

Filippa Ronquist
filippa.ronquist.18@ucl.ac.uk

‘Trust me’: Conventionalism and the Directedness of Promissory Obligation

Abstract: Conventionalist accounts of promising hold that our ability to make promises to each other is, in some fundamental sense, dependent on a social practice or convention of promising. These accounts of promising typically understand the wrong of breaking a promise as the wrong of undermining or free-riding on a useful social practice. This appears to jar with our common understanding of the primary wrong involved in breaking a promise: breaking a promise, we tend to think, wrongs primarily the person to whom the promise was made, not the people who participate in the social practice of promising at large. In this article, I develop an account of promissory obligation that shows that this common critique of conventionalist accounts of promising fails.

1. Introduction

To make a promise is to voluntarily undertake an obligation, by communicating an intention of doing so. In the right circumstances, communicating such an intention creates an obligation to do the thing promised. This obligation is owed to the promisee, and only the promisee can release the promisor from her obligation.

Or so we seem to think. The above characterisation, I take it, accurately captures what people take themselves to be doing when they make promises to each other. Joseph Raz succinctly formulates this account of promising in what he calls the “Obligation Conception” of promises:

The Obligation Conception. “Whoever communicates, in circumstances of kind C, an intention to undertake by the very act of communication an obligation to perform an action and invest the addressee with a right to its performance ought to perform that action and his addressee has a right that he shall do so (unless the addressee releases him from this requirement)” (Raz 1977, p. 211).

The Obligation Conception captures the important insight that, when we make a promise, we understand ourselves to be undertaking an obligation, and not simply communicating an intention, however firm, to act in a certain way. The communication of

an intention, even if done willingly and for the express purpose of assuring another of one's future action, does not amount to making a promise. From here on, I'll take it for granted that the Obligation Conception correctly describes the central rule of promising.

The important question to ask about promising is whether the Obligation Conception can be vindicated as a moral principle. Why think that we, through acts of communication, can create moral obligations in the way described? Raz's own, deceptively simple answer to this question is: *because it is desirable that we should have the ability to do so*. We can explain and justify our 'normative power' to make promises, Raz thinks, by appealing to the value of having such an ability.¹

Raz, I think, is right to suggest that we wouldn't be able to create obligations at will if it weren't, in some sense, valuable for us to do so. But the explanation he offers of the grounds of promissory obligation – that we can create such obligations because it is valuable for us to be able to do so – has never seemed entirely satisfactory to me. And, in fact, Raz himself was not entirely satisfied with this explanation, either. In a late article, Raz suggests that, contrary to what he earlier claimed, the value of having the normative power to promise isn't sufficient to explain why we have an obligation to keep our promises (Raz 2022).² My dissatisfaction, however, is not Raz's own. Rather, my dissatisfaction stems from the fact that the value of a normative power doesn't seem sufficient, on its own, to establish its existence. Raz himself acknowledges that this might be so. There might be additional conditions which need to be in place for normative powers to exist (Raz 2022, p. 169). But Raz does not spell out what these additional conditions might be. To me, this seems a major drawback of his account. It seems as if there could be variations on our normative power to promise that might be valuable to us in much the same way that our 'standard' way of promising is. Perhaps the best way to illustrate this point is to consider the fact that we already recognise different rules of promising in different contexts. For example, promissory rights and obligations in commercial contexts can often be

¹ This value might be intrinsic or instrumental. If having the normative power to promise did not have intrinsic value, the existence of the normative power would depend on whether having such a power would, overall, lead to better outcomes. While this claim drives both Hume's and Rawls's accounts of promising, it isn't obvious to me that Raz's account relies on it.

² This is because the normative case for possessing the power to promise doesn't directly and obviously translate into a normative case for *keeping* one's promises. Raz's solution is to argue that our normative power to promise is more limited than we might think (and more limited than his previous writings suggest). We do not have the normative power to make promises that in no way serve the interests, broadly construed, of promisees. It follows from this that we do not have the normative power to make promises that there is no reason to keep.

commodified, that is, bought or sold to third parties. This is not allowed by the rules of promising that apply in intimate contexts. Of course, the commodification of commercial promises is typically recognised by law, and hence is part of our legal power to contract as opposed to our normative power to promise. But imagine a society which doesn't have laws which allow for the commodification of promissory rights and obligations, and which also doesn't have any social practice to that effect. In such a society, a normative power to make 'commodity promises' might be tremendously valuable. Should we conclude that the people in this imagined society have the normative power to make commodity promises, because it would be valuable for them to have it? Or do they only have our 'standard' power to promise? To me, it seems that we could not speak intelligibly of people in this context as having a normative power to make 'commodity promises' (nor 'standard' promises, for that matter) unless and until these normative powers were, in some sense, recognised and put into practice.

Conventionalist accounts of promising are appealing in part because they provide a straightforward answer to the question of why it should be possible to will obligations into existence in the specific way that promising allows, but not in other ways that we might imagine. The fact that we can voluntarily incur promissory obligations is explained by the fact that we have a social practice of promising which makes it possible for us to do so. The very possibility of making promises – and the intelligibility of our normative powers, more generally – depends, for the conventionalist, on a social practice or convention of some kind.³ While conventionalist accounts of promising, most commonly associated with David Hume and John Rawls, agree with Raz that our ability to make promises to each other ultimately depends on the value of us having this ability, they tend to understand this value rather differently from Raz.

In this article, I offer a broadly conventionalist approach to explaining promissory obligation, in an effort to articulate what I think is a more satisfactory account of promissory obligation than the one offered by Raz. I take as my starting point the idea that promising is a cooperative social practice which delivers important benefits to its participants, and I follow Rawls in arguing that we can understand the wrong of breaking a promise as the wrong of free-riding on a useful social practice (although, as will become

³ It will be important for the conventionalist to explain in what sense, exactly, promising depends on a social practice or convention. Everyone, conventionalists and non-conventionalists alike, agrees that promising is a social practice, in the sense that it is a rule-governed activity that we sometimes engage in. Further, there can be little doubt that the manner in which we make promises – by saying "I promise," or by interlocking pinkie fingers – is a matter of social convention. In these ways, promising is rather trivially conventional.

clear, my understanding of what such free-riding amounts to differs in several important respects from Rawls's). I return to how my account compares with Raz's 'normative powers' account of promising in the final section.

A common objection to Rawls's account of promising – and a common objection to conventionalist accounts of promising more generally – is that they fail to adequately capture the wrong involved in promise-breaking.⁴ The wrong of free-riding on a useful social practice appears to be impartial between all those who contribute to (or perhaps participate in, or benefit from) the practice or institution in question. Because of this, the conventionalist explanation of why it's wrong to break a promise seems incorrect, or at the very least, incomplete. As the Obligation Conception articulates, promises are made to particular persons, who thereby gain a right to performance. Breaking a promise violates this right, and is hence a wrong committed against the promisee, not people in general. The wrong of breaking a promise seems a paradigmatic example of what Michael Thompson (2004) calls 'bipolar' obligation, or what Margaret Gilbert (2004) calls 'directed' obligation – that is, the kind of obligation that one person owes to another person. This kind of obligation is importantly different from an obligation which merely concerns another person, even where that obligation has the same content. (Compare: "I must not harm Jack" and "I owe it to Jack not to harm him.") Promisees are the victims – not mere collateral damage – of broken promises.⁵ This idea, furthermore, is firmly embedded in our practice of making and breaking promises. When we break a promise, we apologise to the promisee, and try to make it up to her, not to everyone else who participates in the promising practice. Even if we think that we wrong other contributors or participants in some way when we break a promise, there seems to be another, special way in which we wrong promisees.

To be clear, the objection against the conventionalist is not that she cannot recognise any ways in which breaking a promise harms the promisee in particular. A promisee can suffer various harms as a result of broken promises, such as reliance costs and dashed expectations, but such harms are in an important sense 'contingent,' and they are not involved in all instances of promise-breaking. Not all promisees expect, or come to rely on, performance. Further, a conventionalist can grant that the rules of our promising practice imply that promissory obligations are owed to promisees, and that these rules (implicitly or explicitly) give promisees a right to performance. But given that not all conventional rules are morally binding, the fact that the rules of our convention specify that

⁴ For two versions of this objection, see Scanlon (1998, p. 316), and Darwall (2013, p. 137-138).

