

How to Believe a Promise: A Defense of Scanlon

Nat Tabris

Abstract

According to Scanlon, in order for me to promise to do X, I must lead you to believe that I will do X. Further, promising is possible in cases where my only reasons for doing X is my promissory obligation. This leads to a circularity problem, as noted by Kolodny and Wallace: you should believe that I will do X because I promised to do X, but the promise depends on your belief that I will do X. I argue that this circularity is not problematic, because your belief that I will do X is self-fulfilling. I then defend the possibility of forming self-fulfilling beliefs at will. Finally, I use my account to explore the background conditions necessary for promising.

This paper is about promising, but I'm more interested in the notion of *believing at will* that I use to defend Scanlon's account of promising. So while I think Scanlon's account of promising is basically right, I'm only going to say enough about it as I need for setting up the problem which I'll spend the rest of my time trying to solve.

Scanlon's account is supposed to explain how you can promise to do X even when you have no prior reason to do X, and even if you're in the 'state of nature' without social obligations or conventions. To this end, he introduces a "principle of fidelity" such that if I intentionally lead you to believe that I will do X because I know that you want this assurance, then I must do X unless you release me.¹

Here's the problem (this was raised by Niko Kolodny and R. Jay Wallace).² If I don't have a prior reason for doing X, then how can I lead you to believe that I will do X? There's a circular dependence: promising is supposed to supply the reason, but the promise doesn't exist until you believe that I will do X. Maybe you believe that I'll do X because I tell you I intend to do X, and you know that I wouldn't lie (Scanlon suggests something like this). But a mere intention to do X doesn't give you reason to believe that I will do X unless you have reason to believe that my intention is more than a whim—you need to believe that I have a solid reason for doing X, but we want promising to be possible where there is no antecedent reason.

If accepting a promise were an action that could be performed at will, one could get around the circularity worry thus: you know that if you accept the promise, then I will do X, and you (at will) accept the promise. If promising involves a social convention, then it's plausible that accepting a promise is an action which can be performed at will. But on Scanlon's account, you accept a promise by becoming assured that I will do what I promise—you believe that I will do X—and forming a belief isn't (usually) an

¹Scanlon (1998), p. 304

²Kolodny and Wallace (2003), pp. 119–154

action that can be performed at will. Thus, if we want to preserve the independence of promising from social practice, we need a different solution.

One cautionary note: you may think that what I call ‘promising’ is not in fact promising (because, e.g., you believe that a promise has force even when not accepted). If that’s what you think, then you’re welcome to suppose that I use ‘promise’ to refer to some (other, non-promissory) sort of obligation created by an act of will in a way which does not essentially depend on a convention for its obligatory force. It is the possibility of creating obligations in this way that I shall defend.

1 Believing at Will

To find the solution, we need to take a digression and consider the possibility of rationally forming certain beliefs ‘at will’.

1.1 The argument for self-fulfilling beliefs

Normally, if you ask me whether I believe that p , I can answer this question by considering my reasons for thinking that p is the case. As Hieronymi puts it, “there are no possible cases in which you answer positively the question whether p but are prevented from believing p . In answering the question positively, one has already, therein believed.”³ If rationally forming the belief that p just is believing that p is the case, then we cannot believe p if we know that p does not obtain (or will obtain, if p concerns the future). For promising, the promisee would not be able to believe that the promisor has strong reason to do X when the promisee knows that the promisor does not, presently, have such reason.

Here’s another way of looking at the question of whether I believe that p . Suppose

³Hieronymi (2006), pp. 54–5

I have a friend Norman who's going to take a test tomorrow. I know that Norman is a pretty good student, and as things now stand, the facts I know about Norman give me reason to believe that the likelihood that Norman will pass the test is 90%. But I know additional facts which complicate the situation. First, Norman has the uncanny ability to know my occurrent beliefs about him just by looking at my face. Second, when Norman is nervous, he does poorly on tests—in fact, I know that if Norman were nervous, then the likelihood that he would pass the test tomorrow is only 20%. Third, Norman will be nervous if he believes that I expect him to pass.

