

Promising by Normative Assurance

Luca Alberto Passi

May 2023

Abstract

This paper develops a new theory of the morality of promissory obligations. T. M. Scanlon notoriously argued that promising consists in *assuring* the promisee that we will do something. I disagree. I argue that it is true that promising consists in *assuring* the promisee, but what the promisor gives to the promisee is not an assurance that they will do something, but that the normative situation is in a certain way.

Keywords: Promising, Moral Obligation, Normative Power, Right, Normativity, Trust.

Word Count: 10.020.

I Introduction

We all have an intuitive grasp of what it is to make a promise. However, when we ask why we should keep the promises we make, our certainties begin to fade. It seems trivial to ask: why must I water my mother's flowers if I promised her I would? A plausible answer: because I promised. But what is the explanation? In this essay, I offer a new theory of the morality of promissory obligations. The guiding idea is that promising should be understood as a special form of gift-giving, whereby the recipient is assured that they will have an unconditional opportunity to take an interest in the promise. The wrongness of promise breaking lies in depriving the promisee of an opportunity that we have convinced her of having and, if any, in harming the interests born out of that conviction. The point of this essay is to say this more slowly.

In §2, I present Tim Scanlon’s *Expectation Account* (EA) of promising.¹ The EA explains promissory obligations on the basis of certain expectations that the promisor intentionally generates in the promisee. I start with the EA because (a) in my view, Scanlon came closest to a correct theory of promising; (b) my positive proposal is a version of the EA, for it is based on what the promisor leads the promisee to believe. In the discussion, I show that the EA is untenable, as it falls victim to two serious objections. Then, I introduce the *Hybrid Account* (HA) of promising, developed by Niko Kolodny and Jay Wallace to solve the EA’s problems.² As we shall see, however, the HA does not solve all the problems of Scanlon’s EA. These problems motivate the search for another story. In §3, I present my positive proposal. In §4, I consider several difficulties for my account. This will allow me to show how my account stands in relation to *Normative Power Accounts* (NPAs) of promising,³ the EA’s historic rivals. NPAs generally attribute moral powers to agents to constitute directed moral obligations by promissory utterance, leaving no room for ‘expectations’ in explaining why promises bind morally. My account combines the main virtues of these competing accounts. Like NPAs, it is extensionally adequate (the EA is not), and, like the EA, it is a ‘perlocutionary account of promising’⁴: the promissory obligation is not constituted by declaration but is explained by certain beliefs that the promisor intentionally generates in the promisee.

2 The State of the Art

2.1 The Practice Account

Scanlon develops his EA in opposition to what are commonly called *Practice Accounts* (PAs) of promising. PAs say that there is a practice (that of promising), which is *useful* (Hume) or *just* (Rawls), and which is shared by members of a certain group. Given this practice, which functions according to certain intentions and expectations, there is a moral judgment that it is wrong to violate its rules. According to PAs, the obligation to keep promises is owed to all those who stand to benefit from, or have con-

¹Scanlon (1998).

²Kolodny & Wallace (2003).

³I think of Shiffrin (2008), Owens (2012), Raz (2014), and Bruno (2020).

⁴This tag is from Pratt (2003).

tributed to, the practice. If one breaks a promise, one ‘wrongs them collectively, either by depriving them of an important benefit or by free-riding on their contributions to the social practice’.⁵

In recent times, however, PAs have gone out of favour. Rightly so, in my view, and for one reason in particular: PAs fail to account for the *directedness* of promising. PAs tell us that it is wrong to break the rules of a useful or just practice, as this would amount to undermining it or free-riding on it, but they do not explain why the promisee has a special moral standing to blame the promisor. The moral standing of the promisee is intuitively different from the standing of all others who participate in the practice, and this difference is due to what I call the *Directedness Intuition*.

Directedness Intuition: The wrong we commit by unjustifiably breaking a promise is primarily a wrong to the promisee, not to the rest of society.⁶

A bystander witnessing the breaking of a promise may disapprove of the promisor’s behaviour and say: ‘Imagine if everyone did what you did!’ But he or she does not have the *privileged* role that the promisee has to blame, forgive or ask for compensation if the promisor breaks the promise. For instance, a bystander cannot forgive me for not watering my mother’s flowers, but my mother can.

2.2 The Expectation Account

One of Scanlon’s central goals is to offer a theory of promising that accounts for the *Directedness Intuition*. To do so, Scanlon argues that the most fundamental wrong we commit when we break a promise is the wrong of failing to meet certain desirable expectations that we have intentionally generated *specifically* in the promisee. According to Scanlon, there is a moral principle, justified in *contractualist terms* (i.e. a principle that no one could reasonably reject), which explains promissory obligations. Here it is:

Principle of Fidelity (F): If (1) A intentionally leads B to expect that A will do X (unless B consents to A not doing X); (2) A knows that B wants

⁵Scanlon (1998: 316).

⁶Kolodny & Wallace (2003: 125). Although I don’t have space to discuss their views here, some PA theorists have tried to show how PAs can account for the *Directedness Intuition* as well. For a recent example, see Melenovsky (2017: 14-16).

to be assured of this; (3) A acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) B knows that A has the beliefs and intentions just described; (5) A intends for B to know this, and knows that B does know it; and (6) B knows that A has this knowledge and intent; then, in the absence of special justification, A must do X unless B consents to A not doing X.⁷

Is Principle F a valid moral principle? Scanlon thinks so. Potential promisees have no reason to reject a moral principle that achieves the value that promises are meant to serve: the value of the assurance being fulfilled. The promisee, Scanlon concludes, has no reason to reject a principle that generates an obligation for the promisor to do X. Similarly, the potential promisor has no reason to reject Principle F. It is also valuable for the promisor to give assurances and bind herself in this way if she *wants* to.

I agree with Scanlon that the role of expectations is essential in explaining the *wrong* we commit when we unjustifiably break a promise. However, I do not think that Principle F is the principle that governs promissory obligations. This is because a theory based on Principle F runs into at least two important objections: the *Circularity Objection* and the *Necessity Objection*. Let us examine them in detail.

