

Promises to the Self¹

Abstract: The received view is that we can't make promises to ourselves in the same way that we can make promises to others, specifically, we can't obligate ourselves by self promise the way we do when we promise others. I argue here that the received view is wrong, and that we can make reflexive promises that entail promissory obligations. In the first section I clear the ground for the discussion by briefly outlining the nature of promissory obligations and proposing a methodology for discerning our intuitions on them in a given scenario. In the second section I introduce a hypothetical scenario wherein the protagonist appears to have made such a reflexive promise. I then defend the scenario against arguments that it doesn't prove the reality of self promise. In the third section I rebut a conceptual argument offered by Hobbes against the possibility of reflexive promises, on the grounds that such promisers could release themselves from their promissory obligations on a whim. I close by briefly outlining the import of self promises for traditional theories of promising, and I argue that the phenomenon calls for a new sort of promissory theory.

© 2006

¹ I would like to thank Tom Christiano, David Schmitz, and the two anonymous reviewers for this journal for their help and suggestions on earlier drafts of this paper.

I – Self Promises and Promissory Obligations

Can we make promises to ourselves? This is a question that has not received much consideration in the large body of philosophical work on promising. And in what commentary there is the answer is uniformly negative. I think this is a mistake, and that the conventional view that we can't make reflexive promises is wrong. I also think that this has some important implications for promissory theory in general. In what follows, I will attempt to argue for the first of these two claims, and to briefly outline my reasons for the second.

Historically, the question of self promise has historically been addressed directly by Hobbes,² and more recently by H.L.A. Hart³, R.S. Downie⁴ and P. S. Atiyah⁵. All of these discussions are very brief, and, with the exception of Downie, who in an aside calls self promises “dubious” and then says no more about them, all of them either explicitly deny that self promises are possible, or (as in the case of Hart) assume this implicitly, in the service of another argument.⁶

² Hobbes, Thomas *Leviathan*, Richard Tuck ed., (Cambridge : Cambridge University Press, 1991) : 184.

³ H.L.A. Hart, “Are There Any Natural Rights?” *Philosophical Review* 64-2 (1955): 175-191.

⁴ R.S. Downie, “Three Accounts of Promising” *Philosophical Quarterly* 35-140 (1985): 259-271.

⁵ P. S Atiyah, *Promises, Contracts and the Law* (Oxford: Oxford University Press, 1981): 54.

⁶ Although there is some record of support for self promises as well, see, e.g. Joseph Raz's appeal to them in his reply to Neil MacCormick in “Voluntary Obligations and Normative Powers – II” *Proceedings of the Aristotelian Society Supplemental Vol.* 59-99 (1972): 97.

I think this consensus is mistaken. I think that we can make binding (in the sense of morally obligatory) promises to ourselves. In what follows I offer a hypothetical case as an argument in favor of the reality of self promise. I then offer a rebuttal to two arguments offered against self promises and end with my reasons for thinking that the reality of self promises has bearing on the central questions of promissory theory.

Let me begin with a few prefatory remarks about promises and promissory obligations. Firstly, since I want to argue that certain instances of (apparent) self promise are in fact promises, and not metaphorical uses of the word, I should make plain what the difference between the two things is.

It is widely accepted that what crucially separates promises from nearby phenomena like expressions of firm intentions or notations of prior decisions is that promises produce obligations, where these other things do not. If I say to my neighbor: “I fully intend (or ‘I have decided’) to mow my lawn this weekend” I haven’t thereby given myself an obligation to do so. Whereas when I say to my neighbor “I promise you that I will mow my lawn this weekend” I have made it obligatory for me to do so. This production of a promissory obligation then is what I must demonstrate my example shows on the part of a self promise.

Secondly, promissory obligations of the sort I’m after are traditionally (with some exceptions) taken to be moral obligations. Whether or not this is accurate is a question I won’t attempt to answer here. Nothing much in my argument turns on this point, and so for the first part of the essay I will speak only of moral obligations. I will briefly take up the implications of this claim in the third section, when I consider the counter-argument to my position.

Third, promissory obligations have at least three commonly agreed upon necessary features that, taken together, set them apart from many other sorts of obligations. So in the interests of lighting the target let me consider these qualities for a moment.

Promissory obligations are *created* by the promiser when she makes the promise, *informed* by the utterance or expression that constitutes the act of promising, and can be *waived* by the promisee. A promissory obligation isn't a standing obligation, but rather one conjured into being by an act of promising. Once created, the promissory obligation is then grounded (at least *prima facie*) in that act of promising. So, if I have a promissory obligation to mow my lawn, I have an obligation to do so *because* I promised. Of course, promissory obligations can be co-extensive with other obligations. I can promise my neighbor to pay my taxes on time, and thus take on a promissory obligation to do something I already have an obligation to do.

What a promissory obligation is an obligation to do is a question also settled by the act of promising. To wit, a promissory obligation is an obligation to do what was promised in the act of promising.

And thirdly the recipient of a promise, the promisee, can cancel the obligation, generally merely by saying so. We can call this last the promisee release condition. So, a promissory obligation is an obligation that the promiser has to do what she promised, because she promised it. And the promisee can release a promiser from her obligation by simply saying so.

Finally on the topic of methodology: Since I will be attempting to make my case by an appeal to intuitive judgments about the existence of promissory obligations in a

given scenario, it would be good if I could use some diagnostic questions to prime the intuitive pump. And it would be better if those questions didn't contain reference to the phenomenon directly at issue, promissory obligations.