⁵ As Thompson puts it, where there is a duty not to mistreat you, which is not owed to you, "You are the occasion, not the victim, of my fall" (2004, p. 340).

promisees have a right to performance cannot, by itself, establish that they do. What the conventionalist struggles to explain is what grounds the directed obligations specified by the rules of our promising practice, since the conventionalist account of why it's wrong to break a promise seems to leave no space for directed obligations of this kind. As a result, some conventionalists bite the bullet, insisting that our ordinary understandings of the wrongs involved in promise breaking are illusory. For example, Liam Murphy suggests that it is mistaken to think that breaking promises wrongs primarily the promisee (2020, p. 458). As far as I am aware, convincing attempts at explaining the directedness of promissory obligation on conventionalist accounts of promising are missing.⁶ In this paper, I try to develop such account.

The plan for the paper is as follows. In Section 2, I give a brief account of Hume's and Rawls's accounts of promising. In section 3, I explore the role played by trust in the promising practice, and argue that the trust the promisee places in the promisor is a *benefit* to the promisor, and therefore in effect a *contribution* that promisees make to the promissory relation. In Section 4, I discuss trust in greater detail. In Section 5, I argue that if we understand promissory obligation as owed to those whose cooperative efforts enable the promisor to benefit from promising, it will turn out that promissory obligations are, more often than not, owed to promisees and promisees alone. In Section 6, I return to the question of the role that the practice or convention of promising plays in my account, and return to the contrast between mine and Raz's accounts.

2. Hume and Rawls on promising

I will take as my starting point the idea that promising is a social practice sustained by our mutual efforts, which benefits those who participate in it in various ways. To understand what kind of social practice promising is, and what benefits it yields, I will first turn to Hume and Rawls.

In his *Treatise*, Hume argues that promising is 'naturally unintelligible' outside of any human practice or convention of promising.⁷ My main focus here will not be on

⁶ C.M. Melenovsky's (2017) argument is the only explicit attempt to solve the problem that I am aware of. But his solution is unconvincing, in large part because he doesn't pay sufficient attention to the kind of convention that promising is. His attempted solution ends up begging the question, since it relies on the idea that promisees have a moral right to promisors fulfilling their promissory obligations, without explaining why promisees have such a right.

⁷ More specifically, Hume argues for two related claims: first, that "a promise wou'd not be intelligible, before human conventions had establish'd it," and second, that even if it were intelligible, "it wou'd not be attended with any moral obligation" (2007, p. 331).

Hume’s natural unintelligibility claim, though I will return to it briefly in the final section. Though promising is ‘naturally unintelligible’ for Hume, it is perfectly intelligible, and capable of giving rise to moral obligations, within the context of a promising convention.⁸ For Hume, the very general problem that the practice of promising helps us solve is the ‘want of mutual confidence and security’ between people. Call this *the assurance problem*. It is important that we find a solution to the assurance problem, since a lack of confidence and security between people is severely disadvantageous for everyone. It is most obviously disadvantageous because it hinders people from undertaking mutually beneficial schemes of cooperation. It is frequently the case that one party of some cooperative scheme has to perform their part of the agreement before the other, and where the other party’s generosity or good will alone cannot be relied upon, we need some other way of creating assurance in order for such schemes to go ahead. Hume illustrates the problem with an example involving two farmers who are considering whether to help each other with their respective harvests. In his example, one farmer’s crops are ripe today, another farmer’s will be ripe tomorrow; neither can harvest all of their crops by themselves (2007, p. 334). It would be beneficial for both if they could agree to help each other out, but if the first farmer has no means of assuring the second farmer that he’ll perform his part of the deal tomorrow, he might not be able to get any help with his own harvest today. If so, both harvests will be lost. So pervasive are situations like this, Hume thinks, that without the invention of a promising practice, nothing short of the benefits of commerce would be lost among mankind.

The obligation to keep our promises, for Hume, derives from the great general utility of the promising practice. By sticking to the rules of the promising practice, we follow a set of rules such that, if everyone follows them, we will all be better off. The wrongness of breaking a promise is the wrongness of violating such generally useful rules. Hume’s explanation of why promise-breaking is wrong is, then, a paradigmatic example of what we might call a ‘two-level’ view of promising. On this kind of view, we first establish that some set of rules would lead to great benefits or utility overall if everyone (or most people) were to follow them. Then, in a second step, we argue that acting in breach of such rules is wrong. The rules of the practice are justified in relation to the practice’s

⁸ In *An Enquiry Concerning the Principles of Morals*, Hume defines convention as “a sense of common interest; which sense each man feels in his own breast, which he remarks in his fellows, and which carries him, in concurrence with others into a general plan or system of actions, which tends to public utility” (1987, p. 95).

overall justice and utility; the rightness or wrongness of particular acts is justified in relation to the rules of the practice.⁹

In the contemporary literature on promising, Rawls is the paradigmatic conventionalist. While Rawls does not speak of the ‘natural unintelligibility’ of promises, he argues that promises are a kind of institutional artefact, defined by a set of constitutive conventions. Much like a home run in baseball, a promise is defined by the rules of the institution of which it is a part. The idea is that we can hit a ball and run around a field, or say “I promise” or interlock pinkie fingers, outside of there being a social practice of baseball or promising, but we cannot *hit a home run* or *make a promise*. To hit a home run is to do whatever the rules of baseball define as a home run; to make a promise is to do whatever the rules of promising define as a promise. Rawls, like Hume, thinks of the primary function of this practice as facilitating schemes of mutual cooperation, and also offers a broadly ‘two-level’ view of promissory obligation. However, Rawls gives a specific interpretation of why it is wrong to violate the rules of a useful social practice. For Rawls, it is always an open question whether the rules of a given social practice or institution give rise to genuine moral obligations. Whether they do is determined by the principle of fairness:¹⁰

Principle of Fairness. A person is “required to do his part as defined by the rules of an institution when two conditions are met: first, the institution is just (or fair), that is, it satisfies the two principles of justice; and second, one has voluntarily accepted the benefits of the arrangement or taken advantage of the opportunities it offers to further one’s interests” (Rawls 1999, p. 96).

Rawls’s explanation of why breaking a promise is wrong, then, hinges on the fact that one is in general morally required to do one’s part in upholding just social institutions, as specified by the rules of such institutions, whenever one has voluntarily accepted the

⁹ In general, ‘two-level’ theories justify the rules of some practice in relation to the practice’s overall justice and utility; and justify the rightness or wrongness of particular acts in relation to the rules of the practice. Rawls provides a well-known account of ‘two-level’ views in morality along this line in his “Two Concepts of Rules” (1955).

¹⁰ Rawls takes the principle from Hart. In Hart’s formulation, the principle holds that “when a number of persons conduct any joint enterprise according to rules and thus restrict their liberty, those who have submitted to these restrictions when required have a right to a similar submission from those who have benefited by their submission” (1955, p. 185). For Hart, unlike Rawls, it is important that the principle apply even when benefits have not been voluntarily accepted, since Hart appeals to the principle to explain political obligation.

benefits of the institution. “We are not,” as he puts it, “to gain from the cooperative efforts of others without doing our fair share” (1999, p. 301). Promisors have an obligation to keep their promises because they have, in virtue of making a promise, voluntarily accepted the benefits of promising. By breaking their promise, promisors would be undermining a generally useful social practice, or free-riding on the cooperative efforts of others.¹¹ Since the principle of fairness specifies that a person must have voluntarily accepted the benefits of a practice in order to incur an obligation to contribute, it becomes important for Rawls to show in what way each and every promisor benefits from the practice. If some promisor manages to make a promise without voluntarily accepting any benefits of the promising practice, she will not be obliged to keep her promises. (At least not in a way that can be explained by the principle of fairness.)

In the rest of the paper, I will follow Rawls in appealing to the principle of fairness to explain promissory obligation, with two important qualifications. First, I will assume that the social practice of promising is sufficiently just or fair to satisfy the first condition of the principle without assuming that this is so because they satisfy Rawls’s two principles of justice.¹² Second, Rawls’s principle of fairness needs further qualification in order to be accepted. There are some social practices which meet the conditions outlined and nevertheless do not appear to yield any obligations for beneficiaries to contribute.¹³ Here,

¹¹ While it seems like only a subset of broken promises (at best) have the power to undermine the promising practice, all broken promises will involve free-riding (on the assumption that each and every promisor voluntarily benefits from the promising practice, in some way).

