



A Paradox of Promising

Holly M. Smith

The Philosophical Review, Vol. 106, No. 2. (Apr., 1997), pp. 153-196.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8108%28199704%29106%3A2%3C153%3AAPOP%3E2.0.CO%3B2-K>

The Philosophical Review is currently published by Cornell University.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/sageschool.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

A Paradox of Promising

Holly M. Smith

For centuries it has been a mainstay of European and American moral thought that keeping promises—and the allied activity of upholding contracts—is one of the most important requirements of morality. On some historically powerful views the obligation to uphold promises or contracts not only regulates private relationships, but also provides the moral foundation for our duty to support and obey legitimate governments. Some theorists believe that the concept of keeping promises has gradually moved to center stage in European moral thought. They see this movement as part of an historical shift from a moral conception in which an individual's duties are mainly externally imposed and unalterable, to a conception in which duties are largely chosen by the individual.¹

In this paper I wish to examine a paradoxical aspect of the practice of promising which suggests that making and keeping promises may not always have the high moral luster usually ascribed to them. I shall argue that a system of moral rules avoids this problem only if it ascribes no moral value at all to the keeping of promises.

To see the problem, let us begin by looking at a typical example of promising.

Recruiting

Dean Allen is conducting an outside search for a new chair of the Chemistry Department. He offers the job to Marilyn Jones, who says she will only accept the position if she has the opportunity to build the department by hiring three new faculty

I am grateful for discussion and comments from Robert Frank, Alvin Goldman, Richard Healey, Davis Schmitz, the anonymous referees for the *Philosophical Review*, and from the colloquia participants at the University of Arizona, the University of Pittsburgh, and the Universidad de los Andes (Colombia).

¹See Lawrence C. Becker, "Social Contract," in *The Encyclopedia of Ethics*, ed. Lawrence C. Becker and Charlotte B. Becker, (New York: Garland Publishing, 1992), vol. 2, 1171. For a useful account of the history of how promising has been conceived in Anglo-American law and morals, see P. S. Atiyah, *Promises, Morals, and Law* (Oxford: Oxford University Press, Clarendon Press, 1981).

members. The Dean makes a commitment to Jones to provide three new lines to the Chemistry Department.

In ruminating about examples such as *Recruiting*, one is forcibly struck by the fact that promising, as a moral institution, gives agents the power to control the moral status of their future actions. By making a promise to do A an agent makes A obligatory for him, and thus typically *changes* its moral status. For example, we can readily understand *Recruiting* as a case in which the promised act, providing three new lines to the Chemistry Department, would have been wrong were it not for the promise. Suppose that there is far greater need for new faculty in English and History than in Chemistry, and hence that in circumstances in which no promise had been made, the dean would have been obliged to provide the lines to these two departments rather than to Chemistry. However, the dean promises these faculty lines to Jones to induce her to accept the chairmanship of Chemistry. Once Jones arrives, it appears that the dean is obliged to carry out his promise. Thus the dean's promise has converted an act that would otherwise have been wrong to one that is obligatory given his promise.

The fact that promising to perform it can make an otherwise wrong act obligatory has been recognized by a few authorities. Sidgwick, for example, notes that "it is clear that a promise may sometimes make it obligatory to abstain from doing what it would otherwise have been a duty to do. Thus it becomes my duty not to give money to a meritorious hospital if I have promised all I can spare to an undeserving friend; though apart from the promise it might have been my duty to prefer the hospital to the friend."² Joseph Raz, similarly, argues that promises establish obligations, and then states that "obligatory acts are . . . acts which the agent must sometimes perform even if they should not be performed on the balance of reasons."³

Unfortunately, the fact that making a promise can alter the moral status of one's future action apparently enables agents to exploit

²Henry Sidgwick, *The Methods of Ethics*, 7th ed. (Chicago: University of Chicago Press, 1907), 305.

³Joseph Raz, "Promises and Obligations," in *Law, Morality, and Society: Essays in Honour of H. L. A. Hart*, ed. P. M. S. Hacker and J. Raz (Oxford: Oxford University Press, Clarendon Press, 1977), 224.

the practice of promising in an abusive manner. Consider the following case:

Nepotism

Dean Baker has funds that can be allocated to one of her departments to hire an adjunct faculty member. As the dean knows, if the funds are allocated to the Classics Department, it will conduct a fair search and hire Dean Baker's cousin, the best available candidate. However, since Classics has far less need for an adjunct faculty member than do other departments, it would be wrong (other things being equal) for Dean Baker to give the funds to Classics. Dean Baker would like her cousin to have employment, so she promises to allocate the funds to Classics, and subsequently carries out this commitment.

In this case Dean Baker wants to allocate the funds to Classics so that it will hire her cousin. Such an allocation would in itself be wrong, but Baker makes a promise to the Chair of Classics to provide these funds. She thus seems to place herself under an obligation to allocate the funds in this manner, and so seems to alter the moral status of her future act. Let us assume (we will examine this more carefully later) that Dean Baker's promise in fact converts her act of allocating the funds to Classics into an act that she is now morally required to perform. *Nepotism* then provides an example of the way in which making a promise can be a very powerful device for manipulating morality in an abusive manner.

Of course, not every promise transmutes the promised act from a wrong act to an obligatory one. Some promises convert an otherwise morally neutral act to an obligatory one, while others merely render more obligatory an act that was already morally required. But in at least some cases it appears that a promise converts an otherwise wrong act into an obligatory one.

The fact that our moral system permits agents to dictate in this manner the moral status of their future actions seems an astonishing power to build into a moral system. It is especially troubling when one notes that agents apparently can exploit promises in order to legitimize otherwise objectionable courses of action. What would we say, for example, about a moral system in which an agent may render A obligatory by simply declaring, "My doing A next week will be, by virtue of this declaration, morally obligatory"? We

would find such a system altogether too subject to abuse, since it would enable us to define for ourselves, in whatever way we found personally convenient, which acts are right and wrong. Yet our own system seems to have something like this very feature built right in, since to promise to do A is something like to state that one by this declaration intends to undertake an obligation to do A, and thereby to place oneself under an actual obligation to do A.⁴ As Sidgwick notes, if such chosen obligations are available without limitation, one could evade any other moral obligation simply by promising not to fulfill it, an outcome he finds absurd.⁵ And yet this, or something close to it, seems to be the very implication of the practice of promising as we understand it.

1. Limitations on the Power of Promising

At this point it is natural to protest, "But there *are* limitations built into our practice of promising that save it from generating such absurdities. Matters are not nearly so bad as you have suggested." Of course it is true that we recognize such limitations. Let us survey the most important ones, and see whether they protect promising from the difficulty I have described.

