The Agreement to Keep Our Agreements: Hume, Prichard, and Searle

by Jan Narveson
University of Waterloo
1988 Learned Societies meetings, Windsor.

Abstract

Does it make sense, and is it at all plausible, to view the moral obligation to keep particular promises and do what is called for by particular agreements such as contracts as being founded on a general "Social Contract" -- i.e., to give a contractarian account of promise-keeping? This paper argues that it does. Borrowing from Hume, David Lewis, Gilbert Harman, and David Gauthier, I provide a sketch of what the "social contract" is (not, e.g., either a real or a hypothetical meeting of all with all) -- namely, a form of commitment, makable by any individual, but with commitment likewise to social reinforcement. Then it is argued that Searle's familiar thesis is in error in that it leaves out the latter factor as well as implicitly calling for a version of the former. Then it is argued that this general understanding of morality can reasonably issue in a specific rule about agreements that would be one among others in a contractarian moral platform.

The Agreement to Keep Our Agreements: Hume, Prichard, and Searle

1. Introduction

An interesting question about the Contractarian account of morals was first (to my knowledge) raised by David Hume with the observation that "the observance of promises is itself one of the most considerable parts of justice; and we are not surely bound to keep our word because we have given our word to keep it."[1] Again Hume, cross-examining a supposed would-be contractarian, "If the reason be asked of that obedience, which we are bound to pay to government, ... Your answer is, *Because we should keep our word*. But, besides, that no body, till trained in a philosophical system, can either comprehend or relish this answer; besides this, I say, you find yourself embarrassed when it is asked, *Why are we bound to keep our word?* Nor can you give any answer but what would, immediately, without any circuit, have accounted for our obligation to allegiance."[2]

Much later, Prichard puzzled over the question of the obligation of promises, concluding that the obligation to keep any particular promise must rest on something that is very like a promise to keep promises, but somehow not *quite*: "The general conclusion which I wish to suggest ... is that promising to do this or that action, in the ordinary sense of promising, can only exist among individuals between whom there has already been something which looks at first like an agreement to keep agreements, but is really an agreement not to use certain noises except in a certain way, the agreement nevertheless being one which, unlike ordinary agreements, does not require the use of language. .... I am suggesting .. a problem for consideration; viz. what is that something implied in the existence of agreements which looks very much like an agreement and yet, strictly speaking, cannot be an agreement?"[3]

Later still, and most notoriously of all in terms of its impact on then-contemporary ethical theory (1964), there was John Searle's famous essay, "How to Derive an "Ought" from an "Is",[4] in which he seems to have[14] argued that one could logically derive the conclusion "Jones ought to pay Smith $5" from the premise, "Jones uttered the words 'I, Jones, promise to pay you, Smith, $5', together only with supplementary factual and clarificatory premises, plus moral premises asserting only that other overriding obligations do not obtain. The point is that Searle's argument, if it worked, would presumably have had the effect of showing that no further *moral* premises, e.g. to the effect that promises must be kept, are required. A fortiori, we would require nothing like a general "promise to keep promises": promises have, intrinsically, the power to create obligations *ex nihilo*. They are, so to say, logically self-confirming.

The question that concerns me here is whether the Prichardian intuition can be sustained in a sensible way, leaving us with a respectable account of the obligation to keep promises that leaves it recognizably a branch or theorem of the Grand Social Contract. Does it, in short, make sense -- and if so is it plausible? -- to hold not only that one among other of our basic moral requirements is to keep our promises, but also that the foundation of this is in some sense a "social contract"? Hume argues that this is to try to have one's cake and eat it too; Searle supposes that the attempt is misguided because unnecessary; and Prichard obscurely intuits that some such thing is needed, but is unable to articulate it to our satisfaction. Can we steer between the Scylla of circularity and the Charybdis of redundancy? I shall propose a rendering that, unlike Prichard's, lacks obscurity and yet has the desired effect of avoiding both of these nasty rocks.