Circularity Objection: The obligation that is explained by the expectation is the same obligation that motivates the expectations.⁸

In cases where I have ‘prior reasons’ to, for example, sell you my horse, I can indicate that set of prior reasons. For instance, I could tell you that I always wanted to sell you the horse, that I need the money, that I refused to sell it to five different buyers because I was waiting for you to buy it, etc. If I am successful in assuring you that I will sell it to you, and other conditions are met, then I trigger Principle F: I now have an obligation to you to sell you the horse. However, promises are often made when there are no prior reasons to do what we are promising to do. If I had no prior reasons to sell you the horse, I could still generate the relevant expectation by saying: ‘I promise I will sell you the horse!’ When the assurance is not provided by reasons given during

⁷Scanlon (1998: 304).

⁸The *Circularity Objection* is nicely presented by Kolodny & Wallace (2003).

a conversation, we often resort to promising. But on what basis can *this* assurance be produced? In order for you to expect that I will sell you my horse, you must believe that I am obligated by Principle F to sell you my horse. But in order for me to be obligated by Principle F, you must expect that I will sell you the horse. There is a vicious *circle* here: Principle F creates a binding obligation only if the recipient believes that the assurer will perform. But for this to happen, there must be something to explain why the recipient forms the relevant expectation when I say: ‘I promise I will sell you the horse!’ What is the rational basis for your expectation in this case? Saying that you could simply form a groundless belief that I will do it is not enough. You need a rational basis for this expectation, and the EA does not seem to be able to explain what this basis might be. If the *Circularity Objection* is valid, then Principle F cannot get started. A principle that cannot be triggered by saying ‘I promise that I will do X’ can still be considered true, but it fails to explain standard cases of promising. Let us look at the other objection:

Necessity Objection: There are intuitively valid promises in which one or more conditions of Principle F are not satisfied.

This objection says that it is not *necessary* to satisfy all the conditions of Principle F in order to incur a promissory obligation. Let us consider two cases of promises that are intuitively valid, but whose promissory obligations the EA cannot explain. Consider:

Soccer Match. A father promises his daughter to go to her soccer match. The daughter does not believe he will go, but she desires him to. After seeing her father’s insistence and firmness, the daughter accepts the promise.⁹

This case violates condition (1) of F. Consider:

Fishing Squad. You promise me that you will come fishing, although I have no particular interest in your coming. Seeing how important is for you, I accept the promise.

⁹Friedrich & Southwood (2009: 272).

This case violates condition (2) of F.

The EA tells us that promisors in these cases do not incur in a promissory obligation, even if these look like intuitively valid promises. Scanlon commits us to a ‘radically revisionist’¹⁰ account of promising. Is there any way to intervene on Scanlon’s EA to deal with these two rather deep objections?

2.3 The Hybrid Account

There have been some more or less successful attempts to rescue the EA from these two objections. In this section, I will look closely at the version of the EA that I think best resists the two objections outlined above. We can call the third proposal the *Hybrid Account* (HA), a theory developed by Niko Kolodny and Jay Wallace.¹¹ HA theorists think it is sufficient to combine the EA with the PA (Rawls-style) to address both objections.

In fact, the *Circularity Objection* can be nicely answered. When I say: ‘I promise I will do X’, I incur a moral practice-based obligation not to *free-ride* on a just social practice — that is to say, I incur a practice-based promissory obligation (as Rawls’s PA tells us). Since I am obligated to do X by the practice-based obligation, the promisee has *reason* to believe that I will do X, and thus forms the relevant expectation, which triggers F. The HA is better than a PA because it accounts for the *Directedness Intuition*. Indeed, the F-based obligation is directed.

HA theorists also claim that they can handle the *Necessity Objection*. The HA shows that in *Soccer Match* and *Fishing Squad* there are valid promises. The HA tells us that we can make two kinds of promises. One in which we have two obligations, where F is triggered; and one in which, if F is not triggered, we just have a Rawls-style obligation not to *free-ride* on a just practice. In this picture, the *Necessity Objection* seems harmless: even if Principle F is not triggered, we still have a moral practice-based obligation to do X. However, for all the promises that do not activate F, the *Directedness* is lost: whenever F is not triggered, the promisee cannot really say that the promisor owes it to them, more than to others, to keep her promise. However, it seems to me that if we grant that one is really making a valid promise, then the *Direct-*

¹⁰Heuer (2012: 836).

¹¹Kolodny & Wallace (2003).

edness should be preserved even in cases where F is not triggered. To make this point more intuitive, consider *Coming to Believe*, a case that violates Principle F's condition (1):

Coming to Believe: I am a visual artist. I have to present my work at an exhibition next month. I talk about this to a friend of mine, Abe, who is unreliable. Abe promises me that he will be there to help me install my work: a reproduction of Van Gogh's *Joseph Roulin*. I want him to help me, but I do not believe that he will. However, Abe turns out to be reliable after all: we have a few appointments to see my work, and he is always on time and shows that he takes my exhibition seriously. I begin to think that he is really committed and that he will come. Now I expect him to come. I begin to rely on the fact that he will be there, for example by not paying someone else to come and help me. Now the promise has *great value* for me.