¹² The principle of fairness, I take it, is compatible with many different accounts of what makes a practice just or fair. Besides, it does not seem to make much sense to speak of the rules of promising as directly *satisfying* Rawls’s two principles of justice. (Are the rules of promising the kinds of rules that could satisfy the difference principle?) Perhaps what Rawls means is that the rules of promising are in some sense compatible with the two principles of justice, even if they don’t directly satisfy them. But even this weaker claim remains far from obvious.

¹³ As several critics, and most notably Nozick (1974) and Simmons (1979) have pointed out. Cooperative schemes come in many variations, and in each case, one must consider the details of the scheme in order to determine whether anyone might have an obligation to contribute to it. It will be relevant, in particular, to consider whether the benefits of the scheme outweigh the costs, including the opportunity costs, to each participant of the scheme, whether the benefits and burdens are distributed fairly, and what kind of good is provided by the scheme. If the good provided by the scheme is sufficiently trivial, or the scheme is sufficiently ill-advised, there may not be an obligation to contribute even if one voluntarily benefits. Further, if some one person or smaller group of people are happy to pay the whole cost of the scheme, since the benefits to them vastly outweigh their costs, or the good in question is a by-product of an activity that people would engage in anyways,

however, I rely on the principle of fairness only to make a specific claim about promising: a promisor who voluntarily accepts the benefits of the promising practice has an obligation to contribute to the practice by keeping her promises.

Before moving on, it will be helpful to pause on why Rawls's account of promising fails to adequately explain the directedness of promissory obligation. There are two ways of understanding the worry. Either one might think that the Rawlsian account doesn't give rise to any directed promissory obligations at all. Alternatively, and somewhat more plausibly, Rawls's account naturally suggests that promisors, as beneficiaries of the practice of promising, owe their promissory obligations to everyone who has helped contribute to the practice – everyone, that is, whose cooperative efforts the promisor would be free-riding on if she accepted the benefits of promising without contributing.¹⁴ On this reading, the problem is not that Rawls's account doesn't generate any directed obligations, but that it over-generates them, and to the wrong people. Do I really owe it to all promise-keeping strangers not to break my promise to meet my friend for lunch? Even more worryingly, if promises are owed to all contributors, but not to non-contributors, it seems as if the people to whom a promise is owed may not include the promisee herself. My friend might be a notorious promise-breaker, and if so, the Rawlsian picture suggests that I might owe it to lots of people, but not to her, to keep my promise.

3. Benefits and contributions

Although both Hume's and Rawls's accounts of promising provide a good starting point for thinking about the benefits of our promising practice as well as the kinds of actions necessary to sustain it, neither Hume's nor Rawls's accounts of what counts as a benefit

independently of whether other people contribute, it is typically out of the question to obligate all beneficiaries to contribute. See Arneson (1982) for further discussion of the conditions that must hold in order for the principle of fairness to generate obligations to contribute. As Arneson rightly points out, voluntary receipt of benefits isn't always necessary to generate obligations to contribute. This is especially important for those who, like Hart (from whom Rawls originally takes the principle), want to use it to justify political obligation. Here, I'll largely leave to one side the question about whether benefits need to be voluntarily accepted in order for beneficiaries to have an obligation to contribute, since promisors typically accept at least some of the benefits of promising voluntarily.

¹⁴ Rawls notes that the intuitive idea underlying the principle of fairness is that “when a number of persons engage in a mutually advantageous cooperative venture according to certain rules and thus voluntarily restrict their liberty, *those who have submitted to these restrictions have a right* to a similar acquiescence on the part of those who have benefited from their submission” (1999, p. 301, my emphasis).

and a contribution to the promising practice is complete. More specifically, I will suggest in this section that neither Hume nor Rawls pays sufficient attention to the role that trust plays in sustaining our promising practice.

In the case of some social practices, the answer to the question of what counts as a benefit and a contribution of the practice is fairly straightforward. Consider a system of tax-financed healthcare. The benefit of the practice is the healthcare that people get access to via the practice. The costs involved are what each person pays towards the scheme, as well as the labour and time involved in organising it. If I do not contribute anything myself, and yet benefit from the health care provided, the sense in which I gain from the cooperative efforts of others is clear. I benefit from the cooperative efforts of everyone who has paid towards the scheme, or contributed their labour or time. In the case of promising, however, the answer is not so simple.

In Hume's treatment of promising, we can distinguish between two sorts of benefits of promising. First, there are the direct benefits that an individual can get in return for a promise. In the case of the two farmers, this is the benefit of getting help with one's harvest, which is a direct result of one's ability to assure the other. But this benefit doesn't exhaust the benefits of the promising practice. Other people can also be expected to benefit, albeit indirectly, when this kind of mutually beneficial exchange goes ahead. The other villagers might benefit from an abundant supply of cheap corn at the market. This kind of benefit is doubtlessly more indirect, but no less important for that. As already noted, Hume thinks of this indirect benefit as nothing short of the benefits of commerce itself – in other words, a benefit not to be underestimated.

But if we, like Rawls, want to use the principle of fairness to explain promissory obligation, we must explain in what way *each* promisor who makes a promise voluntarily accepts the benefits of the promising practice. (Recall: on Rawls's interpretation of the 'two-level' view, the obligation to do one's bit to sustain a practice is triggered only if one voluntarily accepts the benefit(s) of the practice). And if we want to do this, we cannot appeal to the two kinds of benefits just mentioned. We cannot appeal to whatever benefit the promisor gets in return for his promise, since some promises are gratuitous: the promisor gets or expects nothing in return for making the promise. And we cannot appeal to the indirect benefits of living in a society with a stable promising practice, because such benefits are not voluntarily accepted. Rawls's account of promissory obligation must

suppose that there is some such benefit; in spite of this, his own account of what this benefit might be is both brief and unsatisfactory.¹⁵

We can do better than Rawls in this respect. To see what further benefits are involved in promising, it will help to consider the mechanism by which promising typically solves what I've called the assurance problem. Promising is just one of several ways in which we can assure other people of our future actions. If Anna wants to assure her partner that she will begin to see a therapist, she might tell him how much she wants to start therapy, that she has discovered her health insurance will pay for it, that she wants to do it because it will make him happy, etc. Or, she could make a promise. In the latter case, Anna intends to assure her partner by voluntarily undertaking an obligation, owed to her partner, to do what she promises.

But again: why would another's undertaking of an obligation to do something assure us? One possibility goes as follows: if I believe the promisor to be the kind of person who takes her obligations and commitments to others seriously, I will likely be assured by her promise. More specifically, promises can give promisees assurance if the following conditions hold: (1) the promisor, in virtue of having made a promise, understands herself to have undertaken an obligation (2) the promisee knows (1), and (3) the promisee believes that the promisor is the kind of person who takes her obligations to others seriously (or, at the very least, that the promisor is the kind of person who takes *this kind of* obligation, to *her*, seriously).¹⁶ The promisee's belief, expressed in (3), is naturally expressed in terms of *trust*: the promisee will be assured by a promise if she trusts the promisor to be the kind of person who takes their obligations to others seriously. As I will suggest in the next section, trusting someone typically involves a belief in their tendency or disposition to take their obligations to others seriously.

In this sense, then, we might think of trust as one possible mechanism by which promises can solve the assurance problem. Where promisees trust promisors, they will typically be assured by their promises. However, I want to claim not only that trust is one of the mechanisms by which promises can solve the assurance problem, but that it is the

¹⁵ Rawls only makes some brief remarks about promising being a means of creating obligations that "further one's ends" and that there is some general benefit to having such obligations publicly recognised (1999, p. 305-306).

¹⁶ Or, to be more precise: the promisee believes that the promisor is the kind of person who tends to fulfill the obligations *that she thinks she has* towards others. The promisor might be mistaken about some of her obligations, but if the promisee knows that the promisor thinks she has an obligation to ϕ , and the promisee knows that the promisor takes the obligations she thinks she has seriously, this can be sufficient to assure the promisee that the promisor will ϕ .

