

TRUST-BASED THEORIES OF PROMISING

This paper discusses the prospects of a comprehensive philosophical account of promising that relies centrally on the notion of trust. I lay out the core idea behind the Trust View, showing how it convincingly explains the normative contours and the unique value of our promissory practice. I then sketch three distinct options of how the Trust View can explain the normativity of promises. First, an effect based-view, second, a view drawing on a wider norm demanding respect to those whom one has invited to something, and finally, as a new suggestion, a Normative Interest View. This view holds that promising is a normative power that serves our interest in facilitating or enabling the relationship of trust between promisor and promisee. I argue that only those embracing the third view can fully account for the distinctive obligation that results from the giving of a valid promise in all cases.

Keywords: Promises, Trust, Assurance, Normative Powers, Normative Interests

This is the submitted manuscript version of the paper. To cite, please refer to the published version: <https://doi.org/10.1093/pq/pqz086>

1) Introduction

When attempting to explain the specific role that promises play in our lives, it comes naturally to refer to the notion of trust. One thing that promises do, or so it seems, is help us trust each other in situations where we otherwise may not be able to. It was only recently, however, that some philosophers have tried to put the connection between promises and trust truly at centre-stage of a comprehensive theory of promises and promissory obligation. The central idea is that promises are, or at least essentially involve, *invitations to trust*, and that their normative force relates in some way to the value of the trust which they bring about or facilitate.¹

¹ The most important contemporary defenders are Daniel Friedrich and Nicholas Southwood (2009, 2011)

In this paper, I will discuss the prospects of this Trust View of promising in closer detail. I will proceed as follows: first, I will briefly outline what precisely the central claim of trust-based theories of promising amounts to (section 2). In the following, I will argue that the Trust View warrants further philosophical attention, as it has considerable explanatory potential. On the one hand, it gives a powerful and accurate explanation of the normative contours of our practice of giving and keeping promises (section 3). On the other hand, it provides resources for a convincing account of what is to many the most interesting feature of promising: the fact that successful promises are normative, and in particular give rise to an *obligation* on part of the promisor to perform the promised action.² In fact, I believe that understanding promises as involving invitations to trust not only offers resources for one, but rather three different kinds accounts of why we have special reasons not to break our promises. The bulk of this paper will be concerned with outlining and evaluating these three options.

First, I will discuss a view focusing on the effects of successfully giving a promise. This account explains the wrong committed by a breach of promise by reference to the wrong of breaking trust one has knowingly and deliberately raised through the act (section 4). The second construal of the Trust View instead locates the wrong of breaking promises within the broader wrong of showing respect to those one has *invited* to something. As the argument goes, there are good reasons to assume that we have to act in accordance with an invitation to trust we have proffered (and that was consequently accepted), even when trust is not

and Thomas Pink (2009).

² As I understand it, OBLIGATION is a normative concept. If I am under an obligation to φ , I have reason to φ . It is furthermore a *pro tanto* notion: it can be that, all things considered, I ought to act contrary to my obligation, if another particularly weighty consideration disfavors the action I am obligated to perform. Obligations differ from 'ordinary' reasons in their directedness, as well as their stringency, in that they silence at least some competing considerations. One way to make sense of this that I am sympathetic to is through the concept of a 'protected reason', as put forward by Raz (1990: 191).

successfully created (section 5). Finally, I will sketch a third theoretical option that has so far not been appreciated in the literature. Trust, or so I shall argue, lends itself very well to the incorporation into a Normative Interest Account of promising along the lines recently made popular by Joseph Raz and David Owens. On such a view, the wrong-base for breaking promises is an independently created directed obligation, brought into being through a normative power to provide the promisee with warrant for trusting the promisor (section 6). I will argue that all three aspects plausibly create reasons to keep one's promises in ordinary circumstances, but only those embracing the Normative Interest Account will be able to fully account for the distinctive obligation that results from the giving of a valid promise in all of the circumstances in which we intuitively judge it to do so. The Trust-Based Normative Interest Account yields a very attractive comprehensive theory of promising, since it allows us to avail ourselves of the explanatory potential that the Trust View holds while also giving an extensionally adequate account of which promises bind.

2) Trust and Invitations to Trust

Nicholas Southwood and Daniel Friedrich, who recently presented a fleshed-out trust-based theory of promising, spell out the central idea of the Trust View by means of an example:

Suppose that Albert has been unfaithful to his wife, Berta. She is deeply hurt. He is full of remorse. He sincerely promises her that he will be unerringly faithful henceforth. According to the Trust View, in promising to Berta that he will be faithful to her, Albert invites her to trust him to be faithful to her. In virtue of inviting her to trust him to be faithful to her, and having the invitation accepted (or at least not rejected), Albert incurs an obligation not to betray the trust that he has invited in Berta. The wrong involved in his breaking his promise, if he chooses to do so, is a matter of his violating this obligation, that is, of his betraying the trust that he has invited. (Friedrich/Southwood 2011: 278)

The core claim of the Trust View is thus a simple idea: In offering a promise, what we are doing essentially involves *inviting someone to trust one to do something*. To understand this proposal and assess its merits, let me first offer some words on the notion of trust in play. For the purposes of the discussion that is to follow, I will try to content myself with a relatively intuitive, philosophically uncontroversial understanding of trust.

The kind of trust that I am interested in takes the form of a three-place relation: A trusts B to do X. I take this type of trust to be an attitude that involves A having a certain kind of faith or optimism in B to pursue a particular course of action, specifically an optimism in B's character and motivations.³ This last qualification is crucial in order to distinguish trust from mere reliance.⁴ Both of these attitudes involve a kind of optimism that the agent in question will act in a specific, expected manner. However, the way in which this optimism is felt is fundamentally different in the cases of reliance and trust – trust is the more demanding and normatively richer attitude of the two. As an inhabitant of Königsberg in the 1780s, I might be able to rely on Kant walking past my window at the same time every morning. Similarly, I can rely on the neighbourhood burglar taking flight when they catch sight of my growling Rottweiler. In neither of these cases, however, my reliance amounts to anything close to trust. When I trust somebody, I not only expect this person to act in a particular way, but I also expect this person to act out of a certain class of motives. To trust somebody to do something⁵, I not only have to assume that they are competent and able to do this thing, I also have to assume that they are motivated in a certain way, a way that

³ See paradigmatically Jones 1996: 1-4.