According to the EA and the HA, I do not have a special claim against Abe, for I do not expect that he will come when I accept the promise. However, it does not seem to me that the fact that I did not believe he would come prevents me from having a special claim against Abe. After all, I accepted the promise. I wanted him to keep it, but I simply had no control over my expectations. If Abe breaks his promise, confirming my initial doubts, he would be doing a wrong *to me*, and not to the rest of society.¹² The second example violates Principle F's condition (2):

Coming to Care: Kathi is a law student, and her dream is to become a magistrate. In her country, the competition to become a magistrate is tough. At least two years have to pass between completing a law degree and entering the competition for the magistrature. Kathi has no money

¹²The analogue of this case is, in Scanlon's work, *The Profligate Pal*. You lent money to *Pal*, no big deal. *Pal* has promised you that he will pay you back, without generating any expectations, since he is, in fact, unreliable. Scanlon says: 'Does he have an obligation to pay you back? Principle F does not generate any such obligation' (Scanlon 1998: 312). Someone may argue that if I change my mind about *Pal*, then F could be triggered at a later point. However, it is essential for the EA that B's expectation is an immediate response to A's offer, as condition (3) of F stresses. I could express the intention to do X today, understand that this does not create any expectation you, and continue living my life. If you, in a while, at t_1 , changed your mind, and formed expectations by virtue of the intention I expressed at t_0 , I could be bound by Principle F, and this without having any idea of being so bound. Scanlon does not want such a story, for it would trigger F too easily.

and is considering an alternative career. There is no way she could support herself and prepare for the competition. Kathi's aunt, who is a prodigious law scholar, promises Kathi that she will help her prepare for the competition. Kathi does not *want* this assurance, as she does not think that, considering her circumstances, she will be in a position to prepare for the competition. However, after her aunt has insisted, she accepts the promise. Six months later, Kathi unexpectedly wins a scholarship for aspiring magistrate candidates. Kathi calls her aunt and tells her that she is ready to prepare for the competition. Now the promise has *great value* for Kathi.

According to the EA and the HA, Kathi does not appear to have a special claim against her aunt, for she does not *desire* such assurance when she accepts the promise. However, by accepting the promise she seems to get a 'right to the performance', which allows her to rely on the promise made by her aunt when her circumstances change. The aunt might say to her: 'Did you see that? I was right' — but being right does not free her from the obligation to help Kathi. It seems to me that this is a genuine promise, and that the aunt owes it to Kathi, and not to the rest of society, to help her. Our explanation of promissory obligations in these cases should make sense of the *Directedness Intuition*, or so it seems to me.

3 The Normative assurance Account

3.1 Scanlon's Desired Assurance

Now I shall present my positive proposal. The goal is simple: I want to address the *Necessity Objection* while doing justice to the *Directedness Intuition*. Then I will investigate whether my proposal also evades the *Circularity Objection*. It is helpful to start with the very idea of 'giving an assurance'. According to the EA, we give an assurance when we lead the recipient to believe that we will do X and the recipient *wants* us to do X. Here I will use the terms 'desire' or 'want' interchangeably, because the idea of 'want' that Scanlon has in mind is meant to capture various nuances, such as 'desire', 'care', 'want', 'give importance to', and so on. I call what Scanlon has in mind a desired assurance. This is how I define giving a desired assurance:

A gives B a desired assurance if and only if A intentionally leads B to believe that A will do X (unless B consents to A not doing X) *and* A knows that B wants to be sure that A will do X.

There are two key notions at play here: belief and desire. The promisee must form the *belief* that the promisor will do X. Moreover, B must not be indifferent: B must *desire* that A will do X, and wants it so much that she wants to be sure that A will do X. I want to show that giving a desired assurance is not necessary to generate a right to the performance for the promisee, as the EA and the HA claim.

3.2 Unconditional and Normative Assurance

Let us start with *Coming to Care*. Is there a moral principle that assigns an obligation to the aunt to help Kathi? Even if Kathi's aunt is assuring Kathi of something, she is not giving her a desired assurance, but what I call an unconditional assurance. I define giving an unconditional assurance in this way:

A gives B an unconditional assurance if and only if A intentionally leads B to believe that A will do X (unless B consents to X's not being done), regardless of what B desires/wants when the assurance is given.

It does not take much work to derive a principle, analogous to F, where instead of a desired assurance, the promisor gives an unconditional assurance. Even if not reasonably rejectable, this principle would not be able to explain cases such as *Coming to Believe*. So, our strategy is to start with a principle that explains *Coming to Believe*, and then see if it explains the other cases as well. As we shall see, it will.

Let's circle back to Abe's case. Abe makes me a promise, but I do not expect him to fulfil it. Abe knows this, but still decides to make the promise. My scepticism is not necessarily motivated by a negative judgement of his moral character. For instance, I might think that Abe will not come because his tyrannical boss may call him at any time, threatening to fire him if he is not available. Principle F requires the promisee to obtain a factual assurance: the promisee is led to expect that the promisor will do X. In *Coming to Believe*, it is implausible to say that I value the assurance I have been given. I am not factually assured: I do not expect, when I accept the promise, that Abe will come. How, then, to explain my claim against Abe? One possibility

would be to modify F, saying that Abe nevertheless incurs the promissory obligation if he has reason to believe that, having accepted, I expect that he will do what he has promised. But this would be too small an accomplishment. I want to argue for the stronger claim that even if it were common knowledge that I do not expect him to come, Abe can still promise me to come.

It does not look like either of the two assurances we have seen (desired and unconditional) do the job here. However, I submit that Abe can *assure* me of something if I take him seriously. Abe can assure me that if he does not come, then he will be wronging me (unless I consent to him not coming). When Abe says, ‘Okay, well, I know you don’t expect me to come, but I promise I will be there’, he is also leading me to believe that if I change my mind about him or the situation, I will have the ‘opportunity’ to count on that promise, since he has an obligation towards me to come which is independent of what I expect him to do.¹³ Abe gives me what I call a normative assurance. I define ‘giving a normative assurance’ as follows:

A gives B a normative assurance if and only if A intentionally leads B to believe that if A fails to do X, then A will be wronging B (unless B consents to A not doing X).

I think we can consider a further promissory principle, a principle that is still based on beliefs, but on normative ones. Consider:

Principle of Normative Fidelity (NF): If (1) A intentionally and voluntarily leads B to believe that if A fails to do X, then A will be wronging B (unless B consents to A not doing X); (2) B freely accepts this normative assurance; (3) A acts with the aim of providing this normative assurance, and has good reason to believe that he or she has done so; (4) B knows that A has the beliefs and intentions just described; (5) A intends for B to know this; and (6) B knows that A has this knowledge and intent; then, in the absence of special justification, A must do X unless B consents to X’s not being done.