(1) *Promises merely establish obligations.* First, the fact that someone has promised to do A does not in itself establish that A ought to be done, all things considered, or even that A is right, all things considered. At most it establishes that A is obligatory—and obligations at best generate a *prima facie* claim that the obligatory act

⁴See, for example, Richard Bronaugh, "Promises," in Becker and Becker, *Encyclopedia of Ethics*, vol. 2, 1021. However, as Toni Vogel Carey has argued, it is far from clear that a person's *undertaking* an obligation is the same as, or invariably results in, his *coming under* an obligation. She points out, in "How to Confuse Commitment with Obligation," *Journal of Philosophy* 72 (1975): 276–84, that there is a clear intensional sense of "undertaking an obligation" (equivalent to something like "coming to believe that one has an obligation") that does not support this inference, and that promise making is best interpreted as inevitably generating the "undertaking of obligations" only in this intensional sense. If this view is correct (as I am inclined to think it is), then the "analysis" given in the text is inaccurate.

⁵*The Methods of Ethics*, 305. See also H. L. A. Hart's claim that "it would be mysterious if we could make actions morally good or bad by voluntary choice" ("Are There Any Natural Rights," *Philosophical Review* 64 (1955): 175, cited in Atiyah, *Promises, Morals, and Law*, 179).

ought to be done. This prima facie claim can readily be overridden when there are countervailing moral considerations. For example, the dean might promise the romance languages departments to refurbish their aging language laboratory. This promise establishes a prima facie claim that he ought to do so. But if strong countervailing considerations arise—for example, unanticipated budget cuts mean that the dean can only refurbish the laboratory at the cost of laying off six secretaries—the obligation to refurbish the language laboratory would be overridden, and there would be no all-things-considered requirement to keep his promise. Some promises would be wrong to keep.

This is an important limitation on the power of agents to dictate the moral status of their own future acts to their own advantage: it means that no promise to perform it can convert a really heinous act into a morally right one, since the obligation established by the promise would always be outweighed by the countervailing evil of the promised act. For this reason a promise to commit murder or theft cannot create a moral requirement to carry out the promise.⁶ But the crucial point still remains: creating the obligation to keep one's promise adds significant new moral weight to the calculus of what it would be required, all things considered, to do. Often this weight will not be counterbalanced by countervailing moral considerations. Thus it still remains possible that the agent's promise converts a future action into one he is required to do, all things considered, even though the act would have been wrong in the absence of such a promise.

(2) *It is wrong to make certain promises.* The second limitation on the agent's moral power to dictate his own moral future arises from the fact that, as most writers have acknowledged, there are some promises that it is wrong to make. It may be wrong, for example, for Dean Baker to promise that she will allocate funds to Classics to hire an adjunct faculty member. If so, then she ought not to make the promise, and hence she cannot with full legitimacy deploy morality to support a morally pernicious line of action. This

⁶In the law of contracts, a contract will not be enforced if its enforcement is contrary to public policy (*The Restatement (Second) of Contracts*, section 178). For a nuanced discussion of related matters, see Richard Kraut, *Socrates and the State* (Princeton: Princeton University Press, 1984), *passim*, but especially chapter 2.

limitation certainly promises to eliminate some of the abuses of morality that might arise without it.

Notice that this limitation addresses only the *permissibility* of an agent's making a promise. It does not address the question of what moral status the promised act would have if the agent nonetheless made the impermissible promise. If, in such a case, the agent ought to fulfill his wrongful promise, then he still has available a method of manipulating morality in an abusive manner to serve his own ends. However, his use of morality for abusive ends is not *fully* legitimate: he goes wrong on one count (making the impermissible promise), even though he goes right on another (carrying out the promise once made).⁷

But (leaving this complication aside) does this limitation eliminate all the abusive cases with which we are concerned? Discussions of the fact that some promises should not be made commonly cite the lying or insincere promise—a promise the agent plans from the beginning not to carry out—as one it would be wrong to make. Clearly these kinds of cases are irrelevant to our concerns, since insincere promises are wrong because of the unlikelihood they will be fulfilled, not because carrying them out would in itself be morally objectionable.

More directly relevant are cases in which it would be wrong for an agent to promise to do what, on independent grounds, would be wrong for him or her to do. The kind of case standardly described here is a case in which the agent promises some serious transgression, such as committing a murder or robbing a bank.

What is the basis for commonplace statements that an agent should not promise to commit a murder or rob a bank? Someone holding this opinion might subscribe to the following view:

Prohibition Principle: If an act would be wrong, all things considered (independent of any promise to perform it), then it would be wrong, all things considered, to promise to perform that act.

This principle would certainly reduce the abuses we have in mind, since it prohibits individuals from imposing obligations on themselves to do what is independently morally objectionable.

What can be said for or against the *Prohibition Principle*? In con-

⁷For more discussion, see note 31.

sidering this issue, we do not want to produce apparent counterexamples to it by citing cases in which the agent, at the time of promising to do A, had no way of knowing that A would in fact be wrong. Agents typically make promises when they have inadequate knowledge about what circumstances will obtain at the time the promise is to be carried out. An agent who, with good intentions, promises to perform what turns out to be a wrongful act certainly does not seem to be someone who has made a wrongful promise. But this can easily be accommodated by proponents of the *Prohibition Principle*, who can readily say that this agent did what was *objectively* wrong in making the promise, but that the agent's act was *subjectively* right (given his knowledge at the time), so that he is not to be blamed or criticized for his promise. Thus the *Prohibition Principle* does not seem to run afoul of such cases.

But even confining ourselves to cases in which the agent has complete knowledge of the future, only brief reflection is necessary to show there are cases in which the *Prohibition Principle* is mistaken: cases in which it is permissible, or even required, for an agent to make a promise even though the act that carries out that promise would be wrong in the absence of the promise. Some of these cases are ones in which part of the reason the promise is permissible is precisely the fact that the agent will not carry it out. For example, an undercover police agent might promise a Mafia boss to assassinate his rival for suitable payment. It would be wrong for the police agent to kill the rival, but he never needs to, since his promise elicits the payment, which serves as evidence to send the Mafia boss to jail for attempted murder. Here making the promise is morally permissible, because the promise itself has good consequences that are independent of the actual performance of the promised act, which never occurs.