⁴ See Baier 1986.

⁵ Note that in ordinary language, 'trust' is sometimes used in a wider sense. We talk about trust *that* someone will do something. On this usage, the conveyed meaning appears to be only that one *expects* something (I trust that the sun will shine tomorrow). This sense, sometimes referred to as 'predictive trust' (Faulkner 2007: 880-4) will not be at issue here.

specifically relates to *me*. We can call this the motivational condition of trust.

Motivational Condition: A trusts B to φ only if A takes B to be motivated to φ for reasons that bear on A's interests or projects in some way.⁶

How to best spell out this motivational condition of trust in detail is controversial. Classically, it has been assumed that the motivational condition of trust requires a certain kind of goodwill towards the trusting person on the part of the trustee.⁷ For the purposes of the discussion that is to follow, I can remain agnostic about the precise interpretation and content myself with the minimal version of the condition as laid out. Importantly, the condition as stated does not require A taking B to actually be motivated by A's desires and wants. It is enough that the source of B's motivation is something that A herself cares about (allegiance to a common cause, or a shared set of specific values, perhaps). Thus, I can trust Greta Thunberg to do her best in campaigning for much needed political action to stop climate change, even if she does not know of my existence. Furthermore, let me stress that I only take on board the motivational condition as a necessary condition for trust. Whether or not it forms part of the smallest informative set of jointly sufficient conditions, and thus plays a role in our best definition of trust, is something I will happily leave open. In particular, accepting the motivational condition should be compatible both with more narrowly motivational theories of trust⁸, according to which trust itself is some sort of motivational attitude, as well as with a form of cognitivism about trust.⁹ For both types of theories, the

⁶ Note that to be plausible, the motivational condition must be concerned with a subjective notion of interest. A medieval patient can trust her doctor to drain a sufficient amount of blood from her body, even though the blood-letting is objectively not in her interest. Thanks to an anonymous referee for pressing me to clarify this point.

⁷ Baier 1986, Jones 1996. A different way to spell out the idea is proposed by Hardin, who holds that to trust is to believe that the trustee will 'encapsulate one's interest in hers' (Hardin 2002).

⁸ See paradigmatically Jones 1996, but also Holton 1994 and McGeer 2008.

⁹ See for example Hardin 2002 and Hieronymi 2008.

motivational condition should be uncontroversial as a necessary condition for trust.¹⁰

The fact that trust has a motivational condition is also reflected by our different reactions to the frustration of expectations in the cases of trust and reliance. When we merely rely on someone, we will generally feel disappointment when things do not go as planned. When someone we trust does not come through for us, however, we will feel much stronger negative emotions. As Annette Baier puts it, ‘trusting can be betrayed, or at least let down, and not just disappointed’.¹¹ Trust, unlike reliance, thus requires ‘a vulnerability to betrayal if let down’.¹²

To sum up: On the Trust View, offering a promise essentially involves inviting somebody to trust one to do something. It presents a sort of overture for the promisee to (re-)establish trust, and regain a certain kind of optimism in the promisor’s being motivated to perform the promised action in a way that is essentially connected to the promisee’s goals and values¹³. If this invitation is accepted, this can in turn serve to strengthen the trust relationship as a whole, and lead to a more encompassing kind of interpersonal faith between promisor and promisee. Having laid out the core of the Trust View, I will next move on to examine its explanatory potential, starting with its ability to make sense of the overall shape of our promissory practice.

3) The Trust View and the Shape of our Promissory Practices

One important reason to take the proposal of the Trust View seriously is its ability to explain

¹⁰ Admittedly, there is at least one philosopher who does deny the motivational condition. See Hawley 2014.

¹¹ Baier 1986: 235.

¹² Hieronymi 2008: 215.

¹³ Cf. Friedrich/Southwood 2011: 279.

the distinctive and valuable role that promising plays in our lives. If we understand the promissory practice as essentially being connected with establishing trust, we stand to accurately capture many of the central features of how and why we engage in promissory relations. While I believe there to be more, I will content myself with presenting what I take to be the four most important points in favour of the 'Trust View as a theory of promising.

First and foremost, the 'Trust View can explain how the ability to give promises significantly enriches our social lives. There is no doubt that trust is essential to a number of valuable interpersonal relationships, such as friendships and companionships, as the case of Albert and Berta shows. Furthermore, trust is also highly valuable instrumentally. Many mutually beneficial cooperative endeavours between independent parties with diverging aims would not be possible if the parties were not able to trust one another.¹⁴ More interestingly, however, trust itself can constitute a non-instrumentally valuable form of human connection. Take the following evocative example by Richard Holton:

Suppose we are rock climbing together. I have a choice between taking your hand, or taking the rope. I might think each equally reliable; but I can have a reason for taking your hand that I do not have for taking the rope. In taking your hand, I trust you; in so doing our relationship moves a little further forward. This can itself be something I value. We need not imagine that you would be hurt if I chose the rope over your hand; you might be perfectly understanding of the needs of the neophyte climber. But our relationship would not progress. (Holton 1994: 69)

Holton's case shows how the mere fact that grabbing the hand constitutes an instance of trust can count as a reason to perform that action. Doing so enriches Holton's relationship with his climbing partner. We need not think that he is looking to become close friends with his climbing partner, and that coming to trust him is a necessary step in achieving that aim.

¹⁴ Hume famously made a powerful case for our need for promises to establish 'mutual confidence and security' (Hume 1978: 521).

It's simply better to be trusting climbing-partners than non-trusting climbing partners, even if I can easily avail myself of the assurance trust provides by other means (more ropes). In fact, this seems to be true for almost all domains of life. Imagine a person who makes sure there are metaphorical safety ropes everywhere in her life – she builds a system of safeguards and insurance mechanisms so that she never has to rely on other people in the way that is distinctive of trusting. Intuitively, most would find this life lacking in an important regard. One plausible way to interpret this intuition is by pointing out that we, as inherently social beings, not only value the assurance, but also the particular interpersonal way in which the assurance is provided when we trust someone to do something.