¹³Abe’s insistence does not make him more likely to come, but it has the role of settling the matter on the fact that he really intends to promise. Often, potential promisors and promisees do not want to be in a promissory bond with no hope of its requirements being fulfilled. For this reason, promises from unreliable promisors may be ignored, and unless the unreliable promisor insists, it may be commonly accepted that nothing as robust as a promise was made in the end.

In contrast to Principle F, we need to insert a condition, (2^{NF}), which says that B must freely accept this normative assurance. Principle F does not need an ‘acceptance clause’. In cases that Scanlon can explain the acceptance is *pre-emptively given*, since the promisor *knows* that the promisee wants to be assured that A will do X. In our case, however, it is important to make a distinction between cases where the assurance is not only undesired but unaccepted, and cases where it is undesired (which is compatible with giving a normative assurance) but accepted. Cases of the first kind are illustrated by *Creepy Promise*:

Creepy Promise: A male boss walks into the office of one of his subordinates and says: ‘You may not want to now, but I promise you, if you change your mind, I will have sex with you.’

It is reasonable to think that the subordinate does not think it is an opportunity and does not want to have this assurance. If we did not include an acceptance condition, we would have to say that the subordinate could only consent to non-performance.¹⁴ This is undesirable since the subordinate wants the possibility to reject the offer in the first place. The idea is that it is not always true that normative assurance has value for the promisee (i.e. *Creepy Promise*), while it seems more plausible that accepted normative assurance is always valuable for the promisee – if acceptance is genuine. Here ‘valuable for the promisee’ is to be understood in subjective terms: if acceptance is genuine, then the promisee is okay with having this opportunity (even if they could be misguided).¹⁵

¹⁴This acceptance-talk commits me to the view that a promise is complete only if the promisee accepts the promissory offer. Often people accept a promise by saying things like ‘Okay’, other times silence will suffice. In any case, acceptance seems to involve some form of consent and its communication. One might worry that my perlocutionary strategy cannot be adopted to explain the promisee’s acceptance, since consent is often assumed to be a robust normative power. This may foster a *Revenge Argument*: if we have to assume the existence of normative powers to explain the consent required to accept a promise, why not do the same to explain promising – retaining the simplicity of NPAs? As I suggest in the conclusion, however, my strategy may be naturally adopted to explain other so-called ‘normative powers’ – including consent. In §4.1., I will say more on why I consider NPAs untenable.

¹⁵There are difficult cases, such as threats. Responding to something like ‘I promise that you will regret it!’ with ‘Oh, yeah, bring it on then,’ may amount to acceptance. Suppose that acceptance can be genuine in such cases. For example, B judges the promise-threat in his own interests, since, say, he has a great desire to make A pay for it. Even if we meant that the recipient of the threat is here ‘accepting’ the promise-threat, the ‘promisor’ would not have an obligation to carry it out. A promise does not bind absolutely: if the promise binds us not to reconsider our intentions simply because of our convenience, it does not bind us independently of the cost to ourselves or others. For more on this issue, see (Scanlon 1998: 199–201).

Is Principle NF a principle that no one could reasonably reject? I think so. As I have already mentioned, this is what I think Abe is assuring me of: ‘You are not counting on it now, but you have the opportunity to do so if you change your mind: I am obligated to come regardless of whether you think I am coming.’ In *Coming to Believe*, I take this opportunity. Coming to believe that Abe is responsive to his obligations (e.g. I start to positively judge his moral character), together with the fact that he has given me a normative assurance (i.e. I *believe* that he has an obligation towards me to come), leads me to take a genuine interest in the promise, to ‘remain assured’ in Scanlon’s sense as well: I start to believe that Abe will help me and I plan my life accordingly. A normative assurance, whether or not I expect Abe to help me, leads me to believe two things: (a) Abe has an obligation towards me to help me; and (b) I have the *opportunity* to take an interest in the promise if I change my mind, because Abe’s obligation does not depend on my expecting him to comply. This opportunity has value as such, *qua* opportunity, and I would not have it if Abe were not really obligated to help. So, I have no good reason to reject a principle that binds Abe in this way if I accept his normative assurance.

In this picture, the wrongness of unjustified promise-breaking lies in (a) disregarding the interests that the normative assurance has directly nourished in the promisee (i.e. in *Coming to Believe*, the interests that I develop when I ‘change my mind’); and (b) in the absence of these interests, in depriving the promisee of an *opportunity* the promisor has convinced them they have: an opportunity that has value *qua* opportunity. To make (b) clearer, consider this: if Abe phones me ten days before the exhibition and tells me that he is not coming, then, although I had taken no interest in the promise, he is depriving me of the opportunity he led me to believe I had. This is like a gift, we might say, that is taken away after being given. The fact that the recipient does not value the gift is not a good reason to take it away: when I give a book as a gift, I do not have the right to take it back — even if the recipient does not like it or does not read it. Promising, I suggest, is analogous to gift-giving.¹⁶ In a variant of *Coming to Believe* — *Coming to Believe** — I do not change my mind and I am sure Abe will not come. Moreover, suppose that the gallery issues a policy that says that

¹⁶Raz (2014) has much in common with the idea that the promise gives, like a gift, an opportunity to develop an interest in the promise. Unlike Raz, however, I reject the idea that promising is a normative power to constitute obligations by promissory utterance. More on this in §4.1.

I can appoint only one person to help me with my *Joseph Roulin*. The opportunity that the normative assurance gives me is no longer valuable: not knowing whether Abe will come might make me stressed, and it would be more convenient to appoint another person. Principle NF provides for this too, giving me the option to terminate the promissory obligation at will. In yet another variant — *Coming to Believe*** — where I am sure that Abe will break the promise, I accept the promise to prove Abe's unreliability. In this case the opportunity is taken: we could say that I value the normative assurance because it satisfies my interest in proving Abe's unreliability.