But I would argue, with Sidgwick, that a promise can be morally permissible even in cases where the promised act occurs and would be wrong were it not for the promise. I believe that *Recruiting* is a case of this type. We saw that it would be wrong (apart from the promise) for Dean Baker to allocate the faculty lines to the Chemistry Department, since there is far greater need for new faculty in English and History than in Chemistry. However, the dean promises these faculty lines to candidate Jones to induce her to accept the chairmanship of Chemistry, and the consequences of this are very positive: Jones accepts the position and provides years of

strong and fair leadership to the department; her appointment notably increases the stature of the university; and in the ensuing decade she acquires a series of multimillion-dollar grants whose overhead returns significantly benefit the university. The promise of these faculty lines to Jones obtains benefits that jointly outweigh the presumption against the dean's provision of the three faculty lines to Chemistry rather than other departments. The consequences of making the promise are *additional to* those of carrying it out, and this broader set of consequences makes it morally acceptable to make the promise, even though carrying it out would, apart from the promise, be wrong.⁸

These cases show us that the *Prohibition Principle* is mistaken: it is sometimes morally permissible to make a promise even though carrying out that promise would, apart from the promise itself, be wrong.

The *Prohibition Principle* is not the only way to defend the view that some promises are wrong to make. We will examine a more plausible rationale for this idea in section 3, but we will find that even this defense does not bar the problem of abusive promises.⁹

⁸There are two different ways to interpret the claim that (once Jones has been hired) giving the three lines to Chemistry would be wrong, apart from the promise. One interpretation focuses on whether this would be the right act if the promise had never been made. The other focuses on the case in which the promise has been made, and asks whether giving the lines to Chemistry is the right act, leaving aside the moral value it may have because it counts as keeping a promise (and the moral disvalue other acts may have because they would involve breaking a promise). This latter interpretation is the relevant one. To get a clean case here we should imagine that even though Jones's hiring benefits the university, it does not benefit English and History, or compensate them indirectly (for example, through access to increased overhead monies) for loss of the lines that otherwise would have gone to them.

⁹In addition to the view examined later in the text (the view that promises must always be assessed as part of a sequence of actions), there may be alternative rationales for the view that it would be wrong to make certain promises. The fact that most philosophers, in arguing that it would be wrong to make certain promises, cite cases involving extreme transgressions such as murder or theft suggests that they may actually hold a weaker version of the *Prohibition Principle*—perhaps a principle stating that it would be wrong to promise to perform an action that is in itself strongly *prima facie* wrong. Murder or theft are clear examples of types of acts that are strongly *prima facie* wrong. This weaker principle has greater plausibility than the *Prohibition Principle*. However, even it seems suspect: surely there are cases in which it is best, all things considered, to make a promise to do what in itself is *prima facie* very wrong (the case of the undercover

(3) *Promises must potentially benefit the promisee.* The third limitation on our ability to make promises in an abusive manner arises from the fact that a promise is not a solitary act: a promise is always a promise *to someone*. Some authors have held that, for this reason, any genuine promise must be to perform some act that will benefit the promisee or at least an act the promisee wants the promisor to perform.¹⁰ Sometimes this is phrased as a condition that there is no obligation to carry out a promise that would injure the promisee. And all authorities agree that the promisee may always release the promisor from the promise, so that if it begins to appear that carrying out the promise would harm the promisee, the promisee can avoid injury by releasing the promisor from his obligation. If these restrictions are correct, I could not successfully legitimize immoral actions by promising you that (say) I will steal your cash, because (a) I cannot genuinely promise to do what will injure you (or at least such a promise would not create any obligation to carry it out), and (b) even if I could so promise, you could—and doubtless would—release me from the promise, thus eliminating any supposed obligation on my part to commit the theft.¹¹

It does not appear, however, that the first of these alleged restrictions on the power of promising in fact obtains. Raz, for example, argues persuasively that a genuinely binding promise need

police agent is one example; cases in which the strong prima facie wrongness of the promised act is counterbalanced by some even more powerful prima facie positive consideration in its favor would be other examples). Such a weaker principle, if true, would certainly limit the degree to which agents could bend morality to their own abusive ends by deft use of promising. But even if true it would not rule out the possibility of cases in which it is morally acceptable (or even required) for an agent to promise to perform an action that is wrong, all things considered, but wrong for a complex variety of wrong-making features balanced against countervailing right-making features, where none of the wrong-making features is itself terribly weighty on its face. Such cases still raise the specter of being ones in which the agent might use promising abusively.

¹⁰Alternatively, the promisor must believe this at the time he makes the promise. See John Searle, *Speech Acts* (Cambridge: Cambridge University Press, 1969), chap. 3; Thomas Scanlon, "Promises and Practices," *Philosophy and Public Affairs* 19 (1990): 208; G. J. Warnock, *The Object of Morality* (London: Methuen, 1971), 98; Sidgwick, *The Methods of Ethics*, 306–8; and Cicero, *De Officiis*, bk. 1.10.

¹¹Of course in some cases you might not be able to release me—perhaps you are unaware of the possible bad consequences to yourself, or perhaps you are comatose and incapable of releasing me.

not benefit the promisee, or even be believed by the promisor to do so. He describes a plausible case in which a man solicits a promise, hoping and believing that it will be broken, in order to prove to a certain lady how unreliable the promisor is; the promisor is aware of all this and makes his promise intending to disappoint the promisee by keeping it. In Raz's case it is plausible to say that the promisor has an obligation to carry out his promise even though doing so will injure the promisee.¹²

But even if the alleged restriction did obtain, it still would not prevent promisors from using promising abusively. One can always promise one person to do something that will benefit the promisee but still be wrong (apart from the promise) because of the manner in which *other* individuals' welfare might be affected. Sometimes this is direct (for example, I promise my partner in crime to steal your money); in other cases it is more indirect (Dean Baker promises to allocate adjunct funds to Classics; this affects the welfare of students and faculty in other departments experiencing greater student demand). In cases such as this, the second restriction—the fact that the promisee may release the promisor from the commitment—may offer no help either, since the promisee may not care about, or may even desire, the injury to the third parties. Hence even if promises that would harm the promisee were eliminated, this would not eliminate abusive promises that negatively affect the welfare of third parties.¹³

We have now examined three limitations, or purported limitations, on the moral power of promising: the fact that a promise merely establishes an overridable *prima facie* moral requirement that the promised act be performed; the fact that some promises

¹²Raz, "Promises and Obligations," 213–14. David Schmidtz points out that some threats have the same binding character as promises, and appear to establish, in the same way, an obligation to perform the act in question, even though it injures the person threatened. In his example, a coach might threaten to cut a player who skips one more practice. This threat in itself is reason (albeit defeasible) for cutting the player when he skips another practice. The difference between calling this a threat and a promise seems merely to be the fact that the action in question is injurious; the moral quality of the situation does not otherwise seem to differ from that of a promise.