One might worry that the claim that trust and trust-relationships generally have this kind of non-instrumental value leads to bizarre conclusions. Does it mean that we have reason to try and get ourselves to trust as many people as possible, for example by extracting promises about trivialities from them? I think it need not, for at least two reasons. For one thing, we must take into account the potential negative effects of misplaced trust. Since breach of trust is a grave injury, we have reason to be careful about whom we trust to do what.¹⁵ Secondly, it isn't clear that seeking out further trust relationships is something we can do quite as easily as the worry suggests. We cannot bring ourselves to trust just anyone at will – at the very least, the motivational condition of trust will have to be fulfilled. Barring these qualifications, the following appears clearly true: If I am in a situation in which I seek assurance, then, *ceteris paribus*, it is better to receive this assurance by trust rather than other means.

Assuming, then, that trusting can be valuable in the way just laid out, the Trust View

¹⁵ Cf. Baier 1986: 231-2.

can easily explain the significance that promises have in our lives by making reference to the importance of (re-)establishing trust. In the example of Albert and Berta, we have a case in which trust was damaged by Albert's prior action. If their marriage is to continue at all, a rebuilding of lost trust is now crucial. The fact that a promise seems like exactly the right tool to *repair* this situation speaks for the supposition that the function of a promise is to (re-)establish a certain kind of trust.

Second, the idea behind the Trust View is further strengthened by the fact that a valid and sensible promise *necessarily* involves an invitation to trust. This is not to say that in offering an invitation to trust, the promisor necessarily expects the promisee to accept her offer. It certainly makes sense to utter sentences such as 'I promise you, even though I expect you won't trust me'. However, sentences like 'I promise you, but I don't *want* you to trust me' do *not* make sense. This idea is further strengthened by the fact that promises become redundant or indeed even counterproductive if the promisee already trusts the promisor. Friedrich and Southwood illustrate this with the case of a woman going out one night and assuring her partner 'Goodbye, darling. I promise you that I'll be faithful to you tonight.'¹⁶ This is certainly not a sensible use of a promise. Promises seem particularly apt in situations in which trust in specific areas is lacking or has been damaged. Where trust (to) is intact, promises are not needed.

Third, we can explain the special role that the promisee takes in a promissory exchange. As an overture to establish an intimate interpersonal relationship, the directedness that is a distinctive feature of promissory obligation, and the fact that a promise has to be accepted (or at least not rejected) by the promisee are both satisfactorily explained if we understand

¹⁶ Friedrich/Southwood 2011: 279.

the act of promising as extending an invitation to trust.¹⁷

Fourth and finally, the Trust View can make sense of the fact that a promise to φ only seems apt when φ ing would somehow be in the promisee's interest. Promises to do an action the performance of which the promisee could not care less about, like promising one's hairdresser to have pasta for dinner, are certainly infelicitous. The Trust View can offer a rationale for this. As I have pointed out above, trust has a motivational condition. Part of what it means to trust a person to do something simply is to assume that her motivations for doing so will in a suitable way align with our interests and aims. In order for the promisee to be able to have this belief, the action in question simply must stand in *some* sort of relation to her interests and aims. How could the promisor's performance of the action align in any way with the promisee's needs and desires when the promisee is absolutely neutral about performance or non-performance? We cannot by promising invite people to trust us to do things they don't care about because not caring about it makes them unable to trust us to do it in the first place.

The same argument also provides a rationale why other types of promises do not bind. Take the following sentence. 'I promise that if you come any closer, I will bash your head in.' This is not a promise, but rather a threat, and clearly not binding.¹⁸ The Trust View can make sense of this. Just as it is impossible for a person to trust another to do something which she does not care about, it is also impossible to trust another to impermissibly do one harm. Again, the fact that the promisor has to be suitably motivated with regards to the promisee's interest in order to correctly speak of an instance of trusting helps the Trust View

¹⁷ On the uptake condition of promises, see classically Thomson 1990: 302, also Owens 2012: 224-6.

¹⁸ Cf. Friedrich/Southwood 2011: 280.

to correctly capture the limits of what kind of speech-acts actually constitute a promise.¹⁹

4) **Three Interpretations of the Trust View – Effect-Based Obligations**

I have argued that the Trust View does a good job of making sense of how and when we give promises and is therefore worthy of further consideration. Of course, as a comprehensive theory of promising, the Trust View also must give a satisfactory answer to the question of why we ought to keep our promises. In the following, I will lay out three different ways a trust-based theory of promising might do so – an effect-based view, an invitation-view and a view building on the notions of normative powers and normative interests. Ultimately, I believe that each of the proposals represents an improvement over its predecessor, meaning that the third and final construal, on which promising is a normative power to create obligations in order to facilitate trust, is the most attractive way to spell out the Trust View.

Assuming that promises are best understood as involving invitations to trust, why is it that not keeping the promises one has made constitutes a wrong, or more specifically, a wronging of the promisee? At first glance, this might seem like a very straightforward question to answer. In my short discussion of the nature of trust, I have already mentioned one central feature that characterises trust. The disappointment of trust, unlike the

¹⁹ Here it again is important that the motivational condition is concerned with a subjective notion of interest. Just by itself, the Trust View thus does *not* entail the invalidity of promises to perform acts that a promisee desires even though they will ultimately be detrimental to her. Say somebody extracts a promise from you to perform a blood-letting on them, utterly convinced it will do them good. In such a case, you are likely under an obligation not to accept, or (after the fact) to inform the promisee of the objective facts on blood-letting, asking for release. Once the promise is given and accepted, however, it is not clear to me that the promise is not binding, should the promisee refuse to release you. Since promissory obligation is only *pro tanto*, this is of course compatible with the idea that on balance, you should still break your promise. There may be an independent explanation of why promises the keeping of which would objectively harm the promisee are not binding (as there may be for promises to perform immoral acts more generally), but these would be extrinsic to the Trust View.

disappointment of mere expectation, goes hand in hand with experiencing feelings of hurt and betrayal. When trust is broken, the trustor is harmed beyond the undoing of her plans. What is more, in cases of a breach of promise, it is not just any trust that is broken, but trust that the promisor has knowingly and deliberately raised. Given this, one might find it very natural to assume an *effect-based account of the wrong of breaking promises*.

(EBA) The wrong in breaking promises is the wrong of breaking trust one has knowingly and deliberately raised.