2. Contractarian Theory

The contractarian holds that the fundamental principles of morals are in some way the objects of an "agreement", which in turn is the outcome of what is in some sense a "bargain". In trying to make clear and plausible this approach to morals, it is obviously of crucial importance to identify the sense(s) or way(s) in which these things are supposed to be so. For one thing is surely clear enough, and that is that the quotation marks are appropriate: the "agreements" and "bargains" of which the contractarian moral theorist speaks are not

your ordinary examples of these types. The Grand Social Contract is, obviously, a model or idealisation of some kind. Thus, for example, such a "contract" clearly cannot be the result of any sort of actual gathering; it cannot have been "negotiated" en masse, not even by representatives. And there is the familiar problem that any agreement that actually was made by past actual people could not as such bind present nonsigners, even had it managed unanimity at the time. For this and other reasons, it has become standard procedure to hold that the "Social Contract" is hypothetical, a matter of what we would sign if we were in a certain condition.

But if we say this, we invite new problems. For just as agreements made by others do not bind, neither do agreements that we didn't make, however true it may be that we should have made them. If we are to make social contract theory work, it is clearly essential to understand the "agreement" in some other sense than that of a formal promise or literal contract, be it historical or hypothetical. And a further reason, specifically pertinent to the present inquiry, is that we would want the principle of keeping one's contracts to be itself founded on The Social Contract.

We need to know at least two things, therefore. First, we need to know just what actual acts, or mental processes, or whatever else, these things are supposed literally to be in contractarian theorizing. It is no good talking airily of hypothetical bargaining situations and bargains unless one can locate suitable analogues of these things in the real world, where the action is -- and moral theory is, after all, concerned with action. Second, we need to know why and how these "agreements" should be thought capable of playing so fundamental a role. Presumably the idea that they can do so is got from the intuition that ordinary agreements and bargains do have some moral force. So the question is whether the way in which they have it is transferable to the very fundamental or global level, the level at which the contractarian wants to apply these notions, without loss of effect. This in turn divides into two issues: (1) whether the "moral force" that ordinary bargains have is enough to operate at this level, and (2) whether the special circumstances or special properties of the contractarian's fundamental bargains and agreements are such as to preserve that force, or whether instead they undermine it. These basic issues are not investigated in this specialized inquiry, but some sense of direction on them is essential to it as background.
Gauthier's new Morals by Agreement[5] is outstanding for its treatment of these two fundamental issues. His answers as I understand them, are as follows. To the first he replies that moral force is simply the force of practical reason. Whatever reasons we have for doing virtually anything are, in effect, our reasons for being moral. And to the second he replies that there isn't anything terribly "special" about the circumstances of the "fundamental agreement(s)". In contrast with the likes of Rawls, for instance, no special attitude, no "veil of ignorance", is required. All that is required is that we are dealing with similarly rational individuals. In cases departing radically from that paradigm, we simply have no moral duties and revert to unlimited seeking of advantage in whatever ways it is possible.

In the Gauthieran construction, which I am now inclined to accept, the "Social Contract" is a 'contract' in the following sense: people discern that if they interact under the unrestricted motivation of their various separate interests, then they will do worse than could be done if instead they were to adopt certain restrictions or constraints on the pursuit of those interests. A main context in which such restraint is called for is that of social interaction. Here it is often the case that the unlimited pursuit of advantage by each leaves all worse off than if they had instead adopted restrictions on individual pursuit of utility. Constraining conduct promotes maximization, even though maximization is precisely what it constrains.
The secret is in mutuality: each must shape her behavior by refraining from certain erstwhile maximizing actions on condition that the other does likewise. It is the necessity of this condition that makes the "Social Contract" a contract. "I shall do this, provided he does that"; "I shall do that, provided she does this" -- and that goes for both parties. It is rational, so the contractarian argues, to act as dictated by the constraints, which tell us to refrain from out-and-out maximization when interacting with others similarly constrained. Each comes out of it better off, but the condition on which each is better off is that the other has refrained from attempting to become still better off.

It is widely disputed that that is rational, to be sure. But a tamer objective is sufficient for my construction: it is rational to bring it about, by erecting suitable reinforcing devices, that everyone respects those constraints, even when they are likewise imposed on oneself. Moral reinforcement is just such a set of reinforcing devices.