Towards that end, I will couch my diagnostic questions in the language of debt and owing. As the poet said "A promise made is a debt unpaid"⁷ and the comparison is more than just poetically apt. Debts, like promises, are obligations brought into being by deliberate action on the part of the debtor, and thereafter rest on that action. And, like promissory obligations, the obligation to pay a debt is one that can be forgiven by the debt holder, merely by saying so. Moreover what is owed is also determined by the initiating action. This is why we often speak of promissory obligations in terms of debt and owing. If my spouse catches me lazing on the weekend, rather than cutting the grass, after I had promised my neighbor to do so, she might well tell me that I 'owed' it to my neighbor to get out there and do my duty.

So I propose the following set of questions to determine whether or not a promissory obligation is present: 'Does the promiser owe the promisee the promised act?', 'Is the promisee owed the promised act?' And 'Can the promisee forgive the debt and waive his right to the promised act?'"⁸ I don't claim that these are the only questions that might be asked to determine whether a promissory obligation exists, only that in

⁷ Robert Service, "The Cremation of Sam McGee" *The Collected Poems of Robert Service*, (New York: Putnam, 1989).

⁸ Of course, while these questions avoid asking directly about promissory obligations, they do contain reference to promissory terms like promiser and promisee. But this is just a surface feature, for the sake of brevity. We could easily replace these terms with more cautious but clumsier phrases like "The person who made the ostensive promise" or "The person we might call the promisee, were this a promise".

cases where they are all answered in the affirmative, we have a good *prima facie* reason to believe that a promise has been made.

II – An Example of Self Promise

Let me turn now to the example I propose as evidence of the reality of self promise:

A sales manager, in an attempt to get better performance from her staff, says to them at the weekly sales meeting: “I promise you all that I will give \$100 to the salesperson who sells the most volume over the next week” (say the manager has been given a slush fund from upper management for just these sorts of promotions). And further, imagine that the manager is herself a salesperson, and that she means to include herself in the set of people who qualify for the money, and that everyone concerned knows all of this.

If, at the end of the month, it is the manager who has sold the most volume, has she then promised herself the money? I think that she has, but let me argue for this conclusion.

The first thing to consider is what we think of the case where another salesperson, say, Johnson, wins the contest. In such a case, is it fair to say that the manager has promised Johnson the money? Let’s consider our diagnostic questions. Does the manager owe Johnson the money? I think obviously so. Is Johnson owed the money? Again, yes. Can Johnson forgive the debt, and waive his right to the money? Again, it seems obviously so. So it seems to me that the Johnson case is an instance of a promise.

This tells us, at least, that if we want to hold that self promises are impossible, then the above example should help throw some light on why that must be, since the only

difference between the Johnson case and the reflexive case is the reflexive nature of the latter, and if the Johnson case is a promise, and the reflexive case is not, it can only be that quality that is responsible for the difference. But perhaps we can resist calling even the Johnson case a promise, and thus avoid the problem of explaining the difference between it and the reflexive case.

Towards that end, we might start out by noting that the scenario above is better described as a contest, rather than a promise. This I grant, but it doesn't yet help us deny the claim at issue. That the scenario is aptly described as a contest that alone doesn't mean that it isn't also properly described as a promise (or a set of promises) as well. In fact, it seems to me that this example shows that contests in general might best be described as consisting (in their normative aspects, anyway) of a set of promises, made by the prize givers to the contestants, that the contestants will get the prize if they win the contest.

But I don't need to commit myself to this view to rebuff the charge, since the episode can be re-described to make the contest aspect less salient. Imagine that, instead of informing them collectively at the meeting, the manager approaches each salesperson individually and privately promises them that, if they sell the most next week, she will give them \$100. And she also informs them that, if she sells the most, she will take the money for herself. In this version, it seems even more obvious that what the manager does is make contingent promises to the sales staff, and that the salesperson that satisfies the contingency has been promised the money.

Another difference between the Johnson case and more standard promises is that in the former, the promises are contingent upon the winning of the contest. But this fact

doesn't hold out much hope for denying their status as promises, either. If I promise my son a bicycle if he gets all A's on his report card, I think you would be hard pressed to convince him that, when he comes home with such a report, I haven't in fact promised him the bike. The mere fact that a promise might be made contingent on some set of circumstances isn't enough to deny that it is, in fact, a promise.

Granting then that the inter-personal version of the example is a promise, if we want to continue to deny the reality of self promises we must argue that there is a relevant difference between the inter and the intra-personal cases that explains why the latter isn't a promise. But let us first turn briefly to the positive case, and ask again our diagnostic questions.

Because I think the only real difficulty in making out the diagnostic claims is with the first one, I'll ask them here in a slightly different order. Firstly, is the manager owed the money? It certainly seems so to me. After all, if anyone has a claim to the prize money, it's the contest winner, and that's the manager. Can the manager waive her right to the money? Again, it seems obviously so. If she decides, for reasons of morale, or really for any reason, not to accept her prize, she can merely say 'I've decided not to accept the money' and that would be that.

But perhaps we might pause before we answer affirmatively to the first question 'Does she owe herself the money?' Not, I think, because we question the probity of owing something to oneself. It seems eminently plausible that people can owe themselves things in this non-metaphorical way. All that is required is that they have both the

authority to disburse a good, and the right to have that good distributed to them. Instances of such cases are mundane.⁹

Rather, I think the problem we might have here is the following. In the case where the manager wins the contest, we might doubt that she owes herself the money because we don't feel that she has an obligation to pay herself the money, in the same way that she has an obligation to pay Johnson the money in the case where he wins. The reason for this is that, if she were to fail to pay herself the money, we wouldn't feel that she had shirked a duty, whereas in the Johnson case, we would feel that such a failure on her part was such a normative transgression, a blameworthy omission.