This seems like a natural and straightforward way to ground an obligation not to break one's promises, which is also reflected in a number of points in the literature. Take our initial example from Friedrich and Southwood, where they claim:

[...]Albert incurs an obligation not to betray the trust that he has invited in Berta. The wrong involved in his breaking his promise, if he chooses to do so, is a matter of his violating this obligation, that is, of his betraying the trust that he has invited. (Friedrich/Southwood 2011: 278)

This passage suggests it is the fact that not keeping the promise would constitute a *betrayal* of trust that grounds the obligation to perform. The wrong of breaking promises thus could be seen as the wrong of betraying trust that is the result of one's prior deliberate action of extending an invitation through the promise.

EBA takes the Trust View very close to a different account of promissory obligation that has loomed large in the recent literature: Scanlon's *Expectation View*. On Scanlon's account, the wrong of breaking promises is an instance of the 'more general family of moral wrongs'²⁰ of defeating expectations which one has knowingly and deliberately raised, or as

²⁰ Scanlon 1998: 296.

Scanlon puts it, breaching the wider *principle of fidelity* (Principle F).²¹ I will not pursue the question of whether or not EBA ultimately collapses into Scanlon's Expectation View or whether there are perhaps good reasons to favour a view taking trust, rather than expectations, as the central notion. What is crucial for our purposes is that, whether they are identical or not, EBA at any rate falls prey to the same problems that have been raised against the Expectation View. Most importantly, EBA ought to ultimately be rejected because deliberately creating trust and then breaking it simply is not necessary for committing the specific wrong of breaking a promise.²²

Let me illustrate this point with a familiar type of example offered by Joseph Raz.²³ A is out to get B to promise her something in order to prove to a third party that B is unreliable. A tells B of a problem which she needs help with, and of her worries about B's past unreliability. B, somewhat flaky but ultimately good-hearted, promises to do better this time, and actually provide the needed help. A remains sceptical, but, in order to prove B's unreliability, accepts the promise. It seems to me that even if B knows that the reason for A's acceptance of the promise is to prove that B cannot be counted on, B's promise – offered sincerely – is binding. Even if A in no way trusts B to provide the help, B would be wronging A if she did not come through. Promisees often accept promises that they don't expect promisor's to keep for other reasons, as well - out of politeness or kindness to the promisor, for example.

In none of these cases trust is established successfully. If the wrong of breaking promises were *only* the wrong of breaking trust, the promisors should not be bound by their

²¹ Scanlon 1998: 304.

²² For this kind of criticism of effect-based accounts like Scanlon's, see for example Raz 1977, Friedrich and Southwood 2009: 272–6, Heuer 2012, Owens 2012, Ch.9.

²³ Raz 1977: 213.

promises in any of these situations. Yet our intuitions strongly tell us that they are. If we are to find a satisfactory wrong-base for *all* intuitively valid promises, and not submit to a radically revisionist account of what constitutes a promise, we thus must look beyond the mere results of promises, and again take into focus the act of promising itself.

Before I move on, however, a clarification. I have argued that EBA is unsuited as a full account of what it is that makes it wrong to break promises. This does not, of course, amount to the claim that effect-based considerations have no place in our practical deliberation on whether or not to keep a promise. To the contrary: I take it to be undeniable that the breaking of trust can constitute a grave harm to the promisee and that we have a strong moral reason not to betray trust we invited. However, this cannot account for the wrongness of breaking any and *all* promises. A different kind of wrong is needed to explain this. In most common cases, then, we have at least two reasons to keep a promise: the (effect-independent) promissory obligation and the fact that trust was successfully established. This is reflected in our intuitions: the wronging seems graver when the promisee responds positively to the promisor's offer and puts considerable trust in her than in cases in which the promisee remains doubtful.

5) Invitations to Trust and Duties of Respect

While some passages might suggest an effect-based account, on their considered views both Friedrich/Southwood and Pink do not take the successful establishment of trust to be necessary for promissory obligation. As Pink puts it, 'trust need [not] actually be given by the promisee for promises to oblige. It is perfectly possible for a promisee to be sceptical about a promisor without that scepticism in any way removing the promisor's obligation to

deliver.’²⁴ In a similar vein, Friedrich and Southwood explain that ‘according to the Trust View, getting an individual to trust one to X is not what is needed to incur the relevant obligation of fulfilment. Rather, what is needed is to invite an individual to trust one to X and to get the individual in question to accept (or not reject) the invitation.’²⁵ The idea is that having offered the invitation, it is incumbent upon the promisor to show the suitable respect for the invitation and/or invitee, by taking the course of action suggested by the promise. This, in turn, is supposed to be a fact that is true for the larger class of invitations in general, of which promises represent a sub-class. Let us consider this suggestion, which I shall call Respecting an Invitation Account (RIA), in more detail.

(RIA) The wrong in breaking a promise is the wrong of first offering an invitation to trust, having this invitation accepted or at least not rejected, and then not delivering on it.

As Thomas Pink points out, not making good on a successful invitation to trust represents a form of disrespect towards the promisee.²⁶ We can think of other kinds of invitations for which this is the case. Even though it is up to us whom to invite to a party, say, it is generally felt to be disrespectful to un-invite guests who have previously been offered and have accepted an invitation, unless a weighty reason can be given.

In this way, RIA offers a candidate wrong-base for promises that stands to capture the problem cases raised in the last section. In these cases, a valid promise is given, yet no trust is successfully established. RIA can explain why keeping these promises is nonetheless wrong, though our reasons against breaking these promises may not be as strong as in cases where trust is furthermore successfully established. Put briefly, on the Trust View, a valid

²⁴ Pink 2011: 412.

²⁵ Friedrich/Southwood 2009: 279.

²⁶ Pink 2011: 416.

promise necessarily involves an invitation to trust, and thus necessarily gives rise to the respect-based obligations regarding invitations outlined above. Therefore, RIA seems to be able to account for the fact that simply *every* valid promise gives rise to an obligation to keep it.

RIA thus represents an improvement over EBA.²⁷ Nonetheless, it faces its own problems. As I have said, the view reduces the wrong of breaking promises to a broader, more generic class of wrongs: the wrongs involved in not respecting the invitation one has oneself given to another (who in turn also accepted it). This reduction helps explain the promissory obligation by putting it into a broader moral context. However, it also means that the wrong of breaking a promise is therefore no longer a distinctive wrong. After all, a promise is not the only way one can invite another to enter a relationship of trust.