‘standard’ or ‘typical’ way in which promises do so. This stronger claim might seem implausible. Trust, after all, doesn’t seem to be the only mechanism by which promises can solve the assurance problem. There are other reasons why the promisee might believe that the promisor will fulfil her obligation, which do not depend on the promisee’s belief in the promisor’s good character. Most obviously, various enforcement mechanisms would seem to do the trick. These mechanisms might come in many shapes. The promisee, or a third party, might have the ability to enforce promises, or extract compensation or impose penalties for non-compliance. Where this is so, enforcement might displace trust as a mechanism for creating assurance. Where enforcement is systematic and efficient, we need not believe the promisor is the kind of person who takes her obligations towards others seriously in order to believe that she will do what she promised. Belief in her self-interest will, in such circumstances, typically suffice.

It is true that promise-breakers are sometimes socially punished. However, we do not have any institutions which dole out such punishments on a systematic basis. And while exclusion from the promising practice might be an option in some circumstances, as Hume envisaged, such exclusion is only possible where information about promise-breakers’ behaviour is readily available to all potential promisees, and potential promisees are willing to act on it. Promisees, in most circumstances, have no way of forcing the promisor to do what they promised they would, or extract compensation if they don’t. Even more importantly, coercing a promisor to perform, or extracting compensation if they don’t, is often inappropriate or impossible (or both), especially in the context of promises made in intimate relations. Promises, then, are not systematically enforceable. This, I take it, is one of the key differences between the social practice of promising and the legal practice of contract: contracts, unlike promises, are systematically enforceable. (Alternatively: if contracts are promises, contracts are the subset of all promises which are legally enforceable.) I mention contracts only to set them to one side, since my focus here is on promises as distinct from contracts.¹⁷

¹⁷ For a more detailed analysis of the distinction between promises and contract, and a discussion of the tendency of enforcement to displace trust in contract, see Dori Kimel’s *From Promise to Contract: Towards a Liberal Theory of Contract* (2003). Since promising and contract are partly governed by different rules (for example, contractual rights, but not promissory rights, are typically transferable), and since they typically solve the assurance problem in different ways, it seems we would be justified in treating them as two separate practices. My argument, however, does not rely on this claim. If one thinks of promising and contract as part of the same practice, one should read my argument as applying only to those promises which are not systematically enforceable. For

While enforcement appears to be the main alternative to trust as a mechanism by which promises can create assurance, perhaps there are other mechanisms. I can think of two in particular, which might appear to threaten my claim that trust is part of the ‘standard’ or ‘normal’ explanation of why promisees are assured by promises. First, the promisee might believe that people have a tendency or a disposition to keep their promises, not because they have a good character or good will, but for some other reason such as habit or instinct. If it were true that people fulfilled their promises out of mere habit, and promisees knew this, this could give them assurance without trust. (Compare: I can be assured that many of my students, out of habit, will choose the same seat in class week after week without trusting them to do so.) But promise-keeping is more than a habitual or instinctive behaviour. We keep our promises not (merely) because we have a habit of doing so (though we might have such a habit as well), but because we think that we should. And we think that we should precisely because we understand ourselves to be obliged or committed to doing so.

Second, in situations where the promisor has significant good will towards the promisee, the promisee might be assured that the promisor will keep their promise without trusting them to do so. (As I explain in the next section, belief in or reliance on another’s good will is typically insufficient for trust.) It must be noted, however, that if I can rely entirely on another person’s good will to do something, extracting a promise from her to do that thing will generally be superfluous. If I trust my partner to be faithful to me, I will not ask him to promise to be faithful to me. I will typically do so only if I have a reason to doubt him. Thus, even in the context of intimate relations, promises are typically made when we are not able to rely *merely* on another’s good will to do the thing promised. This point also nicely illustrates why it can sometimes look as if promises are a substitute for trust. Promises are a tool which create novel obligations for the promisor; hence, they give promisors a new reason to do whatever is promised, which is based not merely on benevolence or good will towards the promisee. Promises in such cases can look like a substitute for trust, or a tool for rebuilding trust where trust has been damaged. But we should describe these cases carefully. Though promises can serve as a substitute for a *general* sense of trust, promises will typically only work to assure promisees if the promisee trusts the promisor to do the *specific* thing promised, not merely out of benevolence or good will, but out of a sense of obligation or commitment to do so.

further discussion of some the differences and divergences between promising and contract, see Jed Lewinsohn (2022), Encarnacion (2018), Shiffrin (2007), and Dori Kimel (2003).

Neither of these two cases, then, appear to threaten my claim that trust is normally an important part of the answer to why promises solve the assurance problem. However, it is of course true that there are lots of cases of promising which do not involve trust in the way I've claimed. What shall we say about such cases? One well-known example, provided by Joseph Raz, imagines a man who solicits a promise, "hoping and believing that it will be broken, in order to prove to a certain lady how unreliable the promisor is" (1977, p. 213). But in Raz's case, in order for the person in question to be able to prove a point about the unreliability of the promisor, promise-keeping must, in the normal or standard case, provide evidence of reliability or trustworthiness. If this weren't the case, the broken promise wouldn't prove what the promisee intends it to prove about the promisor's character. In this way, Raz's example serves to underscore, rather than undermine, the idea that trust plays an important role in the paradigmatic or standard cases of promising.

Another example of a promise which doesn't involve trust is Thomas Scanlon's example of Profligate Pal. In Scanlon's example, Profligate Pal has been borrowing money from his friends for years, always promising solemnly to pay it back but never doing so. Because of this, he has lost his ability to induce others' assurance by making promises. But he is in great need of money. So he comes to you one more time, full of self-reproach and sincere assurances that he has changed his ways. You give him the money he asks for, after he promises to pay it back – not because you trust him, but because you take pity on him. It would be humiliating for Pal if you refused his sincere promise, and you wish to spare him this humiliation. Like in Raz's case, this promise doesn't involve trust. However, it is only against the background knowledge that promises typically involve trust that my acceptance of Pal's promise can save him from humiliation. By accepting his promise, I pretend to trust him by signalling that I expect him to pay me back because he has promised to do so.

Normally, then, promises involve trust. Without being willing to trust the promisor, the promisee will typically not be assured by a promise, and without such assurance, promises could not serve their typical function. If this is right, we can finally see what benefit each and every promisor typically gets simply in virtue of making a promise: the promisee's trust. And since trust is something that promisors can only get if promisees are willing to give it to them, trust is not only a benefit that accrues to promisors, but a contribution to the practice provided by promisees. In the standard case, promisors cannot assure promisees without their cooperation: the promisee must be willing to trust the promisor in order for the promisor to be able to assure them.

Let me sum up so far. I've argued that trust is typically necessary for promises to assure promisees. If that's right, it complicates the picture of the benefits and contributions of the promising practice offered by Hume and Rawls. There are, my picture suggests, at

least three important benefits that accrue to promisors from the practice of promising: the direct but contingent benefit that they might be able to get in return for making a promise, the indirect benefit of living in a society with a stable promising practice, and the benefit of promisee’s trust.¹⁸ But promisors only benefit from the promisees trust where promisees are willing to trust them. So, while both Hume and Rawls appear to take it for granted that the main way in which people can contribute to a promising practice is to keep their promises, I’ve suggested that they both overlook a central way in which promisees contribute to the promising practice: by placing their trust in promisors.

4. Promising and trust

The idea that trust plays a crucial role in promising will not, I hope, come as a surprise. A number of authors have noted the centrality of trust to promising.¹⁹ And trust is, to some extent, present in both Hume’s and Rawls’s accounts. Hume notes that, by breaking a promise, the promisor “subjects himself to the penalty of never being trusted again” (2007, p. 335), and Rawls speaks of promising as something which can give people “trust and confidence in one another” (1999, p. 306).

I hope that the idea that trust might be considered a benefit to promisors, and a contribution of promisees, carries some intuitive force. This idea is compatible with several different accounts of what trust is. Nevertheless, it will help to sketch one plausible account of interpersonal trust, to show in what ways exactly trust might be considered a benefit to promisors, and a contribution to the promissory relation made by promisees.

So far, I’ve glossed trust in terms of a belief that some person is the kind of person who takes their obligations and commitments to others seriously. Other well-known

¹⁸ These are all benefits that accrue to promisors. The benefits of promising, looked at from the point of view of promisees, are different. In addition to the indirect benefits of living in a society with a stable promising practice, promisees get two sorts of benefits from promises: the delivery of the content of the promise (which is typically something that the promisee has an interest in or takes an interest in, in a wide sense), as well as assurance that the promisor will actually deliver on their promise.

