis that we learn how promissory trustworthiness provides a source of reasons by reflecting on how the promissory dynamic between S and A resembles an intrapersonal dynamic that might unfold within A himself—between the earlier self A_1 , that is A insofar as he forms and retains an intention, and the later self A_p , that is A insofar as he follows through by planning and acting on that intention.⁶ From this angle, S's promise-based obligation to A revealingly resembles A_1 's intention-based obligation to A_p . The parallel most fundamentally lies in a parallel between the interpersonal agreement or 'understanding' at the core of a promise and the shared intrapersonal understanding—shared as between earlier and later selves—at the core of an intention.

The fundamental parallel lies here: both promises and intentions generate obligations that are distinctively sensitive to requirements on responsible redeliberation—that is, on redeliberation whether to do what you've committed yourself to do that manifests an appropriately responsible attitude towards your obligation to do it. We considered the interpersonal dimension of this issue in section 3, as we observed how the promisee's ongoing needs engage both parties' expectations of the promisor's performance. In the intrapersonal dimension, the question of expectation appears to run in parallel: you are rationally required to redeliberate if the future does not unfold as you relevantly expected it would when you formed the intention—either by raising relevant new considerations that you did not consider when you deliberated or by revealing relevant problems with how you considered what you considered in that deliberation. Why do we need these appeals to relevance? We need the appeals to relevance because not every falsification of your expectations affects the rationality of following through on an intention—a point that obviously applies to the promissory case as well.

We can illustrate with the following case and its several variations. Imagine you intend to have a picnic this afternoon at your local park, and consider the bearing of several unexpected developments on the rationality of following through on your intention. As you're preparing to leave for the picnic, (i) dark storm clouds suddenly loom on the horizon, (ii) you can't find your favourite picnic blanket, (iii) you can't get your hair to look right, and (iv) you notice that your neighbour has acquired a fourth poodle. Let's say that development (iv) is the most surprising among these, since surely—you'd naturally assumed—three poodles is quite enough for any single pet owner. Do you therefore reconsider the intention to picnic? Obviously not, since the fourth poodle has nothing to do with your plans.⁷ What then of development (iii)? Well, that probably doesn't matter either, though it's possible your picnic plans require that you be photogenic in ways undermined by the problem you're having with your hair. The missing picnic blanket is probably relevant. Though most picnickers would treat that merely as a matter of means, some might regard the disappointment as rele-

⁶ A more controversial hypothesis would view A's grasp on intention as developmentally derived from his ability to enter into proto-promissory relations with caregivers.

⁷ Change the poodles to black cats, and some might superstitiously deem the development relevant to their picnicking plans.

vant to the plan to picnic itself—in which camp are you? Nearly all picnic planners would regard unexpected storm clouds as requiring a redeliberation whether to picnic. But even here there may be exceptions—it all depends on how important the picnic is to you and on how well you imagine you'll be able to cope with unexpected weather. Perhaps your picnic is a reunion of Navy SEALs who would welcome the challenge.

The case reveals how much we take for granted in intending. We distinguish relevant from irrelevant expectations—that is, expectations whose falsification should lead us to redeliberate from expectations whose falsification is irrelevant to redeliberation—without formulating the distinction in terms of explicit conditions on the intention. We nonetheless have an implicit understanding of what we're up to in intending, and that understanding—shared between intending and acting self—in turn determines how the distinction applies. As we've seen, there is a parallel role for shared understanding—specifically for the understanding informing the promissory agreement between promisor and promisee—in distinguishing circumstances in which the promisor should do what she has promised to do from circumstances in which her doing so would reveal that she has misunderstood the implicit point of the promise.