What can we say about potential promisors? The promisor may want to promise even if she cannot generate factual expectations in the promisee. Perhaps Abe wants to show me that my factual expectations that he would come would be adequate if I had them, that I was *wrong* to be sceptical. Or Abe wants to strengthen our friendship, and gaining my trust is a good way to do so. But Abe has other reasons for wanting to promise. If I countered Abe's promise with a 'No way!', his coming to the exhibition might have a different, negative effect on our relationship. Not only could I not blame him if he did not come, but more than that, it is not certain that his appearance would give him the 'moral credit' he would get if he was successful in making a promise. If he came on the day of the exhibition, declaring that he wanted to help me, I might resent him for forcing his presence into my life in that way. Similarly, a friend who promises to come out to dinner, whose promise is rejected on similar grounds, cannot expect other friends to reserve a spot in the car for her. When we accept a promise, we take on certain commitments. I play a role in bringing about that Abe would wrong me if he did not come and, at the same time, I commit to the fact that if he does come, I would have to let him help me install the *Joseph Roulin*. Therefore, I think there are no good reasons for potential promisors to reject Principle NF.

If what does the work is an interest in having some kind of opportunity, how can we draw a difference between NF-based opportunities and cases where Abe gives me independent reasons to convince me that I have the opportunity to rely on the fact that he will come (e.g., he persuades me that he loves helping with museum exhibitions)? In such cases, Abe does not seem bound robustly as promises require, but in a sense analogous to what is suggested by Scanlon's Principle of Loss Prevention:

Principle of Loss Prevention (L): If one has intentionally or negligently led someone to expect that one is going to follow a certain course of action, X, and one has good reason to believe that that person will suffer significant loss as a result of this expectation if one does not follow X, then one must take reasonable steps to prevent that loss.¹⁷

The crucial difference is that the interest underlying Principle NF is an interest in having *unconditional opportunities* – opportunities whose value does not depend on the factual assurances that invitations to rely produce under propitious circumstances. Quite differently, the opportunity Abe convinces me of having by solely triggering Principle L is *conditional* on Abe not being in a position to rule out that, if he doesn't help me, I will be subject to a great loss. This explains why NF-based obligations are more robust than L-based obligations.

3.3 Towards a Unified Account

I would like now to suggest that Principle NF is the principle that explains all promissory obligations. To do this, I intend to show that NF can explain even cases that F can explain. In addition, I want to show how NF explains cases such as *Coming to Care*.

Let us circle back to the case where I promise to sell you the horse, and you expect me to sell it to you. Such cases are 'standard', as Principle F explains them easily. Principle NF has no problem explaining them either. In such cases you take the opportunity and do so straight away: you develop an interest in the promise immediately. Arguably, you think that I am motivated by my obligations. In fact, in optimal circumstances, normative assurances give rise to factual assurances. If I believe that a promisor has an obligation towards me, then, if I know that the promised act is feasible and I have reason to believe that the promisor regularly keeps (what is commonly accepted to be) their obligation, I will develop a factual assurance as well. I could believe that the promisor will do X because I believe that they are afraid of being blamed, or because I believe that they are genuinely motivated by the moral obligations that they believe they have / that others believe they have, and so on. The list is not exhaustive, and whether it is one case or the other often depends on the context. If we

¹⁷Scanlon (1998: 300–301).

think of *Coming to Care*, it is easy to see why Kathi can develop a factual assurance that her aunt will help her. Since Kathi thinks her aunt is motivated by her obligations to her niece, Kathi develops a factual assurance directly – which then becomes valuable when things turn out well for her.

One might wonder what interest is harmed by promise-breaking, and how an NF-based story relates to Scanlon's EA. On the one hand, in cases where the complex of assurances (normative and factual) is generated, the main interest involved is the one in being factually assured that the promisor will perform, pretty much like Scanlon's view suggests. It is undeniable that the value of a promise is all the greater the more valuable courses of action we can plan on its expected fulfillment, and the more peace of mind we can get from the factual assurance it generates. On the other hand, in cases like *Coming to Believe* and *Coming to Care*, if Kathi and I haven't changed our mind, the main interest harmed by promise-breaking is the one in having an unconditional opportunity to take an interest in the promise.

4 Some Difficulties and Solutions

4.1 Fundamental Difficulties

When triggered, Principle F gives rise to moral obligations to perform. Principle F suffers from the *Circularity Objection*, which can be solved, as the HA does, by reintroducing the role of the social practice. The social practice (Humean or Rawlsian) motivates the expectation that the promisee will do what she has promised, thereby triggering F. The question for my proposal is this: what motivates the normative belief that the promisor has an obligation towards the promisee to do as promised? The answer cannot come from the fact that the promisor is morally obligated by saying, 'I promise to do X'. In that case, what motivates the normative assurance would be what the normative assurance is supposed to explain, i.e. the directed moral obligation. It looks like the *Circularity Objection* previously concerned the factual expectations about the future conduct of the promisor, while now it concerns the normative belief that giving a normative assurance is supposed to generate in the promisee.

Let us look at why 'traditional' approaches do not save Principle NF from the *Circularity Objection*. First strategy: maybe what rescued Principle F can also rescue

Principle NF. One may say that when making a promise to me, the promisor is incurring an obligation not to free-ride on a just practice, and that she would be generating in me, and me alone, the belief that if she does not keep the promise, she would be wronging me. This may seem like a plausible strategy, but it comes up short, because it does not make sense of the *Directedness Intuition*. At best, the promisor generates in me the belief that should she not do X after she promised to do X, she would be wronging me and all who participate in the just practice of promising. It seems to me that if the promisor intentionally told a bystander about the promise, then she too would have the relevant ‘normative belief’, and she too may have the ‘right to the performance’ that should be exclusive to the promisee. We would only have an apparent vindication of the *Directedness Intuition*. As it stands, NF cannot benefit from this strategy.