Norman asks me if I believe that he will pass the test tomorrow. How should I answer? At present, my evidence does tell me that he will pass. But if I believe that Norman will pass, I also know that this belief will causally bring it about that Norman will be much less likely to pass. So I know that if I do believe that Norman will pass, I will no longer be rationally justified in believing that Norman will pass. I think it is clear in this case that I cannot rationally form the belief that Norman will pass when I know that this belief will be self-undermining.

Norman's case shows that I must not consider only the initial probability that p when determining whether or not to believe that p . You might think I should instead do this: First, I consider the current likelihood that p . Then, I consider this evidence about p as *causing* me to have a belief (either that p or that $\sim p$). Finally, I consider what the likelihood that p will be the case once this causal chain has been played out, and update my belief that p based on this final likelihood.

This won't work. This way of calculating would tell me that there is a 20% likelihood that Norman will pass the test, and I would form the belief that Norman will not pass the test based on that likelihood. But this belief would be false—since if I do not believe that Norman will pass, then the likelihood that he will pass is in fact 90%.

Instead, I ought to consider two questions framed with counterfactual conditionals:

(i) If I were to believe that p , would it be the case that p ?

(ii) If I were to believe that $\sim p$, would it be the case that $\sim p$?

When considering these questions with respect to Norman passing the test, I see that the answer to both questions is 'no'. If I were to believe that Norman will pass the test, then it would be the case that he would probably not pass the test. If I were to believe that Norman will not pass the test, then it would be the case that he would probably pass the test. Thus, it is not clear in this case that I can rationally believe either p or $\sim p$.

Consider another case. Sam, like Norman, has knowledge of my occurrent beliefs about himself. And like Norman, Sam is also going to take a test tomorrow, and wants to know if I think he will pass. But Sam is not a very good student; the likelihood that he will pass is only 40%. If he knows that I believe that he will not pass, he will become discouraged and do worse: the likelihood that he will pass if I believe that he will not pass is only 20%. In this case, I rationally ought to believe that Sam will not pass—this belief would be self-reinforcing, rather than self-undermining. But a further question arises: What credence ought I to give my belief that Sam will not pass? Suppose that I believe that I should assign a credence of 60%. As soon as I do this, it will be incorrect: the belief that I ought to assign a credence of 60% is self-undermining, because as soon as I do assign a credence of 60%, the credence I ought to assign is 80% (and I know this). This suggests that my credence that p should be based on the counterfactual likelihood that p given my belief that p , not on the initial likelihood that p .

Now I want to look at one final case. My friend Owen is also going to take a test tomorrow. Given the way things stand now, my initial credence that Owen will pass the test should be 50%. If Owen were to know that I believe that he will pass the test, then this would give Owen the confidence he needs in order to do better on the test.

If Owen were to know that I believe that he will not pass the test, then Owen would become nervous and be less likely to pass the test. Owen, who like all my other friends can see what I'm thinking about him, asks me if I think he will pass the test. What am I to think?

The case of Norman shows that I must consider the credence I *would* assign to a possible belief that p when determining if I am to believe that p . And the case of Sam is meant to suggest that when I am determining my credence for believing that p , I ought to consider the credence I will assign and not the initial credence. My hope is that it is now plausible to believe that when I consider Owen, I should think about the likelihood that Owen will pass given my belief that he will pass, and the likelihood that Owen will not pass given my belief that he will not pass. In considering these two likelihoods, and not considering the initial likelihood that Owen will pass, it seems to me rational to believe that Owen will pass—since in this case it is likely that Owen will pass. And it seems rational to believe that Owen will not pass—since in this case it is likely that Owen will not pass. I am, so to speak, epistemically free to believe either way.⁴ Thus I believe that in determining what to believe about Owen, it would be rational for me to form the belief that Owen will pass.