There are two things that need to be said about this argument. The first is that while it is true that there seems to be an asymmetry between our feelings towards the manager who reneges on Johnson, and one who merely slights herself, the reason for this might not bear on whether or not the manager in fact has a duty to pay herself the money. Sometimes, our judgment about promise breakers is influenced by other normative features of the situation, features such as the nature of the promisee, or the nature of the promised act.

Consider the case where I promise to pay a visit to a very lonely, and much hated and evil, old man in my village. If I tell my neighbors about the promise, and they then come to find out that I have neglected to pay him the visit, they might well fail to think of me as blameworthy for having done so. If pressed, they might admit that I had shirked a duty, but if we ask only whether they feel the same way about me as they would had I

⁹ Consider, for example, the paymaster of an army battalion. When it comes time to pay the soldiers, certainly she owes herself her own pay, and not in a metaphorical way, the way she might owe herself a vacation, but in a straightforward way, the way she owes another soldier his pay.

renege on a morally unobjectionable person, the answer would be 'no'. Similarly, if I promise my neighbor's teen-age son to bring him back a carton of cigarettes from my trip abroad, if I fail to do so people might think differently about me than if I had failed to bring him back something more innocuous, like a souvenir. Again, if we asked specifically whether I had broken a promise, and whether that, considered in isolation, was a normative transgression, the audience would likely agree.

In both of the above cases, what explains the difference in judgment about the renege isn't that in the objectionable cases there wasn't a promissory obligation, but rather that other normative features, such as the vileness of the promisee, or the badness of the promised act, make our judgment about renegeing less harsh, or even approving.

The same sort of thing might underlay the difference between our judgments of the renegeing manager in the Johnson case and in the reflexive case. That is, the reason we might feel that the reflexive manager wasn't (or wasn't as) blameworthy is that the promised act, in the reflexive case, is an instance of self serving, and self serving acts have a generally unsavory normative reputation. This is not to say that there aren't morally correct self serving acts, only that the group, as a whole, suffers a taint, largely because of the prevailing views about attitudes like selfishness and entitlement. And this, I think, goes some way towards explaining why we might feel differently about the manager who breaks her promise to Johnson, and the one who breaks her promise to herself. In the latter case, the manager fails to perform a self serving act, and as such, I think we're much less likely to think her blameworthy. But of course, as in the above cases of odious promisees or bad promised acts, this fact doesn't alter the underpinning

normative fact of the promissory obligation. Rather, it mitigates our judgments about blameworthiness for shirking that obligation.

The second thing that needs to be said about this argument is that, as the promisee, the manager in the reflexive case also has the power to waive the obligation, or forgive herself the duty she has to pay herself the money. This might affect our judgments about her blameworthiness for a failure to pay herself the prize money in two ways. Firstly, we might be less disposed to blame her for not taking the money when she wins because we assume (without knowing the contents of her mind) that she has in fact released herself from the duty. In such cases of self release, the manager doesn't actually break a promise at all, and so it makes sense that we don't feel she has done wrong.

Secondly, in cases where it's clear that the manager hasn't released herself from the promise, we might be less inclined to see this as a breach precisely because taking the money would have been self serving. This intuition is further bolstered by the fact that the manager in fact had the power to waive the money, but failed to exercise it. This compounds the sin of self dealing with another sort of transgression, what we might call a deficit of liberality, or 'meanness', that is, a failure to exercise a normative power to waive an obligation owed us for the sake of graciousness. Consider, for example, a case where I bet my hot-headed brother-in-law a large sum of money (that I know he can't afford) that his team will lose the Thanksgiving-day football game. If I win the bet, and I fail to release my brother-in-law from his obligation to pay me, then I would be exhibiting the sort of pettiness that the manager who fails to waive her right to the money when she wins it does.

To control for these two confounding factors, I propose that we alter the example in the following way: Imagine that the prize for winning the contest is two-fold, the winner gets the money, but they also have to wear a large, ugly pin with the logo “# 1 Salesman!” in vibrant red letters on the front for the next week. And let’s assume that all the salespeople (understandably) hate the thing. In such a case, if the manager wins and fails to keep her promise to make herself wear the pin, I think it’s obvious that such a failure is a blameworthy transgression. But if she has a duty to wear the pin in such a situation, it can only be because she promised that the winner would do so.

So it seems to me that the above example is a case where a putative self promise produces a promissory obligation, and thus is, *prima facie*, a real promise.

III – Arguments *Contra* Self Promise

Now we can turn to arguments against the possibility of reflexive promises. The arguments in the record take two forms: from various sources (e.g. Downie, Atiyah,) we get broad appeals to the intuitive oddness of self promises, and from Hobbes we get the germ of an argument based on the promisee release condition. The latter is the one with which I will be principally concerned, but we should consider the former first, both because it is more prevalent, and because it can be rebutted more quickly. We can use Atiyah’s arguments as our exemplar.