This leads to problems when trying to make sense of promissory offers in certain contexts. Consider, for example, the following very natural exchange:

A: 'I will φ . Trust me on that.'

B: 'Do you promise?'

A: 'Ok, yes. I promise.'

In this situation, B apparently seeks further assurance by making A undertake a promissory obligation, an obligation that goes beyond what is established by the non-promissory invitation to trust. Defenders of an RIA view, on which the uttering of first and the third sentence lead to obligations of exactly the same type, are pressed to explain why it seems so

²⁷ If the respect-based obligations RIA resorts to are defensible, RIA can not only properly account for the mentioned problem cases for EBA, but also for the types of cases that David Owens argues only a normative power view can capture: cases he claims involve 'bare wrongings', such as the one involving the anthropologist Maklay and his servant (Owens 2012: 125–6). In light of this, the availability of the independent objection building on redundant promises I lay out below becomes especially important.

sensible and natural for B to seek greater assurance by demanding a promise from A.²⁸

Thomas Pink, who defends a form of the RIA view, explicitly addresses this problem. He thinks that the difference between ‘Trust me’ and ‘I promise’ is not in kind, but merely in the emphasis with which the invitation is proffered.

What [promisees] are seeking is not a moral obligation, which they may indeed possess already if any such obligation arises at all, but rather explicit acknowledgment from the promisor that through inviting trust a promise was being made and an obligation was incurred. That is, the function of ‘I promise’ and ‘that’s a promise’ is not to express an intention to obligate oneself [...]. Rather the use of such expressions is explicitly to acknowledge that a serious invitation to trust really is being made [...].²⁹

Pink thus denies that the explicit promise in step three is meant to give B any *further* reason to trust A. Rather, it serves to reaffirm and clarify that the invitation in step one really was enough reason to trust A to begin with. But this does not really capture what is going on in the envisaged situation. We can imagine B responding to A’s initial offer by saying something like ‘I wish I was able to trust you’. In response to this, it seems particularly appropriate to offer a promise to *change* the situation, i.e. to give the promisee a *new* reason for trusting the promisor. For all intents and purposes, promises appear to facilitate trust to a considerably greater extent than a verbal exhortation along the lines of ‘No, look, you really can trust me!’ ever could, no matter how emphatically put forward.³⁰

More generally, we can ask defenders of RIA what reason at all the promisee has to trust the promisor in a response to an invitation offered by her. As Annette Baier said, “‘Trust

²⁸ I do not mean to deny that we can sometimes offer promises by uttering the words ‘I will φ . Trust me.’ This is evidently possible. My argument relies solely on the possibility that in some contexts, it will not be clear whether an explicit invitation of ‘I will φ . Trust me.’ constitutes a promise, such that clarification can reasonably be requested to acquire greater assurance.

²⁹ Pink 2009: 414.

³⁰ Of course, accounting for this is not only a problem for RIA, but also for EBA.

me!” is for most of us an invitation which we cannot accept at will—either we do already trust the one who says it, in which case it serves at best as reassurance, or it is properly responded to with, “Why should and how can I, until I have cause to?”³¹. Of course, defenders of RIA will deny that an invitation to trust, understood in their sense, is best described as a simple appeal on the part of the promisor to trust her. Friedrich and Southwood, for example, understand an invitation to trust to also involve ‘signalling to the other individual one’s recognition of the importance that the relevant action has for her, and one’s willingness to be moved on that basis [...]’.³² Pink also highlights it as important that the promisor expresses both her motivation and capacity to deliver what is offered in the promise.³³

But it seems to me that what sets instances of ‘I promise’ and ‘Trust me’ apart is certainly not the higher informational value of the former. This is reflected in the fact that we can promise against the evidence, as Berislav Marusic has convincingly argued.³⁴ Imagine a son that repeatedly fails a driving test, driving his parents to despair. After the third failed try, he sincerely promises his parents that he will pass the test the next time. If a promise communicated that one is motivated and capable of delivering, we would have to classify the son’s promise as insincere. However, this seems to not be the case. As long as it is not impossible for the promisor to *acquire* the necessary motivation and capacity before the time to perform comes around, it seems to me perfectly reasonable to promise, even if the evidence does not support a belief that the promisor will perform.³⁵ Of course, I am not arguing against the claim that the ability to eventually fulfil the promise is a necessary

³¹ Baier 1986: 244.

³² Friedrich/Southwood 2011: 279.

³³ Pink 2009: 409.

³⁴ Marusic 2013.

³⁵ See Marusic 2013: 293.

condition for the validity of the promise. I cannot successfully promise you to reverse the direction of the earth's rotation, for example. What I doubt, however, is that in promising, we actively impart substantial information about our motivation and skills to the promisee. Cases of promising against the evidence are proof of this.³⁶

This gives us reason to doubt the idea that it is this kind of informational value that sets 'Trust me' and 'I promise' apart. If it is not, however, then it becomes unclear how on RIA, promises can effectively fulfil the very purpose the Trust View set out for them – the establishment of new, or at least, firmer trust between individuals. *Pace* Pink, defenders of the Trust View thus have reasons to hold that it is indeed the specific kind of *obligation* incurred in promises that makes the difference here. This specific obligation can serve as the warrant for trust when the simple exhortation is not enough, thus giving potential trustors a further reason to place trust in us, even when we have already verbally invited them to do so.³⁷ In the following, I will try to make sense of this claim drawing on the resources provided by Normative Interest Accounts of promissory obligation.

6) A Normative Interest Version of the Trust View

Normative Interest Accounts, versions of which have recently been defended by Joseph Raz

³⁶ Here is another way one might argue for the claim that promises involve more than only an invitation to trust. Imagine that I am part of a religious sect that prohibits me from giving promises. It seems that I can still emphatically implore you to trust me to do a certain thing, as long as I make clear that I am not making a promise in doing so. Whatever one may say about my reasons for sticking to my religious convictions, I at least do not appear to be conceptually confused here, as Pink would have to hold I am. Thanks to Stefan Riedener for helping me come up with this example.