1.2 The role of the will

Something still feels funny about forming a belief which is justified only if I believe it. There are a couple different ways to bring out what might be wrong in forming such a self-fulfilling belief. Following Moran, one might argue that it is not rational to regard one's belief from the third-person perspective, as if it were a fact in the world.⁵

⁴I draw this notion of epistemic freedom from Velleman (1989).

⁵Moran (2001): "The clash between these two perspectives on oneself is more clearly exemplified in such phenomena as *akrasia*, self-deception, and other conditions where there is a split between an attitude I have reason to attribute to myself, and what attitude my reflection on my situation brings me

One could then argue that the justification required for forming a self-fulfilling belief requires considering the belief from the third-person perspective.

Another way of looking at the problem: It seems that in order to form the belief that Owen will pass, I need some sort of reason to tip the balance. Otherwise, why would I believe that Owen will pass rather than believing that Owen will not pass? Since the reasons that bear on the truth of whether or not Owen will pass are balanced on each side, any reason that tips the balance must *not* be a reason which bears on the truth of the belief. Hieronymi, attempting to explain why we cannot believe at will, argues that it is “conceptually impossible to exercise evaluative control over a belief for extrinsic reasons” because it is only by consider reasons which bear on the “question about whether *p*” that we can immediately control our belief concerning *p*.⁶ If she is correct, then it is impossible to directly form a belief based on reasons which do not bear on the truth of *p*, but it seems we must do exactly this in order to rationally form a self-fulfilling belief.

In response to the first worry, it seems to me that when there is a *causal* connection between a belief and the state of affair which that belief is about, it would be irrational *not* to consider this potential belief from the third-person perspective. In most cases, our beliefs have at most an *evidential* bearing on their truth.⁷ Supposing that I’m a reliable judge of matters concerning *p*, the fact that I believe that *p* may be evidence for *p*; nonetheless, it would be irrational for me to use my belief that *p* as evidence to justify my belief that *p*. When evaluating the reasons for believing that Norman will pass the test, I consider my potential belief not as evidence, but as a cause. This seems to me the crucial difference which explains why considering my self-fulfilling belief

to endorse or identify with,” p. 67

⁶Hieronymi (2006), p. 58

⁷As in Moran’s discussion of Satre’s gambler, who “now realizes that the empirical perspective cannot provide any additional strength of its own, for all of it is borrowed from the strength of the resolution itself,” p. 80.

from the third-person perspective is rational.

What about the second worry? Let me try to explain how it might be possible that reasons for wanting p to be the case bear on forming a self-fulfilling belief that p . The desire that p be the case cannot directly bear on the question of whether or not p is the case, but it can bear on the question of whether or not to do some action which makes it the case that p . For instance, if I want to go to a party, this is a reason for intending to go to a party. Now, if you ask me why I believe I will go to a party, I can cite my desire to go to a party as a reason for *believing* that I will go to a party.

What about the case of Owen? If there were some action that I could intend to do on account of a desire that p be the case, such that this action would bring it about that p is more likely to be the case, then this would explain how the desire that p bears upon the question of whether p will be the case. But the event which makes it more likely to be the case that p is my believing that p , and believing isn't supposed to be an action one can intend and perform at will—beliefs can be formed only if we have reasons which bear on the question of whether or not p is the case.

Clearly there's a sense in which one cannot rationally form a belief at will. Beliefs are beholden to reasons which bear on the truth of the belief. Nevertheless, I think there is a sense in which we do have some control over the formation of beliefs. It seems possible for me to withhold forming a belief, even if I know that I would form a certain belief based on the reasons I have which bear on the matter—for example, I might withhold forming a belief until I have had more time to investigate the matter. Also, it is possible to reconsider a belief by re-evaluating the reasons for which a belief is held. Thus belief formation can be intentional insofar as we can decide whether or not to engage in evaluating certain reasons, although once we do decide to engage in evaluation, we cannot control the outcome.