We should note that on this more sophisticated view, the "Social Contract" requires no founding moral constitutional conventions or the like. Those we deal with either do or do not meet the condition. If they do, then we deal with them in one way (cooperatively, with restraints); if they don't, then we stay in the "State of Nature" and simply do our best against them. The thesis that morality in general is "founded on" such a contract is the thesis that morality has as its raison d'etre the promotion of our general goals, and that its substance will therefore consist in self-imposed restraints whose mutual observance can be expected to work out better in terms of those very goals than would unlimited independent goal-seeking. Its general form is: "I will do x, which I would not otherwise do, provided that you do y, which you in turn would not otherwise do". The essential conditions are what make it contractual. Absent your adoption of the conditions, it is not rational for me to adhere to them; absent mine, it is not rational for you

to adhere to them. But the "contract" takes place in the heads of each contractor, A, in light of A's perception that those A deals with are similarly disposed. To be moral is to have the disposition to act in the constrained ways in question; since this will often conflict with the unlimited pursuit of one's intrests, it requires the sort of internal discipline we associate with morality.

3. Searle

To see how this applies to the question of promising, we turn first to Charybdis, as represented by the now classic article of John Searle, "How to Derive an "Ought" from an "Is". (The foregoing has partly dealt with Scylla, in the form of Hume's opening salvos; we will return to that below.) According to Searle's argument, there is a logical train going straight from the premise,

(1) Jones uttered the words, "I hereby promise to pay you, Smith, $5",

to the conclusion,

(5) Jones ought to pay Smith $5,

via the intermediate steps:

(2) Jones promised to pay Smith $5

(3) Jones undertook an obligation to pay Smith $5, and

(4) Jones is under an obligation to pay Smith $5.

In a previous treatment[6], I argued that whereas we may reasonably accept that (2) follows from (1) and (5) from (4), it is perfectly clear that (4) does not follow from (2) and thus (5) doesn't follow from (1). I start by presenting much the same criticism here.

My reason for denying this is different from some recognized by Searle himself. Searle was careful to observe that (2) does not follow from (1) all by itself: the words could be uttered in jest, or in ignorance of their meaning, for instance. And even where utterance is perfectly sincere, there is the question whether we should say, in innumerable cases where we would deny the conclusion despite recognizing the truth of the premise, that one who in those circumstances utters those words, however sincerely, does not really 'promise' - e.g. when promising against a grossly insufficient background of relevant information). On this point of ordinary language, as I suppose it is, let us simply accept any reasonable view: that he did indeed promise, but other things weren't equal, or that indeed he didn't really promise under the circumstances. Here we have other fish to fry. Namely, we will question whether it really does follow, straight off, that the promisor undertook an obligation?

Here we may also note Searle's agreement that there could be other moral considerations militating against the doing of what one promised despite the uncontroversiality of one's having promised it. This is to say that (5), our conclusion, does not follow from (4). Let us likewise waive any such problems. Assume that we have a promise where no contrary moral requirements are operative. Well, I still want to know whether it follows, absent any prior moral premises, that one has a moral obligation to do x given only that one has promised someone to do x.

Let us re-focus our attention on the transition from (2) to (3): if A promised B that A would do x, then is it the case that A "undertook an obligation to do x"? This depends, in major part, on whether we understand 'undertook' as having only what used to be called "illocutionary force" or not. In promising, Jones implies that he intends -- he at least gave B the impression that he intended -- to do something the result of which is that he has a certain obligation. In normal cases, we may agree, he certainly supposed he was doing that: he was trying to do that. If you promise to do x, then you perform the illocutionary act of (attempting to) undertake an obligation. In thise sense of 'undertook', he undertook an obligation. But in that sense, one can fail to do what one undertook. That's the case here: from none of what we have so far (the utterance of words under the intention that one wind up with an obligation) does it follow that you actually succeed in doing that -- thjat your eally do wind up with that obligation. What more is needed?

Searle's paper inspired a considerable amount of scholarly reaction, most of it leaping to the defense of the Naturalistic Fallacy principle; but this, I suggest, misses the point. The camel's nose under Searle's tent is easy enough to overlook, actually, because in typical cases it will be so obviously present as to require virtually no notice. What wrecks the argument as it stands is that it is plainly essential, at a minimum, that the promisee be understood to agree to the promise. No amount of verbal ritual on the part of the promisor, acting by himself, will create the needed obligation. (In fairness, I must note that Searle in his later treatment of the subject substantially recognizes the necessity of this, saying even that "The notion of an obligation is closely tied to the notin of accepting, acknowledging, recognizing, oundertaking, etc., obligations in such a way as to render the notion of an obligation essentially a contractual notion".[7])

There was a time when people attached great moral significance to the making of vows, which are rather like one-person promises. But it is surely obvious that a vow does not, all by itself, create moral obligation. If I vow your destruction tomorrow, I do not in consequence have any obligation whatever to bring it about. If I vow to paint my hair green or to lose forty pounds by Christmas, no moral obligation results, whatever may be my personal obligation. No one can complain if I "default" or simply change my mind.