At the core of the parallel lies a distinctively temporal norm: in each case, the one relied on—you *qua* intender or *S qua* promiser—bears responsibility to keep track of and do justice to needs that arise partly from how the one relying has planned on the basis of that reliance. The parallel focuses our attention on one key respect in which the two acts look thus forward to future need. When A intends to ϕ at t , A typically gives himself a reason to expect that he will ϕ at t , an expectation that in turn enables him to plan what to do in the meantime. This reason is grounded in an aspect of A's reliability in intending to ϕ at t , where such reliability in intending is very different from reliability in a mere disposition to ϕ at t . And when S promises A that she will ϕ at t , S's reliability in promising gives A a reason to expect that she will ϕ at t , similarly useful in planning. The key to understanding what's distinctive of an intention, as opposed to a mere disposition, and of a promise, as opposed to a mere statement of intention, lies in understanding how temporality informs these distinctive species of reliability.

To clarify this aspect of intrapersonal rationality, we must understand how it manifests a dynamic of self-trust and self-mistrust. What if you are untrustworthy in forming an intention? Then, by my account thus far, you do not get any planning reasons from the intention—so it's useless as an intention. But wait: that's *your* intention! So what should you do? Insofar as you do have this intention, you should follow through on it. But insofar as the intention is not worthy of your trust, it appears that you should not follow through on it. Something has to give. Here's my solution: you have a narrow-scope rational obligation to follow through on your intention but only a wide-scope rational obligation to follow through on the practical judgement that informs it. Yes, insofar as you intend, you ought to follow through on the intention. But insofar as you are not trustworthy in making the practical judgement that informs the intention, you ought not to retain the intention. This complexity in the structure of intrapersonal rational coherence arises not from the rational nexus between an intention and its

execution but from the rational nexus between a practical judgement and the choice or intention wherein you commit to that judgement.

The key question is this: how could you manage to ‘mistrust’ your own intention? When you mistrust a promise, the promise does not thereby disappear: the promisor continues to count as promising, but you mistrust that promise. When you ‘mistrust’ an intention, by contrast, it seems you thereby no longer have that intention. So the question of your trustworthiness—the question whether to trust yourself—poses this more perplexing question: how *could* you manage to mistrust yourself in this respect? If we are to make sense of the analogy with promising, when you mistrust yourself at a time, *t*, your mistrust must target something true of you at *t*—in the way that a promisee’s mistrust targets the promisor’s promise. But if you cannot simply mistrust your own intention—that is, yourself insofar as you have this intention—how can you mistrust yourself? Though I lack space for a full treatment, I’ll sketch an answer in section 5.⁸ I’ll thus explain what it is for intrapersonal trust to be betrayed without thereby being disappointed.

5. Self-Trust Undisappointed yet Betrayed

In section 4 we considered how the self-trust informing an intention could be disappointed yet unbetrayed. It remains to consider how such self-trust could be undisappointed yet betrayed. As before, the answer lies in seeing how the intrapersonal relations at the core of an intention run in normative parallel with the interpersonal relations at the core of a promise. Each normative relation generates these two possibilities. Just as a promisor may disappoint your trust without thereby betraying it, so a promisor may fail to disappoint your trust—that is, may follow through on the promise—while thereby betraying your trust by violating the promissory agreement that informs it. And just as you may ‘disappoint’ the self-trust in your own practical judgement that forms the core of an intention without thereby betraying that self-trust—because you abandoned the intention in rational responsiveness to an unexpected change in your circumstances—so you may fail to ‘disappoint’ that self-trust, by following through on the intention, while thereby betraying it. As in the promissory case, you can betray your own self-trust in following through on an intention by failing to remain true to the intrapersonal agreement that informs the intention.

On the proposal I’ll now sketch, your intention has two elements: a practical judgement, and your trusting commitment to that judgement. When, in a slight revision of Donald Davidson’s (1970: 30) example, you intend to go to sleep without brushing your teeth tonight, you judge, having considered matters to your satisfaction, that you ought to skip brushing, and you commit yourself to that judgement in the diachronically action-guiding way of intention (for example, altering your sink-side preparations). Why distinguish these two elements? Imagine, as Davidson does, that you

⁸ The sketch summarizes the account developed in Hinchman (2009; 2010; and 2013).