In the course of considering possible counter-examples to the utilitarian theory of promissory obligation, Atiyah briefly considers ‘vows’ made to the self, and dispatches them with the following:

“It is very odd to regard a secret vow to oneself as creating an obligation of any kind. If it does, [it] is one without a corresponding right ... [and] which can be violated without risk of legal, moral or social censure.” (Atiyah, 1981, 54)

Atiyah lists two reasons to think that self promises are intuitively odd: that they lack a ‘corresponding right’, and that they can be violated without censure. On the first consideration, I take it Atiyah uses the term ‘right’ in this context to mean a claim on the part of the promisee held against the promiser for the performance of the promised act. If so then his assertion is that reflexive promises don’t (or can’t) produce such a claim. Atiyah doesn’t give any reasons for why he thinks this is so, perhaps because he takes it to be intuitively obvious. But it is this sort of intuition that my example was designed to rebut. Atiyah’s claim that a self promise produces no right corresponds to my first diagnostic question, ‘Is the promisee owed the promised act?’ If someone is owed something, I take it that they have a (*prima facie*) claim to what is owed. In our example, I think the sales manager is owed the prize money, thus she has a right (or a claim) to it.

As to the second consideration, that breaking self promises produces no censure, we need to first distinguish between two possible meanings Atiyah might have in mind. He might mean that since the promises are made secretly to oneself no one would know if they were broken, and thus no one could censure the breach. This we can grant, but it doesn’t reach very far. Firstly, not all self promises need be secret. Secondly and more importantly all secret sins are of this sort, that is they can be committed without fear of censure because they are secret. But we don’t take them to be less of a transgression because of this fact alone.

Alternatively, Atiyah might mean that we don't feel disposed to censure those who break promises to themselves, as we do when they break promises to others. Again, this is the sort of intuition my example was designed to rebut. I argued above that our initial reluctance to say that the sales manager owes herself the money is because we might be less quick to condemn her for violating the agreement if she herself is the potential beneficiary. This intuition in turn needs to be disentangled from competing considerations, and when we do this we can see that a reflexive promise breaker does something wrong.

But perhaps a more clear case is one where the self promise is more ameliorative. Suppose someone with an alcohol addiction takes a pledge of sobriety, adheres to it for a while, and then 'falls off the wagon' and returns to drinking.¹⁰ Is such a person guilty of a wrong that might merit censure? It certainly seems so to me, moreover it seems that the disapprobation we feel for such a backslider is over and above that which stems from our negative judgments about their drinking *per se*. To see this, consider the case where we come across our pledge and his drinking buddy, who has exactly the same drinking habit as the pledge but has never taken the oath, both drinking in a bar. We are right to feel some extra disapprobation at the pledge – over and above whatever general disapproval we might register for his and his buddy's alcohol consumption, because the pledge is committing an extra wrong – he's breaking his word.

¹⁰ This, of course, on the assumption that such an ameliorative pledge is best characterized as a promise to the self. I think this is so, but I don't have space to argue it here. If the pledge of sobriety isn't sufficiently plausible as such we can run the same sort of argument with whatever one might be most disposed to call a self promise (e.g. religious vow, New Year's resolution, etc.) I think the consideration I advance still holds for all such cases.

So neither of the considerations Atiyah advances in the cause of establishing the unacceptable oddness of reflexive promises holds up in the face of closer inspection. Self promises can generate a right in the promisee, as evidenced by the sales manager case, and self promises can produce moral censure if broken, as our lapsed teetotaler demonstrates. Now let me turn the second sort of argument against reflexive promises I find in the literature, the one from Hobbes.

Early on in his discussion of the nature of the civil law in *Leviathan*, Hobbes famously claims that the sovereign of a commonwealth is not and cannot be bound by its laws because he has the power to make and repeal those laws, and since “he is free that can be free when he will”¹¹ the sovereign is not subject to the laws. As an aside, immediately following this, Hobbes makes the same conclusion about individual self-binding: “nor is it possible for any person to be bound to himself, because he that can bind can release; and therefore he that is bound to himself only is not bound.”¹²

Let’s put the argument into explicitly promissory terms: If promisees can waive promissory obligations, and thus release promisers from their duties, it follows that in cases of self promise the promisee, i.e. oneself, can release themselves from the promise. But if this is so, then it seems that self promises cannot truly produce obligations, because obligations are the sorts of things that can compel us *against* (or perhaps, *regardless of*) our will. If self promises are such that the obligations they entail can be abandoned simply by an act of the will, the objection goes, self promises are not productive of moral obligations, and thus are not real promises. In other words, reflexive promises fail due to the promisee release condition.

¹¹ Hobbes, *Lev.*II-26:184

¹² *Ibid*

As an example, consider again the obligation that I have to my neighbor to mow my lawn. That is an *obligation* (the argument goes) precisely because I have a very good reason to do so regardless of how I feel or what I think about it. If I could just relieve myself of that obligation with a shrug of my shoulders, then it would be wrong to speak of it as an obligation. If instead of actually promising my neighbor to cut the grass I had merely promised myself that I would, then I wouldn't have the same sort of persistent reason to do so that I do in the interpersonal case. In the case of self promise I can, as the time approaches, merely release myself from the obligation, so the reason I have to perform the promised act is quite different, in that it is sensitive to my whims and desires.

We can sketch Hobbes' argument like this:

P1: All promisees can waive the promissory obligations of their promisers.

P2: In reflexive cases, promisers, in their capacity as promisees, can waive their own obligations

P3: Obligations are things that cannot be waived by those who have them

Conclusion: Self promises don't produce obligations, and thus aren't promises

This argument has quite a bit of intuitive traction. I can see two ways that a realist about self promises might answer. The first response would be to deny the first premise, that is, grant the point that self promises can't satisfy the promisee release condition but go on to argue that that alone is insufficient to bar them from consideration as promises. Towards this end we might point out that many promises needn't have such a condition, and indeed that one sort of promise, a deathbed promise, is such that a promisee release is

impossible¹³. From there we could argue that promisee release conditions aren't strictly necessary, and go on to try and claim that self promises are enough like promises in other respects to overcome their lack of this attribute.