The other option is to do without beliefs in explaining the obligation. This is what the Normative Power Accounts (henceforth, NPAs) theorists suggest.¹⁸ They say that we have a special moral power to invoke moral obligations by promissory utterance. When I say: ‘I promise to do X’, I constitute a directed moral obligation to perform. If NPAs are right, the role of beliefs can be set aside since we do not need them to explain the promissory obligation. The *Circularity Objection* would no longer have any weight. I will not rely on such a strategy, however, since I believe that NPAs are unacceptable, and for at least two reasons. First, it is indeed mysterious to explain how it is possible to have a power to create a moral obligation by fiat. Hume described the process by which ‘a certain form of words, along with a certain intention, changes entirely the nature of an external object, and even of a human creature’ to be ‘one of the most mysterious and incomprehensible operations that can possibly be imagined’.¹⁹ Second, it seems to me that NPAs give a distorted picture of what is *wrong* with unjustified promise breaking. NPAs postulate the power and say: it is wrong to break the promise because it entails violating a directed moral obligation that we have intentionally constituted. But, as Wallace well puts it, the complaint of the promisee is not so much that the promisor ‘has flouted some claim or other, but that the [*promisor’s*] action shows a lack of consideration for the first-order interest at

¹⁸See, for instance, Shiffrin (2008), Owens (2012), Raz (2014), and Bruno (2020).

¹⁹Hume (2003: 524). For a similar objection, see Prichard (1940/2002: 257).

stake in the assignment of the particular claim at issue'.²⁰ In our case, I suggest, it is the promisee's interest in having the *unconditional opportunity* that the normative assurance leads her to believe she has that justify the assignment of a claim to her. Theories such as Scanlon's and mine can be considered 'perlocutionary accounts' of promising. The directed moral obligation is explained by the effect that the promise causes in the promisee, not by the exercise of a sui generis moral power by the promisor.

4.2 Solutions

Let us explore two possible ways of dealing with the *Circularity Objection* for NF. The first is a 'radical conclusion' and entails the acceptance of a kind of error theory: NF explains ordinary promises, because the common or folk view is, in fact, an NPA-based view. People believe that promises constitute directed obligations, and this conviction is sufficient to generate NF-based directed obligations. The cost of this conclusion is that those who are convinced by my story will no longer be able to accept promises, because they will not believe that promisors can give them normative assurances. However, they may still make promises by deceiving potential promisees. This is a quasi-illusionist theory of promising: 'quasi' because NF is, in fact, morally binding; 'illusionist' because, to trigger NF, the promisee must believe that potential promisors have something like normative powers.

Is it in fact the case that someone who was convinced by my story can no longer receive promises? One possibility remains open. Scanlon's Principle F is a valid principle, and so is the appeal to Rawls-style practice-based obligations to avoid the circularity for F. Consider Seth and John, who have read this essay and are convinced by our conclusions so far. Seth might say, 'I don't believe that you can give me a normative assurance by promising', to which John might reply, 'Look, I don't think so either, but I can still promise you that I will do X, triggering a Rawls-style practice-based obligation, and now you have a reason to expect that I will do X, triggering Scanlon's Principle F'. These would be the promises that we could still receive, even when we realize that NF cannot be triggered. The HA would be the best we can get (with its extensional flaws), and ordinary people could make promises as we intuitively understand them only by virtue of a shared illusion. This does not imply that a story that

²⁰Wallace (2019: 169).

brings in NF does not constitute progress. This may well be the story that explains the dilemma of promising: either you have *Directedness* and adequate extensionality (provided that the illusion of normative powers is alive), or you have *Directedness* and no illusions, but a theory that is not extensionally adequate (i.e. the HA, which fails to preserve the *Directedness Intuition* in cases like *Coming to Believe* and *Coming to Care*).

Formerly, and for some time, I was convinced that the ‘radical conclusion’ was the best we could get. Now, however, I wish to offer the more adventurous reader an ‘optimistic conclusion’. I have come to believe that it is possible to respond to the *Circularity Objection* with a light refinement of Principle NF. This refinement will allow us apply the hybrid strategy, already employed by HA theorists, to avoid circularity. This strategy, as evidenced in §4.1, is unsuccessful when directly applied to our case: if promising entails a Rawls-style practice-based obligation, Principle NF may over generate. To be clear, the problem is as follows: when I tell C (a bystander) that I promised B that I would do X, Condition (1) of NF may be satisfied – and, potentially, the subsequent conditions as well.²¹ Recall Condition (1) of NF: ‘If (1) A intentionally and voluntarily leads B to believe that if A fails to do X, then A will be wronging B (unless B consents to A not doing X)’. Replace C with B, here, and imagine that A is telling C about the promise she made to B. Condition (1) may be fully satisfied because C could also consent to my not doing X – *as far as C is concerned*. My Rawls-style practice-based obligation, as understood here, is owed to C, B, and to all who participate in the fair practice of promising. C’s consent to non-performance would count for little – in terms of relaxing the normative pressure generated by my promise to B – but the fact that C could do so confirms the unfortunate prediction that NF could generate a directed obligation towards C if I told her of my promise to B.

An easy fix might be waiting in the wings. We can rescue the HA’s strategy for

²¹One may suggest that Principle NF’s condition (2) – i.e. ‘C[C/B] freely accepts this normative assurance’ – blocks these results. In one natural reading, however, NF may be triggered in cases where C thinks that A’s Rawls-style obligation depends on C’s acceptance, which C provides. As I see it, this is a problem as long as we do not have a fully worked out theory of what it takes to accept a promise – a theory that plausibly disqualifies C’s ‘acceptance’ as genuine acceptance. For now, it is sensible to make the acceptance condition explicit in the notion of giving a normative assurance, as I shall do in a moment. This is much easier to do than to offer, in the space I have left, a theory of acceptance that satisfies the desiderata of my general project. I hope to develop a satisfactory account of what it takes to accept a promise in future work.