¹⁹ Just to give a few examples: Charles Fried describes promises as “the device that gives trust its sharpest, most palpable form” (2015, p. 8). Daniel Friedrich and Nicholas Southwood (2011) advance what they call the “Trust View” of promising, on which the obligation to keep a promise is explained by the fact that the promisor, by inviting the promisee’s trust, incurs an obligation not to betray the trust that they have invited. Trust is also central to Dori Kimel’s understanding of promises. Kimel argues that while trust is absolutely central to promising, it is less central to contract, since the systematic enforcement of contracts tends to displace trust as the primary mechanism of creating assurance (2003).

accounts of trust suggest that trust typically involves belief in, or reliance on, another's good *will*,²⁰ but to me this appears insufficient for trust. We do not trust people who are reliably good willed, but who do not take their commitments to others seriously.²¹ The person who doesn't keep her promises whenever she judges promise-breaking to be in the promisee's interest is not trustworthy. But equally, the person who commits murder to fulfil a promise to meet her friend for lunch isn't trustworthy. Further, competence (both theoretical and practical) plays a role in trust. The good willed but utterly incompetent do not deserve our trust (Jones, 1996, p. 6-7). We do not trust benevolent fools. Trust, then, typically involves a belief in another's good character, understood as a combination of their good will and their motivation to honour their obligations and commitments, and competence.²²

Such a belief, however, doesn't exhaust trust. As Karen Jones (1996) points out, trust often involves not just beliefs about others' motives and future actions, but also certain affective attitudes. Jones herself characterises trust as involving “an attitude of optimism” towards another's good will and competence in a certain domain. We need not agree that trust involves this particular attitude, however, to appreciate the bigger point. Trust involves an affective element, a *way of seeing* the trusted in a favourable light, which typically corresponds to certain “patterns of salience and tendencies of interpretation” (Jones, p. 11). If I trust my partner to be faithful to me, I will not treat with suspicion his agreeing to meet up with an ex, or his staying out unusually late one night. I will not monitor his actions, nor will I seek out signs or evidence of potential infidelity by reading his messages or looking through his pockets. This affective element of trust is important in explaining both why trust can sometimes seem apt even when it ‘leaps ahead of the evidence’ about another's reliability, but also why trust can give rise to beliefs which are unusually resistant to evidence. When we trust, we have a tendency to interpret another's actions and motives in a charitable light.²³

Trust is rarely global, in the sense that it typically covers specific domains of action. We trust people with specific things, projects or tasks: I trust my colleague to show

²⁰ This is true, for example, of Anette Baier's (1986) account of trust.

²¹ For an account of trust as a form of reliance on others to keep their commitments, see Katherine Hawley (2014).

²² This shows why there is no conflict between my account of trust and Hume's idea that promises are typically necessary only when we cannot rely merely on another's good will or benevolence. I can believe someone to have a good *character*, and be motivated to honour their commitments generally, without believing them to have any good will towards me, specifically.

²³ For further discussion, see Jones (1996, p. 11-12).

up on time for our meeting, but not with the details of my love life. Trust tends to cover specific domains of action. Further, when I trust someone with something, this typically implies that I care about the thing I have entrusted them with, or that it carries some importance to me. The specific thing that promisees trust promisors with is given by the content of the promise: the promisee trusts the promisor to *do the thing promised*. This thing, whatever it is, will typically be something of importance to the promisee, or something she cares about.

What is the relationship between trust and reliance? As many writers, including Anette Baier (1986), have pointed out, to trust someone is different from merely relying on them. I can rely on my flatmate to be out of the house every morning by 8 o'clock, since I know she starts work at 9. To rely on her in this sense – to take it as a given, in my practical deliberation, that she will be out of the house by 8 o'clock – is not yet to trust her. However, trust often involves a particular kind of reliance: reliance on another's good will, character and competence. If I trust my flatmate to keep my secret, this typically implies that I can rely on her to keep it, because I believe her to be motivated by her good will and good character. But does trust always involve reliance? Consider a variation of Scanlon's case of the Guilty Secret, offered by Wallace and Kolodny (2003). Harold is a colleague of yours at your home institution. You are going away to visit another institution, where Harold previously worked. A long time ago, when you were younger, Harold did something embarrassing. He asks you not to divulge this secret to his old colleagues, with whom he has little or no continuing contact. Harold cannot rely on your promise in the sense that there's little he can do about it either way: whether you divulge the secret or not, he will not change his plans or actions. Your promise will not affect his practical deliberation in any way. If this is how we understand reliance, it seems clear that there can be trust without reliance. Harold can trust you to keep his secret without relying on you to do so.²⁴

Does trusting the promisor to keep her promise necessarily involve an expectation or belief that she will keep it? Normally, when we trust a promisor to keep her promise, we expect her to keep it. But on the account I've given here, trust involves a belief in another's good will, character and competence, in combination with a tendency of seeing that person's motives and actions in a charitable or favourable light.²⁵ A promisee can trust the

²⁴ This also makes it clear why the idea that promisees contribute to promissory relations by placing their trust in promisors should be distinguished from the idea that promisees incur reliance costs by relying on the promisor.

²⁵ I should stress that this needn't be true of trust in all circumstances. The account I sketch is meant to capture trust as at it features in the context of promising. Trust, as it appears elsewhere, might be somewhat different.

promisor in this sense without believing that the promisor will fulfil her promise, since she might, for example, believe that the promisor will be motivated by her obligation, while foreseeing some difficult obstacle to fulfilment (which the promisor may or may not be aware of). In some cases, nothing less than a belief in fulfilment will satisfy the promisee, and she may reject a promise which doesn't give rise to such belief. However, we are sometimes happy to settle for less: we might accept a promise without expecting fulfilment, assured by the thought that the promisor will, at least, do her very best to fulfil the promise. In this sense, then, we can see how promises normally involve trust, and can do so even in cases where promisees do not expect, nor come to rely on, performance.

With this account of trust in mind, we can articulate in greater detail why trust should be thought of as a benefit for the person who receives it. If it is right to say that trust generally involves a belief in another's good will, character, and competence, as well as seeing the trusted in a favourable or positive light, trust will tend to make one's interactions with other people both easier and more pleasant. Perhaps it is a good thing, in and of itself, to have others regard one in a favourable light – it seems at least plausible to suggest that this is the kind of thing that, other things equal, makes our lives go better. In addition to this, trust can be often used as a tool by the one who has it, to get further advantages or benefits. If you trust me, perhaps you'll let me borrow your car or use your camera.

Though it seems plausible to suggest that others' trust is the kind of thing that (at least sometimes) has intrinsic value, that makes one's life go better, the claim that trust is a benefit of the promising practice is compatible with the claim that trust is not intrinsically valuable.²⁶ Perhaps trust, in the context of an overall exploitative relationship, isn't intrinsically valuable.²⁷ Such trust can nevertheless be of great instrumental value. If you

²⁶ Compare: The money I might receive in exchange for a promise is a benefit to me, even if we grant that money is of no intrinsic value. And money would still count as a benefit to me even if I were to squander it on things of absolutely no value. That we consider money a benefit (in many contexts, anyways) depends on the fact that it tends to confer certain advantages to the person who has it. We would not, however, consider money a benefit if money could not, in general, be used to buy things of value. But this is compatible with the idea that it is of no intrinsic value, as well as the idea that its instrumental value can be squandered, or remain unrealised.

²⁷ I do not deny that trust is something we often do value for its own sake, and especially so in the context of intimate relationships. But perhaps the intrinsic value of trust depends on the value of the relationship in which it occurs. A closely related point is that since trust can be unjustified or inapt (most obviously, when the trusted is not trustworthy), it seems natural to suggest that only justified or apt trust is intrinsically valuable. But even apt trust can be unwelcome, and it is not always clear in which sense unwelcome trust is of value for the person who receives it. It might burden me to have too many people rely on me in too many ways, and especially where this reliance comes along

trust me, it will be all the easier for me to take advantage of you. Indeed, when a promisor uses a promisee's trust to exploit them, our sense that the promisor has benefitted inappropriately from the promisee's trust is especially strong.