In most cases, consent is obvious, automatic, or can reasonably be assumed and needs no explicit recognition, which no doubt is what makes it so easy to overlook. Indeed, wherever what is promised is

something not obviously downright to the disadvantage of the promisee, it would be reasonable to say that consent can be presumed. Following Baier[8], we could say that promising creates a presumption of obligation; but this presumption is negated if the promisee makes it clear that he wants no part of what is promised. If, to take an example offered by one of the referees of this paper, my grandmother promises me her horsehair divan, which I don't want, then supposing I have enough moral fibre to refuse the old dear (which most of us do not!), then when the delivery truck comes I can quite properly say that I want nothing to do with it. She has no obligation to deliver it to me so far as I am concerned -- and where *I* am the person promised, who else can be more so? My promise to give you an F cannot create a valid obligation on me to give you an F if you don't want an F or don't want one from me, for whatever reason -- apart, of course, from the special case where you have an independent obligation -- as you might, in the former case, to your grandmother -- to accept x's being done. If that's so, then there simply is no such obligation, no matter how fervently I utter the magic words "I promise". If Searle's intention was to argue that the pragmatics of the linguistic act of promising were of themselves sufficient to create the moral obligation that we normally associate with promising, then I suggest that h.

But that's not all. Even a promise with full consent, in the absence of further information, is not sufficient to create a valid moral obligation. Consider the "gangster promise": A promises B that he will assist her in some nefarious enterprise. Here there is both promise and consent -- but is there *moral obligation?* Surely not. There may well be a *sense* of obligation. But the sense in question is in error in the most fundamental respect. When A promises B to assist B in torturing innocent C, for instance, B may well feel let down when A has a change of heart about this detestable deed, and quite possibly B has *been* let down, as well he ought to be. But to suggest that A has here a prima facie obligation which has, however, been outweighed by the greater obligation not to torture people is grotesque. It simply never was the case that A had any moral obligation, *prima facie* or otherwise, to hold poor C down while B applies the hot poker. (On the other hand, my prima facie obligation to bring some flowers to poor sick Martha may be outweighed by my stronger prima facie obligation to get home to dinner on time. There the language of prima facie obligations being overridden is comfortable and appropriate.

4. The Social Factor

What is needed is not only the promise and the consent, but a third thing, viz., what we might call - for the very good reason that that's literally what it is - the Consent of the Community. And the reason is similar, if not quite identical. The existence of a moral obligation is not something that only involves two parties, or twenty; it is instead something that involves the whole relevant community. To go from the utterance by Jones of "I promise to pay you, Smith, five dollars" to "Jones has a moral obligation to pay Smith five dollars", we need intervening premises which give the rest of the community *reason to uphold* the obligation in question. If I, Robinson, a quite uninvolved party, am to accept the judgment that you, Jones, owe Smith five dollars, then *I* have to make a moral judgment, to the effect that I have adequate reason to support Smith against Jones, for example if it eventuates that Jones should be disinclined to pay Smith the specified sum at the appointed time. If *I*, the onlooker, think the promise a valid one, then I will describe Jones' behavior as "defecting", or "reneging", or some such, rather than simply as a perfectly reasonable refusal to part with five dollars for the advancement of Smith. But if I do not? Or if *we* do not? We shall then, no doubt, have some explaining to do to anyone who proposes to make a promise to us, or if we should ever want to make one to anybody else. But that we *logically* might take the option of having nothing to do with "promises" is very clear. Now, Searle's claim was precisely that there is a logical entailment between those premisses and that conclusion, and the present considerations are enough to show that the entailment doesn't hold.