¹³Atiyah notes that during the period of classical contract law it was widely believed (but, as he states, wrongly) that enforcement of contracts and promises was a method of ensuring that there is always harmony between public and private interests. See *Promises, Morals, and the Law*, 77.

are wrong to make; and the alleged fact that a genuine promise (or one that ought to be carried out) must benefit the promisee. We had hoped that these limitations would restrict the practice of promising in such a way that it could not be used by an agent to control the moral status of his own future acts for pernicious ends. Unfortunately, we have found that these limitations do not provide a magic shield against the potential abusive use of promising: they may limit the occasions on which abuses could be carried out, but we found no reason to conclude they form an absolute bar. If abusive use of the practice of promising is to be precluded, we must find the solution in another quarter.

2. Further Considerations

Before searching for this solution, it is worth considering whether promising is unique in harboring potential for this particular form of abuse.

It is easy to think of making a promise as the only way in which an agent can control his or her own moral future. Yet, as a number of authors have noted, making a promise is not unique in this regard.¹⁴ There are many acts an agent can perform that change the moral status of his future acts: he can waive his right to something (thereby making it wrong for him to reserve that thing to his exclusive use); he can injure another person (thereby making it incumbent on him to make reparations); he can release another from a promise (thereby making it wrong for him to insist that the promise be kept); he can beget or adopt a child (thereby incurring parental obligations to the child); he can simply place himself in a position where he physically cannot perform an act that would have been obligatory were he able to do it (thereby relieving himself of such an obligation). We act in a variety of ways that determine our moral future all the time, and I have found no principled way of drawing a satisfactory distinction between promising and these other types of acts.

In these other cases can one's earlier act make obligatory or right an act that would have been wrong apart from the earlier

¹⁴See, for example, the discussion in Annette Baier's "Promises, Promises, Promises," in *Postures of the Mind: Essays on Mind and Morals* (Minneapolis: University of Minnesota Press, 1985), 174–75.

act? Apparently so: for example, I might know that my brother, with whom I am angry, is going to ask to borrow my car; to ward off the necessity of lending it to him, I might loan it instead to a neighbor, thereby waiving my right to it during that period. If I had not already lent my car to the neighbor, it would have been obligatory for me to lend it to my brother instead, and wrong for me to have refused. In another example, a man might father a child, thereby incurring parental duties to the child, and enabling himself to avoid military service. If the new father had not begotten a child, it would have been obligatory for him to join the military on being drafted, and wrong for him to have evaded the draft. These agents' acts render certain of their subsequent acts permissible when they would otherwise have been morally wrong. And in either of these cases, depending on the circumstances, one might judge that the agent has wrongfully exploited the moral power available to him.

It appears, then, that the type of abuse of a moral power with which we are concerned is one which may be most readily seen in cases of promising, but which infects a much broader range of an agent's activities that affect his own moral future.¹⁵

3. Solving the Problem: Phase One

Many authors have praised the value of promising as an institution: it permits individuals to tie down the future and to engage in mutually beneficial cooperative arrangements that would otherwise not be possible.^{16,17} But we have also seen that the practice of prom-

¹⁵Of course there are also many cases in which an act that I perform determines the moral future of *your* act, and may make a future act of yours morally permissible that would otherwise have been wrong. In certain cases I might use this power abusively. Since two agents are involved, however, the potential for abuse in these cases is less than it is in cases where the agent acts to control his own moral future. Hence I shall not directly attend to these two-person cases.

¹⁶See, for example, David Hume, *A Treatise of Human Nature*, ed. Selby-Bigge (Oxford: Oxford University Press, Clarendon Press, 1888), bk. 3, pt. 2, sec. 5; John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971), 344–50; and John Rawls, "Two Concepts of Rules," in *Mill: Utilitarianism*, ed. Samuel Gorovitz (Indianapolis: Bobbs-Merrill, 1971), 183.

¹⁷Hume (*A Treatise of Human Nature*, 522–23) claims that society has an interest in developing and maintaining the institution of promise keeping, and hence that individuals within society have an interest in "punishing"

ising paradoxically carries the seeds of corruption: it appears that individuals may abusively use the moral power of promising to legitimize pernicious acts that would have been wrong were it not for a promise to perform them. One could view this problem as sufficiently serious to call for ejecting promising from our arsenal of moral powers.¹⁸ I shall instead describe ways in which we can interpret the practice of promising that enable us to avoid the problem. Ordinary thinking about promising, like ordinary thinking about many normative practices, is not highly precise. Carefully characterizing and then finding a solution to the problem I have described requires us to impose greater precision on our understanding of promising than already exists in ordinary thought. Because we will be moving from an imprecise to a more precise understanding of this issue, it may be unclear to what extent the proposed solution involves recommending a change in our common understanding of promising, and to what extent it involves making more explicit aspects of that understanding that already exist at a merely implicit level.

In section 1 we examined three limitations on the moral power of promising and found they could not prevent abuses of this pow-

promise breakers by withholding their trust from such promise breakers in the future. But apparently this only means that society has an interest in pressuring promise makers to carry out their promises when they can. Judging by our practice, it does not mean that individuals within society have an interest in *making it possible* for promise makers to carry out those promises.

As we understand promising, the fact that I have made a promise places you under no moral call whatsoever to assist me in keeping that promise. If I have promised my daughter a certain expensive doll for Christmas, and then find myself unable to afford the doll, there is no moral call on you (absent special circumstances) to lend or give me the money to make it possible for me to carry out my promise. Nor does my promise even create any moral call on you not to prevent me from keeping the promise. In a case where I have the money, but there is only one doll of the promised type left in the store, and you want to purchase one as a surprise for your niece, if you are first in line you are under no moral constraint to step aside and let me purchase it instead, so that I may keep my promise.