Why think of trust as one of the contributions that promisees bring to the promising practice? There are two main reasons, I think, why we should think of trust in this way. First, since trust is often an essential ingredient in the way in which promises create assurance, without trust, the benefits of promising will remain unrealised. Without promisees who are willing to trust promisors, promisors will typically not be able to get anything in exchange for their promises, nor will they receive the benefit of trust itself. If by 'contribution' to a social practice we mean any action without which the benefits of the practice would be lost, or severely diminished, it seems clear that we must consider the promisee's trust as an important contribution to the practice. Second, being willing to trust another involves putting oneself at risk of being let down, and of having one's trust broken. To trust someone is to make oneself vulnerable vis à vis the person one trusts. In this sense, too, it makes sense to think of trust as a contribution, or a cost, involved in sustaining the practice of promising. When one trusts, one accepts vulnerability to another's possible (but not expected) ill will and incompetence.²⁸ Since promisees can release promisors from their obligation at any point, we might think of the trust that promisees place in promisors as at least partially willing (more on this point below). But the fact that trust is, at least in a sense, willingly given – and the fact that promisees often *want* to be assured – does not detract from the fact that their trust is an important, and typically essential, contribution to the practice.

It might seem as if my account comes up against the following objection. To trust, as many writers on trust have pointed out, is typically not something that we can do at will (though trust can, over time, be cultivated). Yet, I've spoken consistently about promisees being *willing* to trust promisors. How should we make sense of this? First, as I've already noted, trust tends not to be global. We trust people in certain domains of actions, or with certain specific things. Promises are a tool which can expand the domain of actions in which the promisee trusts the promisor. Thus, before my colleague promised me to comment on the latest draft of my article, I might have trusted her to do many other things, but not *to read the latest draft of my article*. Only after she promises does she acquire an obligation to do so. So if I trust my colleague to be the kind of person who takes her commitments to others seriously, I will trust her to comment on my draft only after she

with an expectation that I will take others' reliance into account when I act. Where this is the case, the possible ways in which I can act without disappointing others are limited.

²⁸ The formulation is inspired by Anette Baier (1986, p. 235)

promises to do so. In this sense, we can see how promises are a mechanism by which trust can be expanded. This also helps to make sense of the idea that the promisee’s trust is a benefit of each new promise: each new promise typically expands the domain of trust.

Further, while philosophers disagree about whether or not promises need to be accepted in order to be valid, it is not controversial to suggest that promisees have the power to release promisors from their promissory obligations (including at the time the promise is made).²⁹ Because of this, a promisee who is aware that a promise has been made to her, and who doesn’t release the promisor from their promissory obligation, at least implicitly signals her willingness to trust the promisor to do the thing promised. In this sense, trust is at least partially willing on the part of the promisee. Promisees can, by choosing whether or not to release a promise, decide whether or not to trust the promisor to do the specific thing promised. This is compatible with the thought that, once both promisor and promisee understand a promissory obligation to have been created, the promisee is not entirely free to choose whether or not to trust the promisor to keep it.

5. Whose cooperative effort?

It’s time to put everything together. If we try to explain the wrong of breaking a promise via the principle of fairness, we need to understand not only in what way promisors benefit from promising, but also *whose* cooperative efforts make the promisor’s benefit possible. That is the question I turn to now.

It seems plausible to suggest that promisors owe their obligation to keep their promise to those whose cooperative efforts they benefit from, when they make a promise.³⁰ On conventional theories of promising, promising tends to be understood as a cooperative effort on the level of ‘society at large’. On this picture, it might naturally seem as if it is the cooperative efforts of all promise-keeping strangers that we benefit from when we make a promise. In this section, I’ll suggest that this picture is largely mistaken.

On one common way of telling the conventionalist story about promising, the promisor would not be able to assure the promisee merely by uttering the words “I promise,” unless there existed in society a widespread social practice that consisted of

²⁹ Just to give a few examples: Gary Watson (2009) and David Owens (2012) both hold that promises must be accepted to be valid, whereas Raz suggests that this isn’t the case: “I may promise someone in person, face to face, to come and visit him in hospital tomorrow. He receives the promise stone-faced, saying nothing and moving not a muscle. My promise is binding, and it requires some imagination to claim that it has been accepted” (Raz 2022, p. 196).

³⁰ I’ve suggested that this is also the most natural interpretation of Rawls’s account of the directedness of promissory obligation.

people actually keeping their promises. On this narrative, the promisee is assured by a promise on the basis of evidence about other people’s promise-keeping behaviour. Only if people in general tend to keep their promises would the promisee be assured by a promise. Were this narrative true, promisors could indeed be said to benefit from the cooperative efforts of everyone in society who keeps their promises.

But now consider a promise made between close friends. Say my best friend promises to come to my birthday party, and I am assured by her promise. Why does my best friend’s promise give me assurance? It seems highly unlikely that the promise-keeping behaviour of strangers will feature in any way in this explanation. Instead, my willingness to trust my best friend will be based on facts about *her*: that I know her to be trustworthy, and to take her commitments to others seriously. My assurance, in this case, has absolutely nothing to do with the fact that people in general, now and around here, tend to keep their promises. And notice, too, that the evidence I have for my friend’s trustworthiness might be unrelated to her previous promises. I can get good evidence of trustworthiness in many ways that don’t involve promising. So, when my best friend promises to come to my birthday party, and I am willing to trust her, it would seem odd to claim that she benefits in any way from promise-keeping strangers in securing my trust. Instead, I want to claim, the only person whose cooperative effort she directly benefits from is my cooperative effort: I am the one willing to trust her.³¹ On the assumption that promises are owed to the person(s) whose cooperative effort makes promising possible, my friend owes it to me, and only to me, to keep her promise. If she breaks her promise, she exploits *my* trust, while she has not benefitted in any significant way from anyone else’s cooperative efforts. She has wronged me, and no one else.

Promises between complete strangers, and promises between the most intimate of friends, are in this respect at opposite ends of one kind of spectrum. When a stranger makes a promise to us, we tend to have no evidence about their trustworthiness other than the kind of evidence we might have about people in general. But when it comes to explaining why we are willing to trust our partners or closest friends when they make promises to us, such indirect evidence is superseded by the direct evidence we have about their past behaviour and character. Further, just as we can get a sense of another person’s trustworthiness in many ways, and not just by studying her promise-keeping behaviour, so we can get evidence of the general trustworthiness of strangers without relying on any evidence which relates to promise-keeping, specifically.

³¹ Of course, my friend’s own past behaviour will play a role in whether or not I am willing to trust her. So, if my friend has kept her promises to me in the past, we might say that she reaps the benefits of her own past good behaviour as well.

What are the implications of this for the directedness of promissory obligation? The picture I have been working up towards is, as it turns out, rather nuanced. I've suggested that promisors who rely on promisee's trust in making promises to complete strangers or distant acquaintances, *may* benefit from the cooperative efforts of promise-keeping people in general, if the reason strangers are willing to trust them is that they have evidence that people, in general, tend to keep their promises. But since I've also suggested that our sense of whether strangers are trustworthy or not is influenced by far more than our sense of other people's promise-keeping behaviour, it is not obvious that those who manage to induce the trust of strangers are always benefitting from others' promise-keeping behaviour, specifically. When it comes to promises made in the context of intimate relationships, the story is rather different. The more information we have about a given promisor, and the better we know them, the less likely we are to rely on any kind of indirect evidence about people in general when deciding whether or not to trust promisors. The perhaps somewhat surprising implication of all this is that, while I might owe my promise to my mother not to be late to dinner to her and her alone, I might owe a promise to a stranger, if I successfully induce their trust, both to the stranger and to everyone who has contributed to create a general climate of trust.

But here's a further reason to think why we might not owe such a promise to all promise-keeping strangers, after all. When it comes to social practices in general, we might want to distinguish between the general social conditions that enable them, and the acts which count as direct contributions to them. Earlier, I mentioned a tax-financed system of health care. I said that the costs or contributions involved in such a system are what each person pays towards the scheme, as well as the labour and time involved in organising it. But there are many other things which are required for a healthcare system to be sustained that are simply part of what we tend to think of as the background social conditions: there must be some general infrastructure, without which neither the medical practitioners nor the patients could get to the hospital, the doctors and patients must share a language in order to be able to communicate with each other, and so on. However, we do not tend to think of these goods as direct contributions to a tax-financed system of healthcare, even though without them, such a system would not function properly. Something similar might be said about our sense of the general trustworthiness of strangers. This might seem more like part of the general background social conditions that make trust, and hence promising,

easier, as opposed to a direct contribution to the promising practice, to be credited to the combined efforts of all promise-keeping strangers.³²

Finally, it is time to deal with an important objection. Sometimes promises are made which do not manage to secure the promisee's trust. And such promises too, it will be said, are binding. How does my account deal with such cases? Consider again Scanlon's case of Profligate Pal. It might seem as if the account I have just sketched must conclude that Profligate Pal, because he isn't able to induce my trust, does not benefit from the practice of promising, and hence has no obligation to pay me back. This is the conclusion that Scanlon reaches (for different reasons); but even Scanlon acknowledges that the result might seem somewhat counterintuitive.