The further conditions required for a promise to be valid, it should be appreciate, are not of the same type as those Searle himself agrees to be necessary, such as that the words be sincerely uttered, in English, and so on. They may, on the other hand, be of the same type as whatever Searle intended by his idea of an "institutional fact" -- but that's because he supplies no clear theory of what such "facts" are supposed to be, leaving the whole idea shrouded in mystery. But a mysterious magical social somewhat isn't what we need. What we need, instead, is in fact an agreement, an understanding. But not merely the agreement Prichard noted, viz., the agreement to use certain noises or marks in certain ways. Certainly we need that, too. But the further thing clearly required that no amount of agreement about the meanings of certain noises, or whatever, can supply, is literally the agreement *to uphold* the promise in question: that is, the agreement that we, any of us, will if asked hold the party that made the promise responsible for doing what that party promised to do.

Prichard is right, then, in thinking that there is indeed an agreement to use words in a certain way. But there is more. There is a disposition to blame, or to go farther and in other ways negatively reinforce those whose behavior does not conform in the appropriate way to the use of language in question, given that the circumstances are right. Absent this disposition, we do not have the "institution" of promise-making and promise-keeping in the community in question.

5. Duty to Keep Promises as Clause in the Social Contract?

Now let us return to Scylla one last time: can it be that the obligation to keep our agreements is due to a distinct, and even an antecedent, "agreement to keep agreements"? Indeed it can. In making any particular agreement, I sugest, there must, if this is to be understood as a moral obligation in the community in question, be an understanding that the words and gestures employed are such that, in the circumstances, we (the members of this linguistic/moral community) will be inclined to visit a certain amount of negative reinforcement upon those who subsequently fail to perform certain acts (viz., those promised), and likewise where appropriate, to supply positive reinforcement for those who do -- especially to those who do in circumstances that made it difficult to do so. This understanding is an agreement, and it is simply not the same agreement as the one undertaken in the particular promise at hand.

It is, of course, as Prichard surmised, not an agreement made in further words. There is certainly no need that it be so; but further, and more fundamentally, since all use of words has the same character, there fundamentally could not be a complete explanation of practices that depend essentially on the use of communication if we stopped at what could be done with pre-existing words. What there is instead is an agreement in a different, though related, sense: the Hume/David Lewis/Gilbert Harman sense: an "agreement" in the sense of a coordinated set of conditional dispositions. It is this agreement in which consists, I submit, the "institution" mysteriously referred to by Searle in his essay[9], and rightly puzzled over by other commentators. Like morality itself -- indeed, as a principal part of morality itself -- this "institution" is informal, a fact which makes it somewhat misleading to use that term.

This coordinated disposition, on the part of any one person, is (or can be) rational and has the relevant kind of "iffiness" to be properly considered an agreement. If verbalized, it would look something like this: "I will support your claim against A, derived from A's antecedent utterance having the form 'I promise B that I will do x', if and only if you will, if the occasion arises, support my claims against whomever, derived from the latter's utterance of similar words in similar circumstances." It is precisely parallel to Hume's famous example of the two men in the boat, where the implicit agreement might be expressed, "I will pull this oar at this time if and only if you pull that one at roughly the same time". This is a reasonable thing to do only if it is to my advantage to be a member of an interacting group of persons with a coordinated mutual understanding of that type. So our reasoning may be further spelled out: " .... Which it is; thus, I support the institution."

To give this arrangement the full character of the social contract as explicated by Gauthier, we should add the clause, " . . . . provided (most) others do likewise." This is in fact nearly superfluous in one major respect in the present case, for what the institution involves is blaming those who do not do as they say they will do, and, e.g., refraining from blaming those who don't when the other parties have defaulted. However, it also involves committing oneself to do what one has promised, and for that the clause is, as Gauthier argues, essential. Were there no rationale for keeping promises, of course, there would also be no rationale for socially upholding and enforcing them. However, the rationale there is is such that those who keep their promises forego something, whereas we bystanders who merely toss praise and blame unto the appropriate bonfires have little to lose. Except, of course, that if no enforcement is forthcoming ever, then given normal humans in our society of interactees, we will soon have no institution left, and then we will forego the benefits of promising and agreeing -- which is a very great deal, indeed practically everything.

The starting point for an institution of promises is, obviously, a social situation -- if such a thing be possible -- in which no gestures or words were understood to have such significance, but in which people have, as they generally do, purposes which could be forwarded by the kind of cooperation made possible by promising. A group of prelinguistic infants more or less exemplifies the first feature of such a situation, as perhaps do many groups of inarticulate animals. Now add a couple of years of normal development in the human cases, and the second feature would also obtain. I conjecture that even without the substantial intervention of adults, such a community will ere long have an institution of what amounts to promising having precisely the features proposed above.