Further, it seems as if I can choose the order in which I evaluate reasons and update

my beliefs. Suppose I know a set of arguments for p and a set of arguments for $\sim p$. I may choose to evaluate the set of arguments for p before evaluating the set of arguments for $\sim p$, and if the arguments for p are sufficiently strong, it would not be irrational for me to update my beliefs before considering the arguments for $\sim p$.⁸ Thus by deciding to begin deliberation by evaluating a certain set of reasons, and by deciding to update my beliefs when warranted by the reasons I have considered, rather than waiting to evaluate additional reasons, I can control the order in which I update beliefs.

Normally the order of evaluation does not affect the final outcome. If I know that the pen I'm looking for is either in the right drawer or in the left drawer, I cannot make myself rationally believe that it is in the right drawer based on the order in which I consider the drawers, since the likelihood that it is in either drawer (in particular) is only 50%. But for a self-fulfilling belief, when I consider the likelihood of it being the case that p were I to believe that p , the likelihood is high enough for me to rationally form a belief that p ; and when I consider the likelihood of it being the case that $\sim p$ were I to believe that $\sim p$, it too is high enough for me to rationally form a belief that $\sim p$. Since varying the order in which I update my beliefs does make a difference for self-fulfilling beliefs, and—I have argued—evaluating and updating is an action insofar as we can control the whether or not to evaluate and the order in which we update, it is possible to intend to update in an order which makes one self-fulfilling belief take precedence over the other. This I believe is more or less the way that practical reasons for forming a self-fulfilling belief can bear on forming that belief.

⁸What makes a set of arguments count as 'sufficiently strong'? Perhaps this is under the direct or indirect control of the agent, or perhaps it is not. I don't think what I've said need commit me to a specific account, although I do assume the constraint that if your rational credence in p is $\leq .5$ given evidence e , then e does not count as a 'sufficiently strong' reason for believing p . I employ this assumption in the next paragraph.

2 Possibility of Promising

If Scanlon's account of promising were true, would it be possible to make promises? Before we can answer 'yes' or 'no', we need to say more about what sort of possibility we are considering. Perhaps we can imagine possible creatures who can rationally form beliefs at will. I'm not sure whether or not this is a possibility—regardless, it is not the sort of possibility we are interested in. On the other hand, we might want to know if promising is possible in some *specific* interaction between two people. This also is not what we're going to examine (although in the next section we'll look at some of the details relevant to answering such a question). We're going to take a middle course: Given Scanlon's account of promising, is promising possible for idealized human rational agents in an idealized environment?

In order to further clarify this question, let me state the idealizations I have in mind. The two rational agents should be ideal in that their process of forming beliefs and intentions should be like that of all other human rational agents, but not hindered by limitations of time, memory, or susceptible to lapses in rationality. Then, let us suppose that each has knowledge of the other agent's beliefs and intentions. Let us now add two background conditions directly relevant to promising. First, the promising agent knows that the promisee has an interest in being assured. Second, the promisee knows that the promisor would regard any promissory obligation as a strong reason for acting. Finally, let us suppose that all of this is common knowledge: agent A knows that agent B knows what agent A believes, the promisee knows that the promisor knows that the promisee has an interest in being assured, the promisor knows that the promisee knows that the promisor would regard any promissory obligation as a strong reasons for acting, and so on.

We're going to start looking at promising from the point of view of the promisee.

I offer you a promise that I will do X. If we explain how you can rationally form the belief that I will do X in part because I regard a promissory obligation to do X as a reason to do X, then we will have solved the circularity problem for the promisee.

Let p be the proposition that I will do X on account of me recognizing a promissory obligation to do X. You know that if you believe that p , then I will know that you believe that p . If you know that I believe that p , then I have a reason to make it the case that p . Thus, you know that forming the belief that p will be causally relevant in making it the case that p .

Since you know this, you can deduce:

- (i) If you were to believe that I will do X, then it will (likely) be the case that I will do X.
- (ii) If you were to believe that it is not the case that I will do X, then it will not be the case that I will do X.⁹

If I am able to rationally form the belief that Owen will pass the test, then you will be able in the same way to rationally form the belief that I will do X on account of me recognizing a promissory obligation to do X—the belief is self-fulfilling because it is a cause of the state of affairs which the belief is about.