³⁷ My objection here has the potential of generalizing. For any theory that attempts to reduce promissory obligation to an obligation that results from a different kind of duty, we can construct a case parallel to the A-B case laid out here. If it makes sense for the promisee to demand a promise even after this different duty has been triggered in a non-promissory way, this is evidence for the normative effects of the promise not being exhausted in those explainable by reference to these other duties. If this holds across the board (which of course remains to be seen), we would be provided with an argument for a non-reductive (i.e. normative power) view of promissory normativity, such as I defend here.

and David Owens, set out from a very intuitive understanding of what we do when we promise.³⁸ The purpose of a promise is a rather obvious one, they argue: it is simply to bind oneself, i.e. to undertake an obligation to act as promised. As such, these accounts are committed to the idea that we have capacities that can be called *normative powers*. Under certain circumstances, we can create obligations or permissions by performing communicative actions whose explicit purpose it is to do just that. This allows us to actively shape some of the elements of our normative situation.³⁹ Of course, the power to promise is not the only instance of such a power. As defenders of Normative Interest Accounts of promissory normativity regularly stress, we are familiar with (and often much less puzzled about) several other normative powers such as ordering, commanding, giving and consenting.

By itself, this idea may not seem to add any interesting new theoretical puzzle pieces. After all, accounts making recourse to normative powers have been familiar at least since the time of the Natural Lawyers. The crucial element which accounts such as the ones of Raz and Owens add to the mix is focusing not only on the power itself, but rather on what underlies this power. The central idea of Normative Interest Accounts is that we can explain why any given normative power has genuine impact on the normative situation by making recourse to a special value it has, or differently put, a special human interest that is served by it. We can call these interests that are served by the existence of normative powers *normative interests*.⁴⁰ Let us apply this to the case of promising. If we understand promising as the power

³⁸ See Raz 1977, 2014 and Owens 2012. Seana Shiffrin (2008) and Ulrike Heuer (2012) also at least express sympathies for Normative Interest Accounts.

³⁹ Raz 1977: 228. This proposal might give rise to worries about *bootstrapping*. I will briefly discuss such worries at the end of this paper.

⁴⁰ This term is due to Owens 2012. The interests at issue can be normative in a strong or a weak way. Either they are intrinsic interests in some normative relation for its own sake (this is what Owens suggests) or they can be extrinsic interests in some normative relation for its value for something else (Raz's proposal, and the one I elaborate here both fall in this second category).

to voluntarily bind ourselves to a certain course of action by undertaking a (directed) obligation to the promisee, then this power must be grounded in an interest in having the ability to bring about precisely this type of normative relation.

Whether this theoretical endeavour succeeds depends crucially on whether the proposed interest stands up to scrutiny. A convincing Normative Interest Account of promising not only needs to point to an interest the satisfaction of which is actually valuable, but an interest that is furthermore fit to ground a power with the specific features of the promissory power as we know it (and not just some other power with less demanding conditions). Raz and Owens of course both suggest interests to fulfill this role, though neither of them focuses on the notion of trust.⁴¹ For Raz, it is our interest in entering relationships of involvement that grounds our promissory power. Owens instead focuses on the ‘authority interest’ of the promisee, i.e. her interest in being in the position of power to decide whether or not the promisor has the obligation to perform what he has promised (by releasing or not releasing him from the promise, which is within her remit). Both of these views have been met with criticism, both regarding their status as genuine human interests and their fit to the contours of the promissory power.⁴²

On the *Trust-Based Normative Interest View* (TNI), the valuable human interest served by promising is our interest in establishing relationships of trust with others. Thus understood, giving a promise involves a special invitation to trust coupled with the exercise of a normative power. In promising, we extend an invitation to the promisee to trust us to perform a certain act, while at the same time offering something that can serve as the necessary warrant for

⁴¹ Owens in fact explicitly rejects the idea that we can explain the binding force of a promise by reference to trust (Owens 2017: 218). He instead argues for a reverse dependence: in order to understand what it means to trust a promise, he claims, we have to have independent purchase on its bindingness (Owens 2017).

⁴² For Raz’s view, see e.g. Anwander 2008, Section 5.3 and Molina 2019. For Owens’ view, see Bennett 2015.

this trust: a directed obligation to perform the promised act. We can capture this idea as follows:

(TNI) The wrong in breaking a promise is the wrong of failing to discharge an obligation created through the normative power of promising, a power which is in turn grounded in our interest in creating warrant for trust.

TNI is able to draw on both the intrinsic value of trust relationships and the close fit between our promissory practices and interest in trust that I have argued for above. Recall again the case of Albert and Berta. Albert's prior misstep has led Berta to doubt his trustworthiness. Where before she did trust him, she is no longer able to bring herself to do so. It is precisely in situations like these where promising can play its important facilitatory role. Berta needs further warrant to trust Albert not to slip up again. Albert wants to give her such warrant. I propose that since being under a voluntarily undertaken, directed obligation can serve as warrant in cases like these, we have a *normative interest* in being able to create such voluntarily undertaken, directed obligations through the exercise of a normative power of promising. Why could only a directed, voluntarily undertaken obligation possibly provide the necessary warrant for trust? Recall the motivational condition: in order to trust an agent to φ , we have to have some reason to believe that the agent we trust is motivated to φ for reasons that are related to our interests in the right way. An obligation that is not directed towards the promisee or not voluntarily undertaken would not necessarily be proof of the right type of motivation.

In fact, TNI thus ties in nicely with some issues that have been raised in the literature on trust. One thing philosophers have been discussing is the question of whether knowing that someone is subject to a norm or social constraint can be a good reason to trust that

person.⁴³ As is regularly pointed out, showing an agent to have reason to comply with norms to avoid blame and/or sanction do not amount to showing that agent to be trustworthy. If it did, this would lead to counterintuitive implications in a number of cases. One such case is that of the sexist employer, who treats female employees well only because he believes that he would face legal sanctions if he did not.⁴⁴ The employer's fear of legal repercussions may be a sound basis for relying on him not to mistreat his employees, yet it surely does not provide the necessary warrant for his employees to *trust* him.

The situation is different in the case of promising as conceived by TNI. If the employer were to sincerely promise his female employees to treat them well (maybe he wants to turn over a new leaf), this is better warrant for his trustworthiness for two reasons.⁴⁵ First, the fact that promissory obligation is voluntarily undertaken demonstrates that, at least in some sense, he *intends* to be under the obligation. Secondly, the fact that the promissory obligation is directed towards the employees, putting them in a special position to demand compliance, is proof of his willingness to do it for *them*. If he fails to keep his promise, his promise-breaking (apart from the damage done by the action itself) will be above all a wronging of all of the female employees to whom he originally offered the promise.