And perhaps we could go further in support of self promises which can't be waived. Many of the most paradigmatic instances of (purported) self promise are ones where, intuitively, there is no possibility of promisee release: pledges of sobriety, devotional vows, New Year's resolutions, etc. These sorts of ameliorative self promises seem essentially irrevocable (e.g. you can 'fall' off the wagon, but you can't simply 'get' off it) and this fits in neatly with the claim that self promises can't be waived by their promisees.

But I think that self promise supporters can and should defend the possibility of promisee release, rather than abandon it in the face of the Hobbesian argument. I think this because of the sorts of issues raised by the case I outlined in the last section. If the sales manager really has made a promise to herself, then I think she obviously *does* have the power to waive it should she choose to do so. But such a power would be denied her if we grant that promisee release from self promise is impossible.

But in addition to the intuitive argument that it seems that the sales manager has such a power there is a more general theoretical reason to retain promisee release as a possibility for self promise, which is also suggested by the example. That is that the possibility of promisee release adds flexibility and power to the practice of promising, and this in turn makes the sorts of arrangements that can be made with promises more subtle and useful. That a promisee can rescind the obligation of a promise means that the

¹³ Although some contractarians claim that such promises aren't obligatory (c.f. Jan Narveson "The Desert Island Problem" *Analysis* 23-3 (1963) :63-67.

promisee has a certain normative power, one whose presence adds to the complexity of the overall arrangement, making more sorts of outcomes possible and foreseeable.

Furthermore, there are norms that attend the exercise of those normative powers such as promisee release, and those norms in turn make possible more sophisticated normative arrangements between promisers and promisees.

One crucial way this is useful is in cases where the motivating reason for a promise has changed or disappeared. Promises are forward looking, and thus it is always possible that circumstances change between the time of the act of promising and the time for the performance of the promised act in such a way as to substantially alter the nature of the promise. Sometimes such changes have the power to completely obliterate the promise, by making its performance impossible or otherwise untenable. If I promise to drive you to the casino next Wednesday, but the casino burns down on Tuesday, then I think it's safe to say I no longer have the obligation to drive you there.

But sometimes such changes don't so much remove the promise as make it otiose, or otherwise change its (rational) standing. If I promise to drive you to work because your car's in the shop and you have no other way of getting to work, then if your spouse later volunteers to take you to work that fact doesn't relieve me of my promise, although it does seem to change the character of the action we would call you 'holding me to' my promise. In such cases, the power you have to release me is eminently useful. Further, in this case it seems that the norms governing the use of the release power would mandate, if not the release, then at least an offer of release, and that you might be liable to criticism on those grounds for failing to do so.

And these same sorts of considerations apply in cases of self promise. Imagine that I vow to myself to go to law school and become a litigator in order to avenge the unjust prosecution of my father. If during the course of my studies I discover that my father was in fact given a perfectly fair trial, and that his conviction was a result of his guilt, then my reason for my having promised myself to go to law school has disappeared. This doesn't mean that my promise has, *ipso facto*, disappeared. Rather, it just means that my promise has been somehow rationally 'orphaned' by the loss of its motivating reason. In such a case, allowing me to release myself from the promise seems acceptable, and certainly better than the alternative, which would be to say that I am unable to release myself, and that even in the face of such a change my dropping out of law school must be described as 'breaking my vow'.

For these reasons I think we ought to maintain the possibility of promisee release from self promises in our response to Hobbes. So rather than deny P1, I think we should deny the theoretical principle that undergirds the objection, the one enshrined in P3: obligations that one can release oneself from at will are not proper obligations. Let us formulate a general principle of this sort:

GP1 All obligations must mandate duties that the person obligated cannot release herself from merely by an act of the will.

GP1 does have some initial plausibility, especially when we consider obligations like those entailed by the moral duty we have to help others in need. I cannot absolve myself of the obligation I have to help a drowning child by any mere act of my will. Such a suggestion seems to belittle the notion of a moral obligation.

But I think that in fact GP1 is untenable. In support of this claim, let me offer up some examples of obligations that one can release oneself from at will that are intuitively proper obligations. Consider a professor employed by a university. In such an arrangement, the professor can quit her job, at any time, and for any reason. Once the professor quits her job, actions that were obligatory for her before her quitting (teaching classes, attending committee meetings) cease to be obligatory. Thus she has the power to release herself from her obligations as a professor at her whim. But this fact does not entail that prior to quitting her duties as a professor were not, in fact, obligations for her. It seems patently reasonable to say that while she is a professor she has certain obligations that follow from her position. The fact that she can abandon the position at any time does not make those obligations any less obligatory, so long as she is still engaged.

It might be objected that the above example is an effect of contractual or legal obligations, and not moral obligations, of the sort promises are meant to be. Skeptics are free to claim that employment obligations are not like moral obligations in just this way—that they can be voluntarily cast aside. We can alter GP1 to reflect this:

GP2 All *moral* obligations must mandate duties that the person obligated cannot release herself from merely by an act of the will.