cannot bring yourself to act on your intention: at some point before going to sleep, you come to mistrust yourself insofar as you have the intention, and the self-mistrust prevents you from acting on it. Do you thereby mistrust the intention itself? As we've begun to see, that doesn't quite make sense. Though you may have worries or reservations about an intention that you persist in holding, if you simply 'mistrust' your intention, in the way that you might simply mistrust another's promise, you thereby cease to hold the intention. An intention manifests trust in your judgement. While there may be special contexts or respects in which you can trust yourself in some substantial way and yet at the same time count as mistrusting that trust, trust in your judgement does not appear to admit of that possibility. You cannot trust your practical judgement 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 judgement. One way to abandon an intention is to abandon the judgement that informs it, but that is not the only way. You might abandon an intention by mistrusting the judgement that informs it. Our question is therefore how this works: how might you mistrust a judgement that you nonetheless retain?

What is a practical judgement? How does a practical judgement guide the formation of an intention? Let's work with the influential account of judgement pioneered by T. M. Scanlon (1998: 25–30; 2007), which assimilates practical commitment to a species of doxastic commitment, equating your all-things-considered practical judgement that you ought to ϕ with a doxastic judgement that you have conclusive reason to ϕ .⁹ The account identifies the specifically practical element in a practical judgement with an element in the content of that judgement: the idea that you have a conclusive practical reason. One might wonder what it is for you to judge that you have a conclusive practical reason, but I'll take that notion for granted. As we've now seen, when you mistrust yourself insofar as you intend to ϕ , you specifically mistrust, not the intention itself, but the practical judgement that informs it. Adopting the Scanlonian view, we'll say that you mistrust your judgement that you have conclusive reason to ϕ .

What is it to mistrust this judgement? For reasons of space, I must state my proposal abstractly.¹⁰ When I speak of 'mistrusting' your own judgement, I don't mean mistrusting your faculty of judgement. If 'mistrusting your judgement' could only mean mistrusting your faculty of judgement, then when you mistrust your judgement you'd be—deliberatively speaking—just stuck. You'd have to stop deliberating and merely wait for the bout of self-mistrust to pass. But we don't think you're just stuck when you mistrust your judgement; we think you can mistrust your judgement in this or that respect and resume deliberating by trusting your judgement in other respects. We individuate these 'respects' with propositions. While self-mistrust typically targets a subject matter, we individuate subject matters with propositions, and a subject matter can be so narrow that it coincides with a single proposition. To say that you mistrust

⁹ I give grounds for rejecting Scanlon's view of practical judgement in Hinchman (2013: Section VI), but there is no harm in assuming it for present purposes.

¹⁰ Again, for more details, see the works cited in note 8.

your judgement in the respect individuated by the proposition *p* is thus to say that you mistrust your judgement 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 judgement on the question whether *p*, it does not follow that you are redeliberating whether *p*. You might be deliberating whether to redeliberate whether *p*. We can put the point like this: to reconsider your judgement that *p* is not yet to reconsider whether *p*. To mistrustfully reconsider your judgement that *p* is to wonder—not idly but with appropriate engagement—whether to reconsider whether *p*.

Why might you raise the issue, wondering whether to reconsider whether *p*? One occasion for raising the issue arises when circumstances have unexpectedly changed since you formed the intention and you wonder whether those changes warrant reconsidering it. Now say you nonetheless persist in your intention without reconsideration, in a way that violates the intrapersonal point of the intention. You form an intention to hike up a mountain trail as a way of enjoying the beautiful weather, but you do not reconsider the intention as the sunshine gives way to a downpour, since you have lapsed unthinkingly into goal-directed determination to reach the peak. By your own lights you ought to have reconsidered your intention—perhaps reaffirming it, with a new aim of reaching your goal, perhaps abandoning it in accordance with the original point of the intention—but you do not reconsider it, rather like the promisor who does what she promised to do despite an unexpected change in the promisee's relevant needs. You don't decide to hike up that mountain in a downpour; you merely persist in your intention despite this change in your circumstances. But the change, given your understanding of the point of your hike, makes your persisting in the intention a violation of the trust that informs it. Given that understanding, you ought to have reconsidered whether to continue hiking up that mountain as soon as the weather changed. You didn't reconsider, and you still aren't reconsidering; it seems that your intention has somehow 'got the better of you.' (Are you afraid of looking like a 'quitter'? Are you so focused on your train of thought that you haven't noticed that you're drenched and shivering?) In thus 'getting the better of you' it betrays your self-trust.