¹⁸This strategy has been advocated by Richard M. Fox and Joseph P. DeMarco, who raise questions about promising that are related, although not identical, to the ones described here, and conclude that “promising seems to be a morally unjustified practice.” See “Putting Pressure on Promises,” *Southern Journal of Philosophy* 30 (1992): 45–58 (the quotation is from page 57), and “The Immorality of Promising,” *Journal of Value Inquiry* 27 (1993): 81–84.

er. Part of the source of this failure is the fact that these limitations apply to the act of promise making, or to the act of promise keeping, *in isolation from each other*. The first step toward solving our problem comes from realizing that these acts should not be considered separately from each other, but rather evaluated as linked acts forming part of a sequence. Whether one should make a promise, or keep it, depends on the moral character of the sequence in which it is embedded, not simply on the nature of the act considered apart from what precedes or follows it. The “unit” we should initially appraise is not the individual act: we must instead enlarge our view, and first appraise the sequence of making the promise and then keeping it, or the sequence of making the promise and then breaking it, and so forth. Only in this manner can we adequately take into account the interplay of these actions with each other. Thus we should adopt the following principle when assessing an instance of promising:

Principle I: In a case where the agent has the sequence options as of t_1 of (a) making a promise at t_1 to do A at t_2 , and then doing A at t_2 , (b) making a promise at t_1 to do A at t_2 , and then not doing A at t_2 , (c) not making a promise at t_1 to do A at t_2 , but then doing A at t_2 , and (d) not making a promise at t_1 to do A at t_2 , and then not doing A at t_2 , the agent ought as of t_1 to perform the best sequence option available to him, and ought as of t_1 to perform each of the acts that is a member of that sequence.

Reflection on the nature of promising may make this principle seem a natural one, because of the overt tie between the acts of making and keeping a promise. However, the idea that acts must be assessed as parts of sequences is a position that I, and several other theorists, have argued for more generally with respect to all actions.¹⁹

¹⁹These are special limited cases of more general principles I have argued for in previous work. There I argued that the proper principle for determining what an agent should do is the following:

S ought at t_1 to perform maximal sequence X starting at t_1 if and only if S has the ability at t_1 to perform X, and X is better than any alternative maximal sequence starting from t_1 which S also has the ability at t_1 to perform.

This more general principle is the correct one, but for ease of discussion we can use the more limited principle in the text without distortion. For statement and discussion of the more general principle, see my “Doing

Principle I states that an agent ought to perform the best sequence option available to him. There are various ways to determine which sequence option is "best." In the discussion that follows I shall assume that the "moral value" of each sequence option can be determined by assigning an individual "moral value" to each action that is a member of the sequence, and then summing these values to arrive at an overall value for the entire sequence. In the typical case the moral value of an individual act will derive partly from consequentialist considerations (that is, the action's consequences), and partly from deontological *per se* considerations (such as the fact that the action would effect a fair distribution, or would involve keeping a promise).²⁰

Carrying out this idea involves assigning numerical values (where the values are measured on at least a ratio scale) to individual acts and to the sequences of which they are members. The assumption that we can do this is of course highly problematic. There are precedents for utilizing such values in deontological contexts. Robert Nozick has argued that numerical values (measurable on an interval scale) must be possible in the case of deontologies in order to balance countervailing considerations against each other.²¹ Moreover, it is relatively easy to imagine mounting a Von Neumann-Morgenstern-type argument for the existence of numerically representable deontological values (again, measurable on an interval scale), since deontological theories, like consequentialist ones, must deal with decisions to be made under risk. Since it is extremely helpful in analyzing our problem to assume that numerical moral values can be assigned to individual acts and to act sequences, I will temporarily assume that this procedure is coherent. However, this assumption is not essential to my argument, and

the Best One Can," in *Values and Morals*, ed. Alvin I. Goldman and Jaegwon Kim (Dordrecht: D. Reidel, 1978), 185–214, and also the earlier "Dated Rightness and Moral Imperfection," *Philosophical Review* 85 (1976): 449–87. (Both of these papers were published under the name 'Holly S. Goldman'.) For a view different from, but importantly related to, that defended in "Doing the Best One Can," see Fred Feldman, *Doing the Best We Can* (Dordrecht: D. Reidel, 1986).

²⁰For a more detailed discussion of the issue of establishing overall moral value for a sequence of actions, see my "Dated Rightness and Moral Imperfection."

²¹Robert Nozick, "Moral Complications and Moral Structures," *Natural Law Forum* 13 (1968).

we may view it simply as a provisional assumption that can be discarded once the analysis has been completed and the solution found. In what follows I will not, of course, try to specify what values determine which sequence of acts ought to be performed. This is determined by the operative normative theory, and different theories will have different implications for how strict the moral constraints are on an agent's choice of sequences.²²

We should note that my discussion assumes that the time at which an agent has an obligation to perform an act may be different from the time at which the act is to be performed (I express the time of the obligation by using the locution 'The agent is obligated *as of* t1 . . .'). As time moves on the agent's choices change, and an act that was obligatory as of an earlier time may no longer be obligatory—indeed may even be wrong—as of a later time. For example, as of 7:00 a.m. the chairman of the History Department ought to meet a graduate student at 10:00 to discuss her dissertation as he has promised to do. However, the chair realizes at 9:30 that he must submit his budget request to the dean at noon, and that he has left the critical papers at home. As of 10:00, then, the chair ought not to be meeting with the graduate student, but ought instead to be driving home to retrieve the budget papers, since submission of the budget request must take precedence over meeting with the graduate student. Hence what an agent ought to do depends on the time of the obligation: as of 7:00 the agent ought to meet with the student at 10:00, but by 9:30 he ought not to, given his intervening failure to bring the budget papers to the office, and his obligation to retrieve the papers in order to submit the budget request by noon.²³

Morally assessing actions as members of sequences in the manner required by Principle I enables us to see clearly why certain instances of promising that intuitively seem morally acceptable are indeed per-

²²The assumptions that (1) the value of the sequence is obtained by summing the values of its component acts, and (2) a prescription for which sequence to perform is derived by selecting the sequence with the highest value, are natural ideas, and hence ones that I shall follow in the text. However, they too are controversial, and might be rejected even by someone who otherwise agrees that acts and sequences can be ascribed numerical values that determine the moral status of these entities.

²³For further discussion of "tensed obligations" see my "Dated Rightness and Moral Imperfection."

missible—even though the agent’s promise renders the promised act morally required when, apart from the promise, it would have been wrong. To see this, consider again the *Recruiting* case. In this case, the dean promises Jones three new lines for the Chemistry Department in order to induce Jones to accept the chairmanship of the department. While it would be bad in itself for the dean to give the three faculty lines to Chemistry rather than to English or History, it is also true that the dean’s making the promise to do so induces Jones to accept the chairmanship of Chemistry, and her accepting the chair produces the substantial benefits described above. Let us assume that the benefits of these events outweigh the losses to English and History of the three lines. The dean has four alternative sequence options: (1) he can promise the faculty lines to Jones and then carry out his promise once she becomes Chair; (2) he can promise the faculty lines to Jones but then renege on his promise once she becomes Chair; (3) he can decide not to make the promise to Jones, hire a much inferior candidate for Chair, and then give Chemistry the three lines, or (4) he can make no promise to Jones, hire the much inferior candidate for Chair, and then give the lines to English and History. Without troubling at this point to assign numerical values to these sequences, we can see that the first sequence, in which the dean makes the promise to Jones and then carries it out, could be morally much better than any of the other sequences. According to Principle I, then, the dean ought to make the promise to Jones and subsequently carry it out—even though giving the lines to Chemistry would be wrong were it not for the promise. Moreover, once the dean has made the promise to Jones, it will be best for him to carry it out, since the damage done by renegeing on it would far outweigh the benefits gained for English and History. This seems to me exactly the right result: just as Sidgwick claimed, sometimes a promise should be made and carried out (even though doing the promised act out would not have been good if the promise had not been made) because the benefits to be secured by making the promise are sufficiently valuable that they outweigh the ill effects that will be produced by carrying it out. But to make this judgment, one needs to examine making the promise and keeping the promise as aspects of a single sequence, not just as two acts in isolation from each other.