6. Mononomicity

J. O. Urmson criticizes contractarianism as an example of what he calls "mononomic" moral theories, that is, single-factor theories, theories which attribute all moral obligation to the exemplification of just one fundamental feature, mentioned in a single ultimate principle. According to contractarianism, Urmson thinks, this is the feature of having contracted to do or avoid the thing in question. And he asks, "Can one seriously believe that the sole moral objection to personal violence is that it is a form of breach of contract? Such a mononomic theory, based solely on contract, would be as incredible as pure utilitarianism."[10]

Now, it is misleading to characterize contractarianism as the view that what is wrong with anything that is wrong is that it involves *breach of contract* in the normal sense of those words. Plainly most acts that would be

condemned by the social contract are not acts we have a some time specifically contracted to avoid. What we should say instead is that whatever is wrong is so because it would be unreasonable, given the facts of interaction, for the members of the group in question, including the relevant parties in the case, to permit an ac of that type in those circumstances; in other words, that it would be strongly reasonable for them to reinforce a rule forbidding it. What makes these things respectively unreasonable and reasonable is that it is in their interest, given the facts about interaction with our fellows, to be disposed not to perform those acts just because we want to, and to be disposed to apply negative reinforcement of various effective kinds to those who nevertheless do.

Clearly this structure holds of a great many other acts and omissions besides making promises. Prohibitions against violence, respect for personal liberty, and so on, also issue, in any recognizable community, from these considerations. Thus the principle of (prima facie) keeping one's promises (if made in suitable conditions, etc.), is just one item among many. Yet the items in the basic set stem from what is unmistakably an agreement, in the sense articulated above, just like the others. The reason we should keep our agreements, then, is precisely that it is reasonable, as a member of the community, that we should do so and be required to do so, provided the other party(s) to the particular agreement in question do their appropriate parts.[11] And this is what was to be shown.

Jan Narveson

University of Waterloo

Revised, May 2, 1988

Notes

1. David Hume: Inquiry Concerning the Principles of Morals, Appendix III (Indianapolis: Bobbs-Merrill, Liberal Arts Press 1957, p. 122)

2. Hume: Of the Original Contract; in E. Barker, ed., Social Contract, New York: Oxford University Press 1962, p. 161; italics Hume's.

3. H. A. Prichard, "The Obligation to Keep a Promise", in Moral Obligation (Oxford, U.K.: Oxford University Press, 1968, p. 179)

4. Philosophical Review, 1964; reprinted in W. D. Hudson, ed., The Is-Ought Question (London, U.K.: Macmillan, 1969), pp. 120-134. Searle significantly modifies the account in his book Speech Acts (Cambridge University Press, 1969), Ch. 8.

5. David Gauthier's, Morals by Agreement (NY: Oxford, 1986)

6. J. Narveson, Morality and Utility (Baltimore: Johns Hopkins, 1967), pp. 190 ff. See also my "Promising, Expecting, and Utility", Canadian Journal of Philosophy, vol. 1, No. 2, December 1971, pp. 207-233; pp. 228-230 deal specifically with Searle. It will be clear, I take it, that the present account differs in essentials from the utilitarian account defended in those two sources.

7. Searle, Speech Acts, p. 189.

8. K.E.M. Baier, The Moral Point of View (Ithaca: Cornell University Press, 1958), pp. 104 ff. Here Baier talks of "presumptive reasons" but the application to obligation is clear enough.

9. Searle, "How to Derive" in Hudson (cf. note 4), p. 130 ff. There is, justifiably, scholarly debate precisely what the Searle thesis came to. His account in Speech Acts paves the way, I suggest, for the present treatment.

9. J. O. Urmson, "A Defence of Intuitionism", Proceedings of the Aristotelian Society, 1975, p. 116.

10. Much of this essay was in part originally included in my essay, "Contractarian Starting Points and 'Moral Force'", presented at the University of Western Ontario Colloquium on Contractarianism, April 24-26, 1987. However, the final version of that essay, to be published in the proceedings of that conference, (Peter Vallentyne, editor), retains only a trace of the present essay.