So what should you do? Even if you continue to judge, without any further reflection, that you ought to hike that trail, at the very least you ought to consider whether to reconsider, thereby withdrawing your trust in that judgement. Reasoning 'upstream' in this way¹¹—should you abandon your judgement because you mistrust it?—manifests responsiveness to the possibility of betrayed self-trust. Such self-mistrust does not, of course, reveal betrayed self-trust, since it is possible that you do care appropriately about what is at stake for you in your deliberative context and that your self-mistrust is therefore mistaken. But it is also possible that your self-mistrust is not mistaken: it is possible that you really have betrayed the invited self-trust relation.

¹¹ For this metaphor, see Kolodny (2005), though Kolodny argues against the possibility of such reasoning (534–9). I reply to Kolodny's specific objections in Hinchman (2013).

The point runs in parallel with the possibility of betrayal in interpersonal trust. If you mistrust a promisor, your mistrust does not itself reveal that the promisor has betrayed the trust that she invites in promising. But your mistrust does manifest responsiveness to the possibility that she has betrayed your trust. In each case, interpersonal and intrapersonal, the responsiveness at the core of trust is a *rational* responsiveness because it targets the possibility that your trust has been betrayed. Responsiveness to betrayal and responsiveness to reasons or rational requirements thus go hand in hand. Trust is crucially unlike other forms of reliance, whether interpersonal or intrapersonal, because of this broad but basic link between trust and rationality.

References

- Baier, Annette (1994) *Moral Prejudices* (Cambridge, MA: Harvard University Press).
- Davidson, Donald (1970) How is Weakness of Will Possible? reprinted in his *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).
- Faulkner, Paul (2011) *Knowledge on Trust* (Oxford: Oxford University Press).
- Faulkner, Paul (2012) Replies, *Abstracta* VI, 117–37.
- Hardin, Russell (2002) *Trust and Trustworthiness* (New York: Russell Sage Foundation).
- Hawley, Katherine (2014) Trust, Distrust, and Commitment, *Noûs* 48:1, 1–20.
- Hieronymi, Pamela (2008) The Reasons of Trust, *Australasian Journal of Philosophy* 86:2, 213–36.
- Hinchman, Edward (2009) Receptivity and the Will, *Noûs* 43:3, 395–427.
- Hinchman, Edward (2010) Conspiracy, Commitment, and the Self, *Ethics* 120:3, 526–56.
- Hinchman, Edward (2012) Can Trust Itself Ground a Reason to Believe the Trusted? *Abstracta* VI, 47–83.
- Hinchman, Edward (2013) Rational Requirements and ‘Rational’ Akrasia, *Philosophical Studies* 166:3, 529–52.
- Holton, Richard (1994) Deciding to Trust, Coming to Believe, *Australasian Journal of Philosophy* 72:1, 63–76.
- Jones, Karen (1996) Trust as an Affective Attitude, *Ethics* 107:1, 4–25.
- Jones, Karen (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.
- McGeer, Victoria (2008) Trust, Hope, and Empowerment, *Australasian Journal of Philosophy* 86:2, 237–54.
- McMyler, Benjamin (2011) *Testimony, Trust, and Authority* (Oxford: Oxford University Press).
- Nickel, Philip (2007) Trust and Obligation-Ascription, *Ethical Theory and Moral Practice* 10:3, 309–19.
- Rose, David (2011) *The Moral Foundation of Economic Behavior* (Oxford: Oxford University Press).
- Scanlon, T. M. (1998) *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).
- Scanlon, T. M. (2007) Structural Irrationality, in G. Brennan, R. Goodin, F. Jackson, and M. Smith (eds), *Common Minds* (Oxford: Oxford University Press), 84–103.
- Walker, Margaret Urban (2006) *Moral Repair* (Cambridge: Cambridge University Press).