If the circularity problem is solved for the promisee, then it is solved for the promisor. By offering a promise, I cause you to be *able to cause* me to be under a promissory obligation. This gives you an ability which you did not have before. We've supposed that I know that you have an interest in having this ability. Thus, it is in my power to assure you (i) by giving you a power to create a reason for me to do X and (ii) because I know that you have an interest in having and exercising this power.

⁹More precisely, replace 'will do X' with 'will do X on account of recognizing a promissory obligation to do X'.

In this idealized case of promising we posited that each party's beliefs and intentions were instantly known to the other party. We posited that the parties were ideal rational agents. We posited that the promisor knew that the promisee had an interest in being assured, and that the promisee knew that the promisor would consider a promissory obligation to do X as a strong reason in favor of doing X. Is promising possible without these idealizations?

3 Objections

Objection: Communicating a promise takes time. Suppose I send you a letter promising *not* to do Y. The promissory obligation doesn't exist until you accept the promise. But what if I do Y before you receive the letter?

Reply: Although the promise doesn't exist until it is accepted, if the promisor makes himself unable to fulfill a promise he has still violated the promise once it comes into existence. Suppose I know that a certain debt is going to come due, and at present I have the money to pay it. Would it be wrong to spend the money before the debt is due, when I know I won't be getting more money to pay off the debt? Yes, because I know that this will prevent me from being able to fulfill a future obligation. Likewise, it would be wrong to make yourself unable to keep a promise.

Objection: If when the promisee receives the offer of a promise, he is unsure whether the promisor is still in a position to fulfill the promise, then the promisee will not be able to form a self-fulfilling belief, because he does not know if his belief will be self-fulfilling.

Reply: In order for the promisee to be assured that the promisor will do X, the promisee must believe that the promisor will be able to do X. Frequently, this requirement is satisfied because the promisee has reason to believe that the promisor has a

reason for making the offer of a promise. Thus, the promisee has reason to believe that the promisor has reason to do his part. On the other hand, this means that it may not be possible to accept a promise when it's common knowledge that the reason I make the promise won't be sufficient to make the promisor do his part. In a moment of sobriety, your friend Tom sends you a letter promising to give up strong drink, but you know that if he were tempted before you received his letter, he would be likely to give in. It seems to me in that in this case we should say that you cannot accept the promise—and I can explain this by noting that you do not believe that the belief that Tom will give up hard drink would be self-fulfilling. Similarly, it is in the power of the promisor to cancel a promise by preventing the promisee from becoming assured. If I write you a letter promising to do X, but before you receive the letter I give you a call and let you know that I don't intend to do X, then it seems to me that I have done something infelicitous, but I did not violate a promise.

Objection: The promisor does not immediately know that the promisee has accepted the promise. The promisor is caused to act on a promissory obligation by the message that the promisee has accepted, not by the promisee's belief. Thus the belief is not self-fulfilling.

Reply: This objection misunderstands the sort of causal connection necessary for a belief to be self-fulfilling. As long as the the belief brings about its fulfillment by a causal chain, it is a self-fulfilling belief. The fact that the effect does not occur immediately after the cause is irrelevant. The fact that other conditions must obtain in order for the belief to cause the fulfillment is also not a problem, as long as the promisee can reasonably believe that these conditions will obtain.

Objection: The belief that the promisor will do X on account of a promissory obligation is not self-fulfilling if the promisor will act on the assumption that the promise will be accepted. Then, accepting the promise does not cause the promisor to act any

differently, so the belief does not cause its fulfillment.