GP2 is more plausible than GP1, but it is still false. To see this, consider marital obligations. It is not difficult to imagine a marriage arrangement wherein one or both parties can dissolve the marriage by whim. Indeed, traditional Moslem marriage is a paradigm example of this,¹⁴ and modern secular no-fault divorce law only a slightly less

¹⁴ For a scholarly survey of divorce in Arabic and African Sharia courts, see *Women, the*

paradigmatic one.¹⁵ But certainly the mere fact that I, as a spouse, am able to release myself from the obligations of my marriage does not entail that those obligations do not, in fact, obtain while I am still in the marriage. It is more than plausible to say that I have an obligation not to sleep with other people while married, even though I could get divorced at any time.

Moreover, the sort of obligation I have not to sleep with other people while married is obviously a *moral* obligation. While it is also (commonly) a legal obligation, and perhaps a cultural one as well, intuitively it is unquestionably also a moral obligation. Witness the sorts of reactions that attend violating the obligation. The aggrieved party and the bystanders would all agree that the cheating spouse was doing something *wrong*, morally wrong, by sleeping with another.

In exactly the same way, it is not difficult to imagine a religion whose doctrines included the claims that proselytizing was a duty mandated by god, but for ordained ministers only, and which also had a mechanism by which an ordained minister could voluntarily cease to be one. In such a church, a minister has a divinely inspired obligation to share the gospel with others, so long as he is a minister. But when he resigns his commission, that obligation disappears.¹⁶ Still and all, he has the obligation while a minister despite the power he has to release himself from it. And it is obvious that, at

family, and divorce laws in Islamic history Amira El Azhary and Eliazbeth Warnock Fernea, Eds.

(Syracuse NY: Syracuse University Press, 1996).

¹⁵ For a brief history of contemporary American No-Fault divorce law, see Herma Hill Kay “Equality and Difference: A Perspective on No-Fault Divorce and Its Aftermath”, 56 U. Cin. L. Rev. 1- 80 (1987).

¹⁶ Such minister-only but voluntary duties are not a flight of fancy. Most churches have duties that fall solely on the clergy, and most also allows their prelates to voluntarily leave the priesthood.

least from the perspective of the religion, the obligations the minister has while a minister are moral obligations.

What these examples show is that the principle that undergirds the objection to realism about self-promises cannot stand. The fact that it is within the power of an obligated person to release herself from the obligation does not entail that the obligation never was at all, as the counterexamples to GP1 and GP2 above demonstrate.

In response to the foregoing we might object: In the cases described (quitting a job, getting divorced, leaving the priesthood, etc.) what the protagonist has done is not ‘waive’ an obligation, but rather end a reciprocal relationship. Further, it is the existence of this relationship, and more specifically, its reciprocal nature, that is the ground of the obligations so ‘waived’.¹⁷

On this understanding, the duties to remain faithful to a spouse, or to perform the tasks of your employment, are duties in light of the fact that one has entered into a reciprocal arrangement with others, an arrangement of mutual benefit, and the fact that one is enjoying (or has enjoyed) the benefits of the arrangement. In particular this last is what grounds one’s duties to perform the tasks specified in the arrangement. I have a duty to remain faithful to my spouse because I enjoy the benefits of the spousal arrangements, i.e. the marriage. When I divorce my spouse, I don’t ‘waive’ those duties, rather I dissolve the relationship, and since such dissolution means I lose the benefits of the arrangement, I simultaneously lose the (normative) reason I had to do the duties I did before.

¹⁷ This objection was first brought to my attention by Thomas Christiano, in his comments on an earlier version of the paper.

This explanation seems most plausible in the employment case, since there is something vaguely cold-blooded about the same thoughts in the context of marriage or the priesthood. At least we can say that it sounds plausible that the ground of the duties I have to perform the tasks of my employment is the fact that I enjoy the benefits of the employer/employee relationship, which is explicitly reciprocal.

But even in these cases, it seems to me that the ground of the duties can't be the enjoyment of the reciprocal benefits of the arrangement, for the following reason: the provision of those reciprocal benefits is not a necessary condition of the obligations obtaining. In other words, I still have my work duties, even if my employer reneges on his obligations to me under the terms of my employment.

To see this, consider the (unfortunately) not-merely-hypothetical case of a civil servant in a poor country, whose government occasionally neglects to pay his salary, sometimes for months on end. During these interruptions, I think we can safely assume that the employer is abrogating its central reciprocal duty of the employment relationship, i.e. remuneration. If there is anything that counts as a benefit that employees enjoy that might ground their duties in the manner described above, surely pay is one such thing. But just as surely it is true that so long as the civil servant doesn't quit his job, he retains his work duties, despite this failure on the part of the employer.

Granted, what the jilted employee gets as a result of the failure is a *reason* to quit, indeed perhaps even *the right* to abrogate his duties, but what he does not get is the *lifting* of those duties. And this can be seen when we contrast the worker who remains on the job with one who has quit. If the former doesn't go to work, he is doing something that the latter is not, but what could this be but 'neglecting his duty'? He might be said to

have a justification for his actions, or perhaps an excuse, but the fact that such terms apply at all is evidence that what he is doing is (*pace* those apologies) a normative transgression.

Perhaps we might counter that the enjoyment of benefits in the reciprocal relation should be measured over the length of the relationship, and thus is not to be assessed at every point in time during the contract. But this proposal suffers from at least two very difficult problems.

Firstly, it makes an assessment of whether an employee has a duty to perform his assigned tasks impossible until the end of the relationship. Since, if the ground of the duty is reciprocal benefit, measured over the length of the relationship, then there must be some amount of benefit under which the obligation doesn't hold. But we won't know whether that limit is breached until all the data are in, i.e. until the relationship is ended. This means that we won't be able to answer the question of whether a current employee, at any point in the relationship, actually has the duties, since we won't know the final tally until the end.¹⁸ But this is absurd.