TNI is therefore able to provide a strong explanation for the existence of our practice of giving and keeping promises in its familiar form. We have a need for trusting and being trusted, but often find ourselves unable to (sensibly) do so for lack of warrant. Being able to make a normative difference through promising helps 'cover' this lacuna by allowing the promisor to voluntarily undertake a directed obligation through the offering of the promise.

⁴³ Hardin 2002: 53.

⁴⁴ Potter 2002: 5.

⁴⁵ Of course, 'better' warrant here does not yet entail sufficient warrant.

How does TNI fare when pitted against the two rival views discussed above? To find out, we have to focus on the following question: which wrong-base for promise-breaking does TNI provide for? The answer is a simple one, and one that it shares with all other versions of Normative Interest Accounts: breaking a promise is wrong, because – having promised – the promisor fails to discharge an obligation owed to the promisee.⁴⁶ That is all. The reason to keep the promise derives solely from the fact that the obligation obtains, independently of any effect a break of promise and/or trust would have on the promisee's interests. The obligation not to break promises on TNI is thus not one that directly derives from the value of our promissory practice via some general duty to contribute to the project of upholding valuable institutions.⁴⁷ Neither is the reason a self-regarding one, derived from my personal interest in being in a position to use the promissory convention to create further valuable interpersonal relationships of trust.⁴⁸ The wrong-maker of an unjustified breach of promise simply is the fact that a valid promise has been given. TNI thus classifies as a non-reductive account of promissory normativity. Nonetheless, human interests still play a crucial role. But instead of the individual obligations, it is the *power* to establish this promissory obligation that TNI grounds in an interest shared by both promisor and promisee – the trust interest.

The explanation of promissory obligations inherent in Normative Interest Accounts thus can be helpfully described as having two levels.⁴⁹ Since the idiom of normative powers does not always lend itself to maximum clarity, let me try to express the central claims of TNI in a different and hopefully more familiar way: as a claim about moral principles and

⁴⁶ Owens 2012: 148f.

⁴⁷ See paradigmatically Rawls 1971. See also Kolodny and Wallace 2003.

⁴⁸ See paradigmatically Hume 1978.

⁴⁹ Raz 1977: 219.

their grounds.

On a first level, TNI postulates a general moral norm requiring us to perform those actions which we have sincerely and validly promised to perform. Call this the Promissory Principle.

(PP): If you have made a valid promise to φ , then you have an obligation (to the promisee) to φ .

The fact that the kind of dependence specified by the Promissory Principle holds *just is* the fact that the agent in question has the kind of normative control over her situation that is crucial for the normative power of promising. Since it is generally within our power to render the antecedent true (possibly in concert with the promisee, as the promise potentially has to be accepted to be valid), (PP)'s truth means that it is generally within our power to bring about a certain kind of normative change. Note that the truth of (PP) by itself does not commit one to the existence of normative powers. Reductive accounts like RIA or EBA can in theory also subscribe to (PP).⁵⁰ What sets TNI apart from these, and qualifies it as a genuine normative-power-view, is the further claim that (PP) is a non-reducible principle. Its truth is not merely an implication of a different, wider moral principle. As a result, the correct answer to the question 'Why do I have an obligation to φ ?', on TNI, simply is 'Because you made a valid promise to φ .' We can capture this by adding an explanatory clause to (PP).

(PP-NR): If you have made a valid promise to φ , then you have an obligation (to the promisee) to φ because you have made a valid promise to φ .⁵¹

On a second level, TNI provides a story of why the promissory principle obtains. The

⁵⁰ Whether a reductive account will be able to account for (PP) in practice depends on the reduction base that it proposes. I have argued that EBA fails to do so, while RIA has better prospects to succeed.

⁵¹ Here, 'because' is meant to indicate a full explanation.

principle is fully grounded, so the idea goes, in our interest in having precisely the kind of normative control that is afforded by (PP). On TNI, (PP) is not just some brute fact, but can instead be explained with recourse to the trust interest, along the lines laid out above.

This two-level-structure of Normative Interest Accounts has some important advantages. In contrast to EBA and RIA, TNI does not reduce the wrong of breaking promises to a more general kind of wrongdoing. This allows TNI to avoid the problems that have plagued EBA and RIA. While EBA's suggested reduction base led to problems accounting for the bindingness of certain promises, TNI's commitment to (PP) as a non-reducible principle affords it full extensional adequacy 'for free' – what it takes for a promise to be binding, on the view, just is for it to have been given under conditions of validity, exactly as intuition suggests. Furthermore, my elaboration of TNI's basic structure has already made clear how it can make sense of the cases that proved problematic for RIA. On TNI, we confer *further* warrant for trust by incurring a new obligation as the result of the exercise of a normative power. This obligation is different in kind to the obligation incurred through the explicit invitation. As such, TNI can easily account for the fact that it seems eminently sensible to demand a promise even after a sincere non-promissory invitation to trust has been offered.

Nonetheless, TNI still affords us an interesting theory of why we give and keep promises, by providing a story of how the *power*, not the individual obligation, is grounded. As such, TNI still qualifies as a genuine version of a Trust View. Because the promissory principle (PP) is ultimately grounded in the trust interest, the notion of trust plays a crucial role in the explanation of promissory normativity. This not only affords us the kind of illuminating explanations of the particularities of our giving and keeping of promises that I have laid out above. It also serves an important vindicatory purpose, allowing us to draw a

principled line between the promissory power and those putative normative powers which do not seem to have the same genuinely normative impact. We can thus show what sets apart the promissory power from, say, the putative power to obligate another to enter into combat by issuing a challenge to a duel. The first is well-grounded in a genuine human interest, the trust interest, while the latter is founded on a conception of honour that we would on reflection not judge to be valuable.

While TNI thus enjoys strong initial plausibility, and furthermore seems well-equipped to face the mentioned extensional worries, it may be subject to some problems pertaining to the general strategy of Normative Interest Accounts. Two problems bear mentioning in particular. First, TNI might be found to involve a worrisome kind of bootstrapping. Hume and Prichard have historically fielded powerful challenges to conceptions of promising on which we can bring about obligations at will. Both were worried that such creation of obligation from ‘nothing’ is incomprehensible.⁵² Since TNI is committed to the idea of promising as a normative power, it may appear to be subject to this kind of worry. However, the availability of a grounding explanation making recourse to a genuine human interest may serve to significantly dampen the impact of this objection against a Normative Interest Account such as TNI, since it grants important argumentative resources that simpler non-reductive accounts are lacking.