But the case of Profligate Pal does not present a difficulty for the picture I have drawn. While we should not recognise, as the *only* benefit of promising, whatever promisors can get in exchange for their promises, we should not deny that very often promisors *do* benefit in contingent ways from making promises. One of the points I have been trying to stress is that the benefits and the contributions of our practice of promising are much more varied and complex than we might first think. Where the promisor benefits in some way from making a promise, this benefit can explain – via the principle of fairness – their obligation to keep their promise. In the case of Profligate Pal as Scanlon describes it, Pal does benefit from his promise, albeit not via the usual mechanism. He gets the money he asks for, not because you trust him, but because you don't want to humiliate him by rejecting his promise. Had he not offered his sincere promise to pay you back, it wouldn't have been so humiliating for him if his request had been turned down, and hence, Pal might not have gotten the money. So Pal does benefit from invoking the practice of promising.

What about those cases in which some person doesn't benefit *in any way* from making a promise? Imagine the case of Unfortunate Una. She tries to make a gratuitous promise to me, without getting anything in return. But for some reason, she is unable to get my trust. Is Una's promise to me still binding, and owed to me? Without specifying any further details, there are a number of different ways in which we might respond to this case. First, we might note that, if I tell Una that I don't trust her to keep her promise, this will (at least in certain contexts) be sufficient to count as a rejection of her promise. Rejected promises, I take it, do not create any promissory obligations.³³ Further, the reason why I

³² Such general social trust will also make it easier for people to form close interpersonal trusting relationships. If we start from a default position of distrust, the hurdles to developing close, trusting friendships are much higher than if we start from a default position of trust.

³³ As already noted, there is widespread agreement that promisees can release promisors from their promissory obligations. (This is also expressed in the Obligation Conception of promising, which I

am unwilling to trust Una seems to matter. Is Una, like Pal, a notorious promise-breaker? Has she continuously exploited my trust in the past? If so, we might explain her promissory obligation, and the fact that it is owed to me, on the basis of this past benefit. But if Una has never before broken a promise (as far as I know), we might instead want to start asking some questions about *me*. Why am I unwilling to trust her? If I, out of ignorance or prejudice, think of people like Una as untrustworthy for no good reason, my unwillingness to trust her begins to look like an unjust way of excluding Una from a generally useful social practice. Those who are excluded from a social practice in this sense, do not, in general, have to obey its rules.³⁴ Finally, consider the case where I, for some reason unrelated to wanting assurance or trusting Una, nevertheless accept (or fail to reject) her promise.³⁵ In this case, I am happy to grant that Una, if she breaks her promise, might harm me in various ways – she might have promised to do something of great importance to me, and I might suffer some harm when she fails to deliver this – but her promise-breaking, it

have been taking for granted from the start). If promisees can release promisors from their promissory obligations, they can presumably do so from the very moment a promise is first made. So we might think of a ‘rejected’ promise as a promissory obligation that the promisor is immediately released from.

³⁴ Some people might will resist the claim that it’s possible to be excluded from promising in this way, since it seems to imply that one can lose one’s normative power to promise. But it seems plausible to me that one could lose this power: just like it takes two to tango, it takes (at least) two to promise.

³⁵ There are complicated issues about when, exactly, a communicated promise succeeds in creating a promissory obligation. The important point to note here, however, is that there are promises which appear perfectly valid and binding, even though promisees don’t trust promisors to fulfil them.

seems to me, would not necessarily *wrong me*.³⁶ Most obviously, this is because in breaking her promise, Una has not violated my trust.³⁷

6. Conventionalism, again

It might seem as if conventionalism drops out of the picture altogether on my view of promissory obligation. My picture is certainly compatible with conventionalism, but it might seem that it is rather trivially so. This conclusion, however, would be mistaken. While I have not focused on defending the core conventionalist claim – that promising is, in some important sense, impossible or unintelligible outside of a social practice of promising – I nevertheless think that a convention or practice of promising is necessary to make sense of the account I have offered. As I noted in Section 3, promises can give promisees assurance if the promisor, in virtue of having made a promise, understands herself to have undertaken an obligation, the promisee knows this, and the promisee believes that the promisor is the kind of person who tends to fulfil her obligations to others. As I see it, the convention or social practice of promising is necessary to explain the first step: why promisors think of themselves as having undertaken a (genuine) obligation by making a promise.

To see why, it will help to compare the role that convention plays in my account and the role that it plays on other conventionalist (or ‘hybrid’) accounts. Several writers on promising suggest that a convention of promising is necessary to explain why promisees

³⁶ Although there is a complication here about the ‘baseline’ of harm, and whether or not we think of harm in counterfactual terms. If you are simply deprived of a potential benefit, have you been harmed? If Una does nothing wrong in breaking her promise, it might seem more accurate to say that you’ve missed out on a potential benefit, than to say that you’ve been harmed. If that’s so, we can only make sense of the idea of someone being harmed by a broken promise if we presuppose that people should keep their promises. (But note that we might be able to fall back on a more general ‘two-level’ explanation of promises to explain why people should keep their promises. Perhaps we have a (non-directed) obligation to follow the rules of promising, if we choose to invoke them, whether or not we benefit in each and every instance when we invoke the rules. If so, Una does something wrong if she breaks her promise, and this wrong might harm you even though it does not *wrong you*.)

³⁷ The idea that a promise-breaker wrongs the promisee by inappropriately benefitting from her trust and then breaching it is naturally expressed in terms of the promisor ‘violating the promisee’s trust.’ However, my account does not depend on this being a distinct type of wrong or harm. Broken promises wrong promisees if and when promisors benefit inappropriately from benefits provided by promisees, independently of how, exactly, this might be said to have harmed the promisee (e.g. by impairing her ability to trust again).

are assured by a given promise, or why they expect performance. The promisee would only be assured or expect performance if people, in general, kept their promises.³⁸ On the account I have suggested, the convention of promising is not necessary in this sense. General promise-keeping behaviour *may* contribute to why promisees are assured by promises. But regular compliance within a given social group is not strictly necessary to create the relevant assurance. Promisees will typically be assured by promises if and when they deem promisors to be trustworthy, and promisees can gather evidence about this in ways that don't involve promising at all. Promissory obligations, after all, are only one of many different kinds of obligations and commitments. Because of this, my account suggests that we can be assured by promises, especially those given by people we know well, even if people in general break their promises most of the time.

What need the practice or convention of promising to explain is why people take themselves to be able to create genuine obligations by uttering certain words or performing certain actions. And this is not just a matter of the meaning of words, or the intended effects of our speech acts. We could all agree that when I say “wingardium leviosa” and point my wand to an object, I've uttered a spell which ‘means’ that the object I point to will levitate, and that I've done so with the intention of making the object levitate.³⁹ Nevertheless, my words will not cause any objects to levitate, no matter what the words mean and how sincere my intention. I might want to have the power to make objects levitate, and perhaps it would be good for me if I did, but to think that they would levitate merely because it would be good or valuable if they did would be wishful thinking.⁴⁰ With promising,

³⁸ This is how the convention of promising features in e.g. the ‘hybrid’ account of promising offered by Wallace and Kolodny (2003). Wallace and Kolodny use the social convention of promising to explain why promisees come to expect that a given promise will be kept. They appear to assume that such an expectation could only be generated if people, in general, adhered to the rules of promising.

³⁹ Though if I don't believe that I have the power to make objects levitate by mere say-so, we may question not only the sincerity of my intention, but also my ability to form it.