Secondly, since there must be some limit of benefit at which the obligation moves from being grounded (i.e. extant) to not, and since that limit is based on the length of the relationship, and since the length of the relationship is based (at least partly) on the will of the employee, on this view employees could determine whether or not they ever had their duties by quitting at different times. If civil servant A, during a salary interruption, quits just prior to the limit being reached then he always had his duties, since the overall

¹⁸ Indeed, it seems worse than this, for the problem isn't merely epistemological, but rather metaphysical. It's not that we can't *tell* whether an employee has the duties, but rather that there is no *fact of the matter*, until they have completed their record.

relationship was sufficiently beneficial to him. However civil servant B, who quits just after the limit is reached, never had the duties. But again, this is absurd.

And all these considerations go doubly for the cases of marriage and the priesthood. It's even more obvious in the case of a spousal relationship that the withholding of reciprocal benefits doesn't obviate the jilted spouse's duties. If a spouse cheats, or moves out of the house, for example, this action doesn't mean that the other partner is thereby relieved of their marital duties. Again, they might have reason to abrogate those duties, perhaps even justifying reasons, but they still have the duties. As before, compare those spouses so abandoned, who remain married, to those who get divorced. If the former take a lover, they do something that the latter, similarly disposed, do not, they 'break their vows'.

Another possible objection is that the sorts of examples I have given so far concerning the obligations of spouses, employees and ministers are relevantly different from promissory obligations, in that we waive those former obligations 'indirectly', by resigning a certain role or position. When I divorce, for example, I waive my obligation of fidelity by ending my role as a spouse.¹⁹

In support of this, we can note that the actions necessary for the sort of waiving that occurs in the examples, e.g. resignation and divorce, are properly described as ending (or leaving) a role. And that as such these sorts of actions (typically) result in the disappearance of the whole complex of rights, duties, privileges and powers that attend the role. So when I quit my job, I don't merely waive my obligation to come to work, I also lose my right to read the company's internal documents, my power to order their employees around, my parking space, etc.

¹⁹ This objection was pointed out to me by an anonymous referee for this journal, for which I am grateful.

Making a simple promise (the objection continues), doesn't seem to involve taking up a 'role', at least not of the same sort as that of spouse or minister. And if making a promise isn't assuming a role, then waiving a promissory obligation can't be accomplished by resigning that role.

Let me begin my response by noting that in terms of the dialectic, this objection doesn't reach my original argument *contra* Hobbes that we can have moral obligations that we can waive. Rather it concedes that point, but goes on to note (what it claims is) a salient difference between the *sort* of waiving occurring in my example cases, and the sort that would occur in a case of reflexive promise. This is a counter to an argument not so much made, but implied in my actual argument: That the examples I list aren't just counter-example demonstrations of obligation-waiving to falsify the universal Hobbesian claim, but rather positive examples of the possibility of reflexive-promise waiving, i.e. something like analogies of such promissory reflexive waiving.

Now of course supporters of reflexive waiving don't need to make the further claim. They can simply concede the point that the examples aren't proper analogies, and be satisfied with the concession on the Hobbesian claim. Having said that however, let me try and supply an example without roles of the sort discussed above:

Suppose you have a pile of scrap wood cluttering up your small yard, while I have a large empty expanse of land in the country. You ask me to keep your wood at my place, and I agree, and you move the wood to my place. Later on, as I am trying to sell the property, I decide that the wood is an eyesore and I ask you to remove it.

The above seems to me an utterly typical arrangement of the sort reached between people all the time, with a similarly typical ending: two people agree on an arrangement of unspecified duration which either can rescind at any time without penalty, and then at

some later time one of them does so. But what is also obvious is that the above agreement comes with attendant moral obligations. Until I tell you otherwise, I have a moral obligation to keep your wood, because I have agreed to do so. This fact is reflected both in our social practice and in the law. If, for example, someone were to vandalize or steal it while it was in my possession you would certainly have a moral claim against me, and very likely a legal one as well (although the precise outlines of such things vary by jurisdiction).

And just as obvious is the fact that I can waive those obligations, at any time, for any reason, by simply telling you that I wish to. So what we have is an instance of moral obligations that can be waived without resigning a role.

Of course, it is open to the opponent to claim that my keeping your wood constitutes the assumption of a role, perhaps the role of ‘landlord’ or some such, and thus the waiving here is another instance of role resignation. But if we take that tack I don’t see how we could arrive at a (principled) definition of ‘role’ that would include wood-keeping while excluding promising.

But perhaps this argument is pointing us to another sort of objection, one that starts from the opposite assumption to the one above, and takes it that promising *is* a role, and that the role itself prohibits reflexive waiving. The wood-keeping example highlights the difference between a promise and an agreement. It isn’t violating the agreement to ask you to remove the wood (although it would be for me to dump it without your permission). Yet it would be breaking a *promise* to do so, had I so promised. Promises, unlike agreements, can’t be revoked by the promiser- that much seems intuitively clear.²⁰

²⁰ Scanlon, for example, uses this fact as one of the reasons to go beyond the traditional expectation-based view of McCormick and others, cf. *What We Owe To Each Other* (Cambridge: Harvard University Press,

But if this is so, it would seem that reflexive promissory waiving can't be allowed, since it amounts to allowing promisers to rescind their promises.