Principle I subjects the making of promises to a stringent *moral* test. Many authors²⁴ have pointed out that individuals frequently

²⁴Raz states, “Voluntary obligations [of which those generated by prom-

make promises as a technique for enhancing their personal welfare—for example, in order to ensure his own election, a presidential candidate may promise his rival an appointment to a cabinet position if the rival bows out of the primary race. Indeed, a promise can often be seen as simply a private deal between two parties to procure their individual ends. But why should the full force of morality support the enforcement of such private deals? We now have an answer to that question. Agents are not simply free to make promises in order to promote their own private projects, but must consider the overall moral impact these promises will have. Principle I implies that some promises would be wrong to make. It also implies, despite some philosophers' statements to the contrary, that making some promises can be morally obligatory.²⁵ Adopting Principle I is an important first step in dealing with cases, unlike *Recruiting*, in which it would be an abuse of morality to make a promise and carry it out.

4. The Problem Deepens

However, accepting Principle I does not yet give us a way of prohibiting every genuinely abusive use of promising. Indeed, we are now in a position to see that the problem is worse than we have thought so far. Hitherto we have described the problem as consisting in the fact that in certain cases morality seems to imply that it is permissible for an agent to make and carry out a promise to perform an act that would be wrong were it not for the promise, even though we feel intuitively that such a line of action is pernicious. We are now in a

ises are an example] are the one exception to the rule that rules facilitating realization of the agent's goals do not impose obligations" ("Promises and Obligations," 224); Hume's analysis of the morality of promise keeping famously describes the practice of promising as arising in order to facilitate individuals' pursuit of their own self-interest (*A Treatise of Human Nature*, bk. 3, pt. 2, sec. 5); Prichard notes that the promotion of our own welfare frequently requires the cooperation of someone else, and that promises are devices for securing such cooperation (H. A. Prichard, "The Obligation to Keep a Promise," in *Moral Obligation and Duty and Interest* (Oxford: Oxford University Press, 1968), 174–75); and John Rawls states, "Promising is an act done with the public intention of deliberately incurring an obligation the existence of which in the circumstances will further one's ends" (*A Theory of Justice*, 347).

²⁵See John Rawls, *A Theory of Justice*, 346, and Jan Narveson, "The Desert Island Problem," in Gorovitz, *Mill: Utilitarianism*, 280.

position to see that morality not only permits agents to make such promises, but can actually require them to do so! To see this we need to examine such a case in more detail. The *Nepotism* case, described on page 155, will serve this purpose.

In what I shall call a “standard uncomplicated case of promising,” of which *Nepotism* is an example, the agent initially has available four alternative sequences of actions.

- (1) promising at t1 to do A at t2, and then doing A at t2;
- (2) promising at t1 to do A at t2, and then doing not-A at t2;
- (3) not promising at t1 to do A at t2, and then doing A at t2;
- (4) not promising at t1 to do A at t2, and then doing not-A at t2.²⁶

In *Nepotism*, A is allocating the funds to Classics, and not-A is allocating them elsewhere.

At t1 the agent performs the first action in one of her four available sequences (either promising to do A, or not promising to do A). Following the performance of this act, at subsequent time t2, the agent must choose between the following acts to be performed at t2 (in the standard case, the acts available at t2 do not change depending on the agent’s act at t1):

- (5) doing A at t2;
- (6) doing not-A at t2.

Each of the acts that makes up one of the four initial sequence options has its own moral value (this typically arises out of a combination of its consequences and its deontological features). In what I shall call a “simple problematic case of promising,” neither the act of promising to do A nor the act of not promising to do A has any moral value. We also need a notion of an act’s “underlying value”: this is the moral value an act possesses by virtue of its morally relevant consequences and deontological features, *excluding* the pure deontological fact that it would involve keeping or breaking a promise. In a simple problematic case of promising, the “un-

²⁶Of course, in most realistic cases, it will be possible to describe the agent as having a finer-grained set of options: for example, he can promise to do A, or promise to do B, or promise to do C, and so forth. In addition, the agent could have very different alternatives at t2, depending on whether the promise was made at t1. But for our immediate purposes in the text, we can concentrate simply on the core case depicted there.

derlying” moral value of not-A is greater than the underlying moral value of A, and neither of these values depends on whether the agent promises to do A. Beyond its underlying value, an act can have further moral value by virtue of being a case of keeping a promise or breaking a promise (these values are ones that, in a typical case, play a strong role in making it true that the agent ought to keep the promise once made). Thus, if the agent has promised to do A, the total moral value of A consists of its underlying value together with the value it derives from being a case of keeping a promise. *Nepotism* is a simple problematic case of promising.

A reasonable assignment of value to the acts involved in *Nepotism* is as follows: The value ascribed to A (allocating the funds to Classics) incorporates the positive underlying value of making courses available to students and securing employment for the cousin, but the negative value of engaging in a nepotistic act. The value ascribed to not-A (allocating the funds elsewhere) incorporates the positive values of making courses available to even more students and securing employment for some other individual.

Nepotism I

A = allocating funds to Classics; not-A = allocating funds elsewhere; P(A) = promising to do A; not-P(A) = not promising to do A; K(P) = keeping a promise; B(P) = breaking a promise.

Moral values	Underlying values
P(A) = 0	A = 2
not-P(A) = 0	not-A = 8
K(P) = 10	
B(P) = -10	

Options as of t1

1. P(A) & A
2. P(A) & not-A
3. not-P(A) & A
4. not-P(A) & not-A

Options as of t2	
If P(A) at t1	If not-P(A) at t1
1. A	1. A
2. not-A	2. not-A

Should Baker make the promise to do A? The way to answer this question, given what I have said so far, is to calculate the overall moral values of the four sequences available to Baker at t1, and then prescribe to Baker the sequence having the highest total value. Carrying out this procedure gives us the following values for Baker's choices as of t1:

1. P(A) [0] + A [2 + 10] = 12*
2. P(A) [0] + not-A [8 - 10] = -2
3. not-P(A) [0] + A [2] = 2
4. not-P(A) [0] + not-A [8] = 8

Of these four, sequence 1 (promising to allocate the funds to Classics, and then so allocating them) has the highest value, and is the one Baker ought to select (expressed by the *). As of t1, then, Baker ought to make the promise at t1 and then carry it out at t2.