Secondly, one might worry that TNI’s project of grounding the promissory power in the fact that it serves some human interest or value represents an objectionable kind of *wishful thinking*.⁵³ As we have seen, TNI seeks to ground the promissory principle in the value of its obtaining. In most domains, such reasoning seems clearly fallacious: evidence for the

⁵² Hume 1978: 524, Prichard 1940/2002: 257.

⁵³ See Anwander 2008: 131-2.

fact that it would be good if some natural law obtained clearly does not constitute evidence for the actual obtaining of this law. Normative Interest Accounts such as TNI are committed to the claim that things work differently in the normative domain. They are, however, not alone in holding that view. Recently, there have been a few attempts to defend the validity of inferences following the schema [it would be good that p, therefore p] for at least some specific classes of normative propositions.⁵⁴ Still, the claim remains a contentious one, and more needs to be done if one ultimately defends a view along the lines of TNI.

Both of the problems just outlined require detailed and carefully argued answers. Unfortunately, providing these is not a task I can pursue here.⁵⁵ However, I do believe that the prospects for satisfactorily meeting these challenges are ultimately good. In this paper, I hope to have at least made a successful case for the claim that further exploration of the merits of TNI is a task well worth pursuing. If the view indeed is defensible, I believe it is not only the best candidate for a theory of promissory obligation building on the notion of trust, but furthermore a strong contender for the most convincing theory of promising in general.⁵⁶

Humboldt-Universität zu Berlin, Germany

⁵⁴ Enoch 2009, Preston-Roedder 2014, Sayre-McCord ms.

⁵⁵ I attempt to answer the wishful thinking worry in Bruno ms.

⁵⁶ For their extremely helpful comments on earlier versions of this article, I would like to thank John Colin Bradley, Francesca Bunkenborg, David Heering, David Owens, Stefan Riedener, Thomas Schmidt and Lukas Tank, as well as two anonymous referees for the Philosophical Quarterly. I also benefitted from useful feedback from audiences at the 7th Humboldt-Princeton Grad Conference, the XXIV. Deutscher Kongress für Philosophie, and Thomas Schmidt's Ethics Colloquium at Humboldt-Universität zu Berlin. Work on this project has been supported by the Einstein Stiftung Berlin.

References

- Anwander, N. (2008) *Versprechen und Verpflichten*. Münster: Mentis.
- Baier, A. (1986) 'Trust and Antitrust', *Ethics*, 96: 231-260.
- Bruno, D. (ms.) 'Value-Based Accounts of Normative Powers and Wishful Thinking', Unpublished Manuscript, Humboldt-Universität zu Berlin.
- Enoch, D. (2009) 'Wouldn't It Be Nice If p, Therefore, p (for a moral p)', *Utilitas*, 21: 222-224
- Faulkner, P. (2007) 'On Telling and Trusting', *Mind*, 116: 875-902.
- Friedrich, D. and Southwood, N. (2009) 'Promises beyond assurance', *Philosophical Studies*, 144: 261-280.
- (2011) 'Promises and Trust', in H. Sheinman (ed.) *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press.
- Hardin, R. (2002) *Trust and Trustworthiness*. New York, NY: Russell Sage Foundation.
- Heuer, U. (2012) 'Promising' (Parts 1 and 2), *Philosophy Compass*, 7/12: 832-851.
- Hieronymi, P. (2008) 'The Reasons of 'Trust'', *Australasian Journal of Philosophy*, 86: 213 – 236.
- Holton, R. (1994) 'Deciding to Trust, Coming to Believe', *Australasian Journal of Philosophy*, 72: 63-76.
- Hume, D. (1738/1978) *Treatise on Human Nature*. L.A. Selby-Bigge (ed.), Oxford: Oxford University Press.
- Jones, K. (1996) 'Trust as an Affective Attitude,' *Ethics*, 107: 4-25.
- Kolodny, N. and Wallace, J. (2003) 'Promises and Practices Revisited', *Philosophy and Public Affairs*, 31: 119-54.
- Marusic, B. (2013) 'Promising against the Evidence', *Ethics*, 123: 292-317.
- McGeer, V. (2008) 'Trust, hope and empowerment', *Australasian Journal of Philosophy*, 86: 237 – 254.
- Owens, D. (2012) *Shaping the Normative Landscape*. Oxford: Oxford University Press.

- (2017) ‘Trusting a Promise and Other Things’, in P. Faulkner and T. Simpson (eds.) *The Philosophy of Trust*. Oxford: Oxford University Press.
- Pink, T. (2009) ‘Promising and Obligation’, *Philosophical Perspectives* 23: 389–420.
- Potter, N. N. (2002) *How Can I be Trusted? A Virtue Theory of Trustworthiness*. Lanham, Maryland: Rowman & Littlefield.
- Preston-Roedder, R. (2014) ‘A Better World.’ *Philosophical Studies* 168: 629–644.
- Prichard, H. A. (1940/2002) ‘The Obligation to Keep a Promise’, in *Moral Writings*, J. MacAdam (ed.), Oxford: Clarendon Press.
- Rawls, J. (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Raz, J. (1977) ‘Promises and Obligations’, in P. Hacker and J. Raz (eds.) *Law, Morality and Society*. Oxford: Oxford University Press.
- (1990) *Practical Reason and Norms*. Princeton: Princeton University Press.
- (2014) ‘Is There a Reason to Keep a Promise?’, in Klass, G.; Letsas, G. and Saprai, P. (eds.) *Philosophical Foundations of Contract Law*. Oxford: Oxford University Press.
- Sayre McCord, G. (ms.) ‘On a Theory of a Better Moral Theory and a Better Theory of Morality’. Draft of May 2018 obtained from <https://philosophy.unc.edu/people/geoffrey-sayre-mccord/> accessed 10 December 2019
- Scanlon, T. (1998) *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Shiffrin, S. (2008) ‘Promising, Intimate Relationships, and Conventionalism’, *Philosophical Review*, 117: 481-524.
- Thomson, J. J. (1990) *The Realm of Rights*. Cambridge, MA: Harvard University Press.