But supporters of reflexive waiving can agree that promisers can't rescind their promises without giving up the position, by pointing out that promisees *can* waive the obligations of promisers. As we saw in P2 of the sketch of Hobbes' argument in section III above, what happens in the reflexive case is that the promiser, in her capacity as promisee, waives the obligations of the promiser (i.e. herself).

This sort of dual-role activity might sound odd, but it is actually quite mundane. Our army paymaster of the previous example (fn. 8) has (at least) two roles, one of paymaster and one of a soldier of her rank (say Captain). The former role has powers (and duties) the latter does not, and so she can do things in her capacity as paymaster that she would be forbidden from in her capacity as a captain - find out the salary and paycheque figures of all the enlisted soldiers at the base, as an example. And while it's perfectly reasonable to treat the claim that Captains aren't allowed to ascertain the pay of random enlisted soldiers as true, in cases where Captains are also paymasters it's false. Similarly, it's reasonable to claim that in general promisers can't rescind their promises, but in reflexive cases the rule doesn't hold because of the dual role of the promiser as promisee.

But what are we to make of the residual intuition that obligations that can be shrugged off aren't proper obligations? I think that this still has a grip because of our desire to criticize those who abandon obligations improperly. When a spouse decides to divorce for a passing sexual urge, when a minister abandons his calling out of irritation with the laity, when anyone exercises the power to release themselves from an obligation

too lightly or for the wrong reasons, it is always open to us to criticize them for the exercise of that power. What is not open to us is to criticize them for failing to perform the acts that were obligatory for them before the exercise of the power.

We can criticize the president of a bank for resigning in the face of a hostile board, rather than rising to the challenge, but we can't criticize her for not chairing the board meeting after she has resigned, since it is no longer her duty to do so. Similarly, in the case of self-promises, we can criticize someone who releases himself from a promise for what we think are inadequate reasons, but not for failing to perform the actions that he had promised himself to do after he has released himself from the promise. If I promise myself to complete law school, and then later release myself from the promise and quit school, I might be criticizable for my decision to quit, but not for not showing up to my contracts class after I have quit.

This is the same distinction between the shirking of a duty, and the proper exercise of a power to rid oneself of a duty, that we saw earlier in the cases where the sales manager failed to waive her obligation to pay herself the winnings, and the case where I fail to release my brother-in-law from his ill-advised wager. In those cases, I noted that we might well find the protagonists' actions blameworthy, not because they breached a promissory duty, but rather because they failed to exercise a normative power properly. In the earlier cases the protagonists fail to use their powers when they ought, in the cases above of divorce and resignation they exercise their powers when they ought not. But in all the cases the efficacy of the power is never in question, and neither is the reality of the obligation over which the power is held.

IV – The Upshot of Self Promise

So it seems that the Hobbesian objection does not stand, and this clears at least one more hurdle on our path to making self promises theoretically respectable. As I said at the outset self promises have been largely overlooked in the literature on promising, and what attention they have received is largely negative. But I hope that my arguments above have demonstrated that, far from being an obvious non-starter, the idea of reflexive promising is in fact quite well supported. Furthermore self promises, if I am correct, play an important role in the promissory universe.

Self promises have import for promissory theory in two ways. Firstly, the fact of self promises is a window into the larger world of what we might call non-paradigm promises, where the paradigm is the sort of promise I make to my neighbor to cut the grass, i.e. interpersonal, and generally made to reassure the promisee of the promiser's performance. I think that the traditional theories of promising suffer from a blinkered focus on these sorts of promises, and that the resulting theories are ill-equipped to handle alternative sorts. And there are quite a few different sorts of promise, besides the kinds mentioned so far. Pledges of sobriety, New Year's resolutions, Oaths of office, all of these and more are, I think, species of promise, but none of them fit easily into the mold of promise envisioned by the traditional theories.

The second way that self promises impact traditional promissory theory is an instance of the first, in that the obligations generated by self promises aren't easily explained by the sorts of grounds that are invoked to explain paradigm promises. Briefly, theories of promissory obligation are generally divided into two broad classes, those that

ground the obligations in the expectations of the promisee (and the promiser's responsibility thereof),²¹ and those that ground them in the utility of the social convention of promising.²² As I have argued at length elsewhere²³ these grounds don't readily support self promises.

In the case of expectational theories, it seems unlikely that the reason a self promiser might be obliged to keep her promise is that by promising, she fostered an expectation in herself that she would do what she promised. Not only is this sort of expectation building not the purpose of self promises, unlike paradigm ones, but in cases of renege on self promises, the promisee is responsible for the fact that the promise went unfulfilled, and as such, can't claim that wrong was done to her, which wrong is typically taken to be the ground of the promissory obligation on expectational theories.

Conventional theories fare little better. They usually explain promissory obligations as derivative on other obligations to maintain (or not take unfair advantage of) the promising convention, where these more basic obligations to the convention follow from its utility in fostering social cooperation. But self promises aren't essentially cooperative in this way (consider, what extra cooperative advantage has the sales manager who includes herself in the contest over one who doesn't?). As such, conventionalists will have a difficult time explaining why self promises are promises at all, or at least why, if they are promises, they are grounded in the cooperative utility of the promising convention.

²¹ Cf. Scanlon (1999): 295-316.

²² Cf. John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971): 344-348.

²³ Allen Habib, *The Authority Theory of Promises*, Doctoral Dissertation, University of Arizona, 2007.

So the above arguments for the reality of self promises isn't, I submit, merely a call for a minor amendment to the traditional theories of promising, but rather a reason to abandon them and start a search for a new theory that can accommodate self promises and their non-paradigmatic brethren.