We can then ask what Baker ought to do as of t2 if she does what she ought to at t1, and find the answer in the same fashion, by examining the values of the sequences then open to her:

t2 (after doing P(A) at t1)	
1. A [2 + 10]	= 12*
2. not-A [8 - 10]	= -2

As of t2, Baker ought to allocate the funds to Classics as she promised.

This analysis, however, gives us highly objectionable answers in telling us that Baker ought at t1 to make the promise to give the adjunct line to Classics and then, having made this promise, ought at t2 to carry it out. It seems intuitively clear that, from the moral point of view, as of t1 Baker ought not to make a promise to allocate the line to Classics, and then ought to allocate it elsewhere. It is the availability of the option of making a promise that leads to the distorted prescriptions in this case. If for some reason Baker

lacked the option of making the promise at t_1 , then morality would dictate that she ought to allocate the funds elsewhere, since the values of her choices at t_2 would be as follows:

1. A [2] = 2
2. not-A [8] = 8*

But Baker *does* have the option of making a promise, the promise is morally required, and once made must be carried out. When philosophers have discussed the problem of making immoral promise—for example, a promise to commit a murder—they have concerned themselves with the possibility that a promise, once made, might convert an otherwise wrongful act (the murder) into one that is morally required. What they have not noticed is that, on the most natural analysis, the problem created by the moral power of promising is actually much worse, since it can result in situations such as this one in which the agent is morally *required* to make such a counter-moral promise.

Since this problem arises because the option of making a promise can illegitimately create conditions in which making that very promise, and then fulfilling it, are morally required, let us call this the “bootstrapping” problem.²⁷ To better understand the bootstrapping problem, and what we need to do in order to solve it, we need a more precise criterion for why the above set of moral prescriptions for Baker is intuitively unacceptable. Since we are concerned with how the availability of the moral power of promising changes the moral character of Baker’s set of alternatives, I suggest the following criterion:

Principle II: A set of prescriptions for a sequence of actions, one member of which involves making a promise, is morally unacceptable if the agent, in fulfilling those prescriptions, performs a course of actions the underlying value of which is less than the underlying value of the course of actions she would perform if she fulfilled the prescriptions that would obtain if making a promise were not a possibility.

Applying the test of Principle II to *Nepotism*, we can see that if Baker carries out her prescriptions, she will make the promise at t_1 , and

²⁷I owe the term “bootstrapping” to Alvin Goldman.

then carry it out at t_2 . The underlying value of this course of actions is 2 ($P(A) [0] + A [2] = 2$). (Recall that the “underlying value” of an action includes all its moral value except that associated with its being a case of keeping or breaking a promise. Hence doing A has an underlying value of 2, not $2 + 10 = 12$.) If no promise were possible, her prescribed course of actions would involve not making a promise at t_1 , and then not giving the adjunct line to Classics at t_2 . The underlying value of this course of actions is 8 ($\text{not-}P(A) [0] + \text{not-}A [8] = 8$). Clearly the underlying value of the second course of action [8] is superior to that of the first course of action [2]. Hence, according to the test posed by Principle II, the set of prescriptions to make the promise and carry it out is morally unacceptable, just as we intuitively feel. *Nepotism I* is a genuine bootstrapping case.

5. Solving the Problem: Phase Two

What has led to the bootstrapping problem, and how can we avoid it? Clearly we do not want to reject promising altogether as an acceptable moral power, and just as clearly we do not want to prohibit making promises in cases, such as *Recruiting*, in which the agent does best to make the promise and carry it out, even though the promised act would have been wrong except for the promise.

The solution I shall argue for takes a different tack: it denies that keeping a promise has any positive moral value to be taken into account when determining what acts are to be performed. Jan Narveson has suggested that “common moral consciousness” holds that there is moral value in keeping a promise.²⁸ But I believe it is precisely this assumption—plausible as it is—that leads to the bootstrapping problem. We must reject it, and concede that while *breaking* promises (per se) has moral disvalue, *keeping* promises (per se) has no positive value at all.²⁹ Let us call this proposal the “Moral Costs” solution, since it interprets duties as imposing potential moral costs, but no potential moral goods, on agents who under-

²⁸Narveson, “The Desert Island Problem,” 279–80. In earlier work I too took the line that carrying out deontological duties, such as keeping a promise, had positive moral value (see my “Dated Rightness and Moral Obligation,” 460).

²⁹I am grateful to Alvin Goldman for convincing me to take a closer look at this solution than I was initially inclined to take.

take them. In describing *Nepotism I*, I followed what Narveson describes as common understanding (and what may be common philosophical practice) in ascribing a negative value (-10) to Baker's act of breaking her promise qua breaking a promise, and positive value (+10) to her act of keeping her promise qua keeping a promise. But if we take the alternative view that while breaking a promise makes an act worse than it would have been if it had never been promised, nonetheless keeping a promise does not make an act any better than it would have been if it had never been promised, the prescriptions in simple problematic cases such as *Nepotism* change significantly.

We need to examine whether this counterintuitive denial that keeping promises has moral value is an acceptable proposal. But before doing so, let us see how it solves the bootstrapping problem. On the Moral Costs analysis of the *Nepotism* case, it appears as follows:

Nepotism II	
Moral values	Underlying values
P(A) = 0	A = 2
not-P(A) = 0	not-A = 8
K(P) = 0	
B(P) = -10	
Values of options as of t1	
1. P(A) [0]	+ A [2 + 0] = 2
2. P(A) [0]	+ not-A [8 - 10] = -2
3. not-P(A) [0]	+ A [2] = 2
4. not-P(A) [0]	+ not-A [8] = 8*

On this way of analyzing the case, Baker ought as of t1 not to promise to allocate the funds to Classics and then allocate the funds elsewhere. Then, if she does not make the promise, she is required as of t2 to allocate the funds elsewhere (see below). If she carries out what she is required to do as of t1 and then again as of t2, she will not make the promise and will allocate the funds elsewhere. As we have seen, this is intuitively the set of prescrip-

tions we want. By ascribing no positive value to keeping a promise, and negative value to breaking a promise, we avoid the bootstrapping problem in this case: Baker is not given an undesirable set of prescriptions, but rather is morally prohibited from making a promise that would then require her to perform a morally inferior act at t2. She may not co-opt morality to support her pursuit of pernicious ends. These prescriptions do not violate the test posed by Principle II, since they do not require the agent to make a promise.³⁰

It is true that if Baker wrongly makes the promise at t1, she then comes to be required at t2 to carry out this promise.