NF by placing a constraint on the definition of giving a normative assurance such that it is necessary that ‘A leads B to believe that, *if B accepts*, then, if A fails to do X, A will be wronging B’. If this modification is welcomed, it is not sufficient for A to inform C about the promise made to B for A to trigger NF with respect to C. The ‘if C^[C/B] accepts’ component of condition (1) would not be satisfied, given that C is not asked to accept anything. Indeed, A is already bound to C by the practice-based obligation – and telling C merely informs her that A is obligated towards C, B (also in NF’s sense), and to all who participate in the practice of promising. Thus, NF would not be activated with respect to C. Let us consider a more precise definition of what is required to give a normative assurance:

A gives B a normative assurance if and only if A intentionally and voluntarily leads B to believe that, *if B accepts*, then, if A fails to do X, A will be wronging B (unless B consents to A not doing X).

In light of this, let us consider a minimally tweaked version of NF:

Principle of Normative Fidelity (NF)*: If (1) A intentionally and voluntarily leads B to believe that, *if B accepts*, then, if A fails to do X, A will be wronging B (unless B consents to A not doing X); ... *and thereafter as for NF*.

NF* plausibly allows us to derive the benefits of the HA in explaining how promises can be regularly given. As in the hybrid view developed by Kolodny & Wallace, the promisor has two obligations: the first is a Rawls-style practice-based obligation towards all involved in the practice of promising; the second is an NF-based obligation that A has only towards B, if B rests normatively assured, thereby vindicating the *Directedness Intuition*. This second obligation is explained by the normative assurance, which is in turn explained by the Rawls-style practice-based obligation: B has reason to believe that, if she accepts, then, if A fails to do X, A will be wronging B, among others, if B has not consented to A not doing X.

We can then explain how we can bind ourselves to a bystander by telling her of a promise made to someone else. If I tell C of my promise to B to come to dinner, with the commonly acknowledged intention of leading C to believe that I will come to

dinner (e.g. C wants me to be there), then I may have a Principle L-based obligation towards C. The opportunity I convince C of having would not be unconditional, but conditional only on my not taking reasonable steps to prevent C's great loss if I did not come.

4.3 Applying the Account: Other Concerns

At this juncture, it is appropriate to address other important concerns regarding the implications of my account. I start with the issue of False Positives. Suppose that A intentionally and voluntarily leads B to believe that, if B accepts, then, if A fails to do X, A will be wronging B (unless B consents to A's not doing it) – *although this is not in fact the case*. I call this the False Positive issue because there appears to be an obligation where there should not be. However, as I shall demonstrate presently, different interpretations of this worry yield different results.

In some cases, A is not bound on the grounds that the promise conflicts with other moral principles that take precedence. For example, if a wife promises her husband that she will have sex with him in the evening, even if the husband accepts, she is not obligated, because, as discussed in footnote 15, promises are not absolutely binding. As Scanlon argues, promises do not override the potential costs to us and to others. In this case, the cost of giving up one's sexual freedom is arguably too high, and for this reason the principle that prevents one from giving up one's sexual freedom takes precedence over promissory principles.²² Although such cases raise important questions for any theory of promissory obligations, I wish to focus now on cases that present particular challenges for my proposal.

Another approach may be adopted in interpreting the objection, whereby the scenario is similar, but no conflict arises with other moral principles that take precedence. Suppose there is no social practice of promising. Suppose Leonardo has a personal code of honor. This code has several rules, including: 'R1. If I don't do what I say I will do, and you accept, then I will wrong you (unless you consent to my not doing it)'. Additionally, suppose that, if invoked, this honor code does not generate any Rawls-style practice-based obligation, since Leonardo is the only participant in this practice. Suppose also that this is true: 'If Leonardo invokes R1, Leonardo's

²²Here I am following Scanlon (1998: 199–201).

sets of moral obligations remain the same.’ That is, we rule out, by assumption, the idea that Leonardo is exercising a real normative power by invoking the code, and so we grant that *it is not the case that, by invoking it, he will wrong anyone if he fails to perform*. Leonardo attempts to convince Sofia that, based on his honor code, if he does not water her strawberry garden, and she accepts, then he will wrong her. Sofia has no reason to believe this. As we have discussed, without the Rawlsian mechanism in place, NF* is circular. However, Sofia ultimately believes that Leonardo has an obligation to her and accepts Leonardo’s offer. One might correctly conclude that Leonardo is not promising, because he is invoking his personal code of honor and not the social practice of promising. Would Leonardo nonetheless be bound to water Sofia’s strawberry garden? I think so. Principle NF* may be triggered even if no promise has been made. As we have seen, one may trigger Scanlon’s Principle F by listing several reasons that she has for doing X. Principle NF* is analogous, as it is practice-independent. However, to trigger NF* in non-promissory ways, one must appeal to the possibility of changing the normative situation through other means (e.g. honor codes such as Leonardo’s). If Sofia believes Leonardo, I see no reason why he should not have an NF*-based obligation towards her: after all, he is convincing her that she has an unconditional opportunity to take an interest in the fact that he is going to water her strawberry garden.

There is a third, more interesting way in which the objection raised by False Positive can be cashed out. Suppose I break my friend’s vase and convince her that I did so negligently. In this case, it does not seem that I incur an additional NF*-based obligation, even though I seem to be giving a normative assurance: by taking the blame, I seem to be assuring her of the fact that, if I do not compensate / apologize, I will be wronging her. In these cases, however, I am giving not a normative assurance but rather a factual assurance. By virtue of the moral norms commonly accepted, if I convince someone that I have acted in a certain way, then that person will believe the normative situation to be a certain way. The normative situation changed with my negligent conduct (which led to the vase’s breakage), and my factual assurance makes it salient that a disjunction of the descriptions that entail my responsibility for the breakage is true. Furthermore, it is hard to imagine how NF* could be triggered once the acceptance condition is made explicit in NF*’s condition (1). When I assure my

friend that it is my fault, I am not doing so on the condition that she *accepts* my normative assurance: If I say to her, in this scenario, ‘If I don’t repair the vase, I will be wronging you,’ I am presupposing that I have acted in a way that entails that I have an obligation to compensate, irrespective of whether she accepts. At most, my friend may concede that I do not need to compensate her for the breakage.