Reply: Suppose the promisor is over-confident that the promise will be accepted, and that the promisee knows this. In this case, the promisee knows that regardless of accepting the promise, the promisor will do X on account of believing that he has a promissory obligation to do X. Here it seems to me that the belief that the promisor will do X on account of believing that he has a promissory obligation to do X is not self-fulfilling. But in this case, the promisee does not need the belief to be self-fulfilling: he knows that it is already true, so he has reason to believe it. What about the belief that the promisor will do X on account of *actually* having a promissory obligation to do X? There is a sense in which this belief is self-fulfilling: insofar as an agent cannot act for a reason which does not exist, creating the reason enables the agent to act for that reason (rather than act for the reason that the agent believes the reason exists). On the other hand, there is a sense in which the belief is not self-fulfilling: the agent acts because he believes that the reason exists, whether or not it does exist, so the existence of the reason does not make a difference. I'm willing to grant that if this is true, it may not be possible for the promisee to accept the promise. Consider a case where the promisee explicitly rejects the promise, but whatever the promisee tells the promisor, the promisor still persists in believing that the promisee really does accept the promise. In this case, there is no promissory obligation. On the other hand, if telling me that the promise is not accepted would make a difference in my actions, then the actual existence of a promissory obligation to do X does make a difference in whether or not I do X, and in this case we can say that your belief that I will do X (on account of actually having a promissory obligation) is self-fulfilling.

Objection: What if you don't have a reason to believe that I will regard a promissory obligation to do X as a strong reason for doing X?

Reply: I think in this case, promising would not be possible. If you believe that I am

a scoundrel, you will not be able to accept my promise, because my offer of a promise gives you no reason to think that I will do what I say. Promising depends on some degree of prior trust: you must believe that I will regard a promissory obligation to do X as a strong reason for doing X.

Objection: What about promising in a “state of nature” when I have no prior reason to believe that you are trustworthy?

Reply: Even if I have no reason to think that *you* in particular will regard a promissory obligation to do X as a strong reason to do X, it may nevertheless be reasonable for me to believe that you will. Consider beliefs based on testimony: it seems reasonable for me to believe that *p* on the grounds that you asserted *p*, even if I have no special reason to believe that you are honest. Likewise I think it is reasonable to presume that you are trustworthy, unless I have some reason to doubt this.

Objection: In most cases of promising, the parties involved have no notion of self-fulfilling beliefs, and yet they are still able to make and accept promises.

Reply: This is true—furthermore, in most cases of promising, the parties have never realized that there might be a circularity problem. In most cases, we simply intend to use promising to offer assurance, and accept the offered promise as a reason for being assured. The promisee does not need to consider the fact that his belief is self-fulfilling: he knows that accepting a promise does give the promisor a reason for fulfilling the promise, without considering how exactly the reason is created. The promisor does not need to consider the possibility of forming self-fulfilling beliefs: he knows that promisees are able to accept promises, and that if the promisee accepts the offered promise, then there will be an obligation to fulfill the promise.

On the other hand, if one of the parties believes that there is a circularity problem with promising and that the circularity problem does not have a solution, then this party will not be able to offer promises. If the promisee believes that the promisor

believes that there is a circularity problem, then he will not be able to accept a promise.

I have argued that Scanlon's account of promising does not make promising impossible for reflective rational agents who understand the nature of promising. Nonetheless, there are situations in which promising is not possible. Promising is not possible when it is common knowledge amongst the parties involved that they do not believe that promising is possible. Promising is not possible when the promisee does not already grant some degree of trustworthiness to the promisor. Promising is not possible when the promisee does not believe that the promisor will do his part in making it possible for a satisfiable promissory obligation to arise. I have tried to show how these limitations on promising do not pose an explanatory problem for my claim that self-fulfilling belief plays a part in promising.

References

- Hieronymi, Pamela. 2006. "Controlling Attitudes." *Pacific Philosophy Quarterly* 87:45–74.
- Kolodny, Niko and Wallace, R. Jay. 2003. "Promising and Practices Revisited." *Philosophy & Public Affairs* 31:119–54.
- Moran, Richard. 2001. *Authority and estrangement: An essay on self-knowledge*. Princeton University Press.
- Scanlon, Thomas. 1998. *What We Owe to Each Other*. Harvard.
- Velleman, J. David. 1989. "Epistemic freedom." *Pacific Philosophical Quarterly* 70:73–97.