⁴⁰ One might think the example is inapt. It might seem obvious that physical facts cannot be willed, but that the case is not so clear in the case of all moral facts. Of course, the case is clear enough with regards to *some* moral facts: I cannot will it to be the case that cheating on my partner is a decent thing to do, no matter how hard I try. Whether or not it is the case that some moral facts can be willed, I do not think that the voluntarily incurred obligations owed to others is the kind of moral fact that can simply be willed into existence (even if we successfully manage to communicate such a will. Consider again the example of the power to create ‘commodity promises,’ from Section 1. I imagined a society which doesn't have either a legal or a social practice which establishes the power to create promissory rights and obligations which can be bought and sold. It doesn't seem

however, things are different. To think that the words “I promise” can create obligations is not mere wishful thinking. As I understand it, this is because once we start acting *as if* promises create obligations, they do. We can begin acting as if promises create obligations (e.g. by holding each other accountable for keeping promises) whether or not they do. And if we start acting in this way, we will soon have created a useful social practice. Once such a social practice exists, it will be true that, when we make promises, we have an obligation to keep them. On the account I have given, this can often be explained by the fact that promisors, by making promises, induce the promisee’s trust, and promisors who break promises inappropriately benefit from such trust.

This brings us back to my initial dissatisfaction with Raz’s ‘normative powers’ account of promising. Though many writers contrast Raz’s normative powers account of promising with conventionalist accounts of promising, I find such a juxtaposition unhelpful. Raz’s ‘normative powers’ account of promising, as he himself makes clear (at least in his later work), is compatible with many forms of conventionalism. As Raz puts it: “Can Jane promise Rose to do her shopping next week unless Rose too understands what promises are and that they bind (or at least that that is a common view)? Such thoughts may encourage the view that normative powers exist only when there is, in the relevant population, a common understanding that they do. True or false, variants of this view are consistent with the definition here offered” (2022, p. 169).⁴¹ Of course, I’ve offered a rather different picture of the value of having the power to make promises than that offered by Raz, stressing some rather concrete benefits that accrue to promisors in virtue of having this ability. I’ve also explained the wrong of breaking a promise in a way that is different from Raz’s. Nevertheless, on my picture, too, the existence of our normative power to promise ultimately depends on the value of us having this power. For those who are more convinced by Raz’s account of the value of having the ability to promise, as well as his account of why its wrong to break a promise, it is nevertheless possible to combine this

unreasonable to think that it might be valuable for people in this society to have the power to make commodity promises. Nevertheless, this doesn’t seem sufficient to establish that they do. And besides, what would we *mean* if we simply insisted that they did have such a power? No one would understand me if I tried to make a ‘commodity promise.’ And even if I patiently explained to the promisee what I meant by this, my intention to undertake such a promise would still, it seems to me, be unsuccessful.

⁴¹ Many conventionalists would think that a common understanding of the rules of the practice isn’t sufficient to establish that the practice exists. People must also (at least sometimes) act in accordance with this common understanding in order to realise the potential benefits of the practice, which ultimately ground any obligations that the practice gives rise to.

part of Raz’s account with my picture of the role that convention plays in explaining why we can will obligations into existence in the specific way that promising allows, but not in any of the other ways that we might imagine.

I will end by noting what, if anything, is left in my account of the idea that the wrong of promise-breaking is the wrong of undermining or free-riding on a useful social practice. My account suggests that the idea that promise-breakers could do much to undermine our practice of promising is severely misleading. It is misleading, first, because our willingness to trust our close friends and family to keep promises is at best tenuously related to whether people, in general, tend to keep their promises. Promising could probably survive as a social practice between some people, even if most people, most of the time, broke their promises.⁴² Second, it is misleading because promise-breakers can be excluded from the promising practice, at least by those who get wind of their promise-breaking behaviour.⁴³ It seems more likely, then, that a perpetual promise-breaker would be excluded from the practice of promising than succeed in undermining it. To some extent, however, I have sought to vindicate the idea that promise-breakers are free-riders on a useful social institution. I have tried to show that while it is possible to think of promising as a cooperative practice on a social scale, we can also think of each and every promise as a cooperative effort between promisor and promisee, in which the promisor typically induces the promisee’s trust. We might say, then, that promise-breakers free-ride on the social practice of promising primarily in virtue of the fact that they free-ride on the promisee’s trust.

⁴² In other words, I don’t think general compliance is necessary for a social practice of promising to exist. What does seem necessary, however, is that the rules of promising be generally known, that people think of those rules as rules which *ought* to be followed, and that at least some people, some of the time, follow those rules.

⁴³ Out of the three benefits of the promising practice that I have stressed here, only one benefit – Hume’s idea of the indirect benefit of ‘commerce’ – is nonexcludable, in the sense that it is typically not feasible to exclude any given person from enjoying it. But given that, in our society, the legal practice of contract has to a great extent, replaced promising as the assurance-creating mechanism of choice in commercial transactions, it might seem more appropriate to thank all the contributors to our legal contracting practice, rather than the contributors to our promising practice, for much of the benefits of ‘commerce.’ The two remaining benefits of promising – trust, and whatever you can get in exchange for your promise – are (to some extent) excludable, as Hume himself noted.

References

- Arneson, Richard J. (1982). The principle of fairness and free-rider problems. *Ethics* 92 (4):616-633.
- Baier, Annette (1986). Trust and antitrust. *Ethics* 96 (2):231-260.
- Darwall, Stephen (2013). *Honor, History, and Relationship: Essays in Second-Personal Ethics II*. Oxford, GB: Oxford University Press.
- Kimel, Dori (2003). *From Promise to Contract*. Bloomsbury Publishing.
- Encarnacion, Erik (2018). Contract as Commodified Promise, 71 *Vanderbilt Law Review* 61.
- Fried, Charles (2015). *Contract as Promise: A Theory of Contractual Obligation*. Oxford University Press USA.
- Friedrich, Daniel & Southwood, Nicholas (2011). Promises and Trust. In Hanoch Sheinman (ed.), *Promises and Agreements: Philosophical Essays*. Oxford University Press.
- Gilbert, Margaret (2004). Scanlon on Promissory Obligation: The Problem of Promisees' Rights. *Journal of Philosophy* 101 (2):83-109.
- Hart, H. L. A. (1955). Are There any Natural Rights? *Philosophical Review* 64 (2):175-191.
- Hawley, Katherine (2014). Trust, Distrust and Commitment. *Noûs* 48 (1):1-20.
- Hume, David (2007). *A Treatise of Human Nature: Volume 1: Texts*. Norton, David Fate & Norton, Mary J. (eds.) Clarendon Press.
- _____ (1987). *An enquiry concerning the principles of morals*. Schneewind, J.B. (ed.) Indianapolis: Hackett Pub. Co.
- Jones, Karen (1996). Trust as an affective attitude. *Ethics* 107 (1):4-25.
- Kolodny, Niko & Wallace, R. Jay (2003). Promises and Practices Revisited. *Philosophy and Public Affairs* 31 (2):119-154.
- Lewinsohn, Jed (2022). By Convention Alone: Assignable Rights, Dischargeable Debts, and the Distinctiveness of the Commercial Sphere. *Ethics* 133 (2):231-270.
- Melenovsky, C. M. (2017). Promises, Practices, and Reciprocity. *Philosophical Quarterly* 67 (266):106-126.
- Murphy, Liam (2021). The Artificial Morality of Private Law: The Persistence of an Illusion. *University of Toronto Law Journal*, Volume 70 Issue 4, pp. 453-488. Available at SSRN: <https://ssrn.com/abstract=3784977>.
- Nozick, Robert (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Owens, David (2012). *Shaping the Normative Landscape*. Oxford, GB: Oxford University Press.

- Rawls, John (1955). Two Concepts of Rules. *Philosophical Review* 64 (1):3-32.
- _____ (1999). *A Theory of Justice: Revised Edition*. Harvard University Press.
- Raz, Joseph (1977). Promises and Obligations. In *Law, Morality, and Society: Essays in Honour of H. L. A. Hart*, ed. P. M. S. Hacker and Joseph Raz. Oxford: Oxford University Press.
- Raz, Joseph & Heuer, Ulrike (eds.) (2022). *The Roots of Normativity*. Oxford: Oxford University Press.
- Scanlon, Thomas (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- Shiffrin, Seana Valentine (2007). The Divergence of Contract and Promise. *Harvard Law Review* 120:708-53.
- Simmons, John (1979). The principle of fair play. *Philosophy and Public Affairs* 8 (4):307-337.
- Thompson, Michael (2004). What Is it to Wrong Someone? A Puzzle about Justice. In R. Jay Wallace, Philip Pettit, Samuel Scheffler & Michael Smith (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Clarendon Press.
- Watson, Gary (2009). Promises, reasons, and normative powers. In David Sobel & Steven Wall (eds.), *Reasons for Action*. Cambridge University Press.