This does not rule out the hypothesis that NF* may be triggered in non-promissory ways. For one – the above example of Sofia’s strawberry garden illustrates just that. Nor do I mean to suggest that a promisor cannot give a normative assurance without saying ‘I promise’. Exasperated by my opposition, Abe might go so far as to say, ‘Even if you don’t believe that I will come, if I don’t come to the exhibition, I will be wronging you.’ If I accept Abe’s normative assurance, it appears that he is making a promise to me, and, arguably, the social practice of promising allows for this. What is important to note is that cases like my friend’s vase involve factual assurances instead of normative assurances. The fact that my friend develops the belief that I have an obligation to her is parasitic on the factual belief of which I have convinced her of. This factual belief, together with the commonly accepted norm that one must repair what one breaks, leads her to believe that I have an obligation to compensate.

Here is a last worry, False Negatives. Suppose that A and B live in a strict caste society, and due to internalized feelings of inferiority, B does not believe that she could hold any rights against A, including promissory rights to performance. So, when A promises B to do X, B does not rest normatively assured. Thus, NF* is not triggered. Is this bad? A critic might say that it is, for we would like to say that there is a directed obligation in these cases: A is bound to B only. Let us proceed in order. If one accepts the ‘radical conclusion’, the reply is straightforward: A is not bound. It is at least necessary for B to believe that A has an obligation towards her in order for B to even consider the unconditional opportunity represented by A’s offer. If one goes for the ‘optimistic conclusion’, things get more interesting. Does the social practice of promising, in the strict caste society, give A–(superior) the right to promise B–(inferior)? If it does and B remains in disbelief, A would only have a Rawlsian practice-based obligation to do X. If it doesn’t and B remains in disbelief, A would not be bound – as for the ‘radical conclusion’. These seem reasonable predictions. Two natural considerations follow. First, the EA and HA do no better than my pro-

posal in such cases, if by better one means explaining a directed obligation that binds A to B only. Second, I doubt that we have good reasons here for wanting a directed obligation that binds A to B only. If B does not even consider the opportunity of taking an interest in the promise, how could A harm B's interests by failing to perform? So, we should accept the conclusion that, *at most*, A would be free-riding on a just social practice if she failed to perform.

Let us take stock. First, we developed an extensionally adequate theory of promising without appealing to a robust conception of normative powers. Even with our hybrid strategy (Rawls' PA + Principle NF*) borrowed from Kolodny and Wallace, we are not committed to the existence of normative powers. Second, with respect to Kolodny and Wallace's HA, my proposal succeeds in doing justice to the *Directedness Intuition* even in cases where the promise does not generate factual assurances, and/or the promisor knows that the promisee does not desire to be so factually assured. Third, through the hybrid strategy and the refinement of NF, we can address the circularity worries for my proposal. This way, both the *Necessity Objection* and the *Circularity Objection* are accommodated.

In closing, I want to suggest that my strategy may be adopted to explain other moral practices. To give an example, it could be argued that, by forgiving, one convinces B that she will have the unconditional opportunity not to be blamed for what she is being forgiven for. The obligation not to blame those we have forgiven may be explained analogously to promissory obligations, and this without positing the existence of normative powers, nor having to rely on factual expectations, which often entail extensional inadequacy. Theories based on the notion of normative assurance – and / or versions of it adapted to the cases under study – may prove to hold promise for a general theory of the morality of our social practices.²³

References

1. Bruno, D. (2020) 'Trust-Based Theories of Promising,' *The Philosophical Quarterly* 70/280: 443–463.

²³I am especially grateful to Johann Frick, Michal Masny, Niko Kolodny, Crescente Molina, Ezequiel Monti, David Owens, Joseph Raz, Tim Scanlon, Paulina Sliwa, Jay Wallace, and to two anonymous referees. I would also like to thank Darwin College, Cambridge, and the University of California, Berkeley for their generous support, in the form of a D. Hugh Mellor Philosophy Studentship and Berkeley Fellowship, respectively.

2. Friedrich, D. and Southwood, N. (2009) 'Promises Beyond Assurance', *Philosophical Studies*, 144: 261–80.
 3. Heuer, U. (2012) 'Promising-Part 1' *Philosophy Compass* 7: 831–851.
 4. Hume, D. (2003) *Treatise of Human Nature*. Mineola, NY: Dover Philosophical Classics.
 5. Kolodny, N. and Wallace, J. (2003) 'Promises and Practices Revisited', *Philosophy & Public Affairs*, 31: 119–154.
 6. Melenosvky, C. M. (2016) 'Promises, Practices and Reciprocity', *The Philosophical Quarterly*, 67/266: 106–126.
 7. Owens, D. (2012) *Shaping the Normative Landscape*. Oxford: OUP.
 8. Pratt, M. (2003) 'Promises and Perlocutions,' in Matt Matravers (ed.) *Scanlon and Contractualism*, 79–102. London: Frank Cass & Co., 2003.
 9. Prichard, H. A. (1940/2002) 'The Obligation to Keep a Promise', in J. MacAdam (ed.) *Moral Writings*, 257–65. Oxford: Clarendon Press.
 10. Rawls, J. (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.
 11. Raz, J. (2014) 'Is There a Reason to Keep a Promise?', in G. Klass, G. Letsas and P. Saprai (eds) *Philosophical Foundations of Contract Law*, 58–77. Oxford: OUP.
 12. Scanlon, T. (1998) *What We Owe To Each Other*. Cambridge, MA: Harvard University Press.
 13. Shiffrin, S. (2008) 'Promising, Intimate Relationships, and Conventionalism', *Philosophical Review* 117(4), p. 481–524.
 14. Wallace, J. (2019) *The Moral Nexus*. Princeton N.J.: Princeton University Press.
- University of California at Berkeley, USA*