Values of options as of t2

If P(A) at t1	If not-P(A) at t1
1. A [2] = 2*	1. A [2] = 2
2. not-A [8 - 10] = -2	2. not-A [8] = 8*

But this too may be correct: doing what is wrong often changes our future obligations. If I wrongly injure someone, I then become obligated to compensate him when I otherwise would have had no such obligation. Similarly, if I wrongly make a promise, I can then acquire an obligation to carry out that promise when otherwise I would have had no obligation to perform the act in question. If the damage to be done by carrying out the promise is sufficiently great, then of course I should not do so. But in some cases this will not be so, and *Nepotism II* is one of them.³¹

³⁰We can also see that the underlying value of the prescribed course of actions (not making the promise, and then not allocating the funds to Classics) is 8, whereas the highest underlying value of a course of actions involving making a promise (making the promise, and then not carrying it out) is 8, which is no higher than the underlying value of the course actually prescribed. Hence no underlying value is lost by this set of prescriptions.

³¹A case such as *Nepotism* is likely to have additional aspects not described in the text. For example, if the dean promises the adjunct position to Classics, the department will advertise the position, schedule and admit students to classes to be offered by the adjunct, and so forth. These would all be additional consequentialist reasons (stemming from detrimental reliance on the promise) for Baker to carry out her (wrongful) promise once made.

It is true, however, that even if there are no such additional consequentialist reasons for Baker to carry out her promise, the analysis in the text implies that she has an obligation to do so once the promise is made. By

It is readily ascertainable that the Moral Costs solution enables us to derive the right prescriptions not only in *Nepotism*, but in all simple problematic cases. It has this effect because of a combination of factors, most conveniently explained in terms of the following four sequence-types that are available at t1 to any agent in a simple problematic case:

making the promise and then carrying it out, Baker performs acts at t1 and t2 that morality condemns as of t1. However, by wrongfully making the promise, she has placed herself in a position where morality now requires her to perform the act that it earlier condemned. As Robert Cummins points out, it may seem counterintuitive to hold that allocating the funds to Classics is morally right from *any* temporal perspective. A clever agent working within such a moral system has the opportunity to manipulate morality to achieve illegitimate ends by making deft (if wrongful) promises. The Moral Costs solution does not preclude this result.

I think one might plausibly feel that this sort of result should be barred. Principle II could be expanded to capture this intuition. One way to bar this sort of situation would be to take the position that a promise wrongly made generates no deontic obligation to carry it out. On this view the only obligation that could arise from a wrongful promise of this type is one generated by potentially detrimental reliance on the promise (reliance by the promisee or others, such as the students just described in *Nepotism*). I suspect that the only wrongful promises we would want to treat in this manner are those that are wrongful in part because of the inferiority of the promised act. (For example, a promise that is wrong because it is insincere surely generates a deontic obligation to carry it out, whether or not there is detrimental reliance.)

Carrying out this idea coherently within the context of Principle I is complex. Perhaps the best procedure to follow would be (1) to assess, in exactly the same manner that we have used so far, whether it is obligatory as of t1 to make the promise and then carry it out, but then (2) to assess, without ascribing any disvalue to breaking the promise as such, whether it is obligatory as of t2 to carry out the promise. (The analysis as of t1 must incorporate disvalue to breaking the promise in order to obtain a prohibition against making the promise.) Using such a procedure, one derives a prescription for Baker, as of t1, not to make the promise and then to perform not-A; but if she disobeys this prescription and makes the promise anyway, she is then instructed as of t2 not to carry it out. These prescriptions assume there are no additional consequentialist disvalues arising from detrimental reliance on the wrongful promise; if there are significant disvalues of this type, then Baker would be instructed as of t2 to carry out the promise.

Interestingly, Atiyah (*Promises, Morals and Law*, 3, 9) points out that in the early English common law, a promise was only legally actionable if it was a promise to do something the promisor ought to have done anyway, and that French jurist François de Connan (1508–51) argued that a promise creates no obligation unless it is actually relied upon.

Options as of t1

- (i) P(A) & A
- (ii) P(A) & not-A
- (iii) not-P(A) & A
- (iv) not-P(A) & not-A

Although the exact values associated with these sequences will vary from case to case, nonetheless certain regularities will hold that ensure that sequence (iv) is always prescribed. First, since the underlying value of A is less than that of not-A, and neither promising nor failing to promise have any moral value, sequence (i) will always be worse than sequence (iv). Second, since sequence (ii) involves negative value for breaking the promise and is otherwise identical in value to sequence (iv), sequence (ii) too will always be inferior to sequence (iv), in which the same act is performed at t2 without breaking any promise. Thus, since the two sequences involving making a promise are worse than one involving not making a promise, the agent will be required not to make a promise. Of course if no promise is made, there is no existing promise to affect the values of the agent's choices at t2. Thus sequence (iv) is always better than sequence (iii), since not-A is better than A when no promise has been made. Hence the Moral Costs solution dissolves the bootstrapping problem for all simple problematic cases.³²

Fortunately the Moral Costs solution also permits us to derive the prescription we want in non-bootstrapping cases, such as *Recruiting*. To see this, let us assign numerical values to the various acts in the *Recruiting* case. This case is useful, not just because it is a non-bootstrapping case, but also because it enables us to see how the Moral Costs solution handles cases that are more complex than simple problematic cases of promising. These cases have additional complexities because (1) making, and not making, the promise are acts that themselves have moral value (typically via their consequences), and (2) the values of the promised act and its alternative are different in the context in which the promise has been made than they are in the context in which no promise has been

³²If the agent is not required to make a promise, his set of prescriptions cannot violate Principle II.

made.³³ In *Recruiting* we must ascribe moral values to the acts of making and not making the promise to Jones, since they have consequences independent of whether the promise is carried out (for example, making the promise induces Jones to accept the chairmanship). We must also note that the dean's possible acts at t2 have different consequences depending on whether he has promised Jones he would give the faculty lines to Chemistry. If he makes this promise and carries it out, his giving the lines to Chemistry has the additional slight effect of increasing trust in his integrity. If, on the other hand, he makes this promise and breaks it, his failure to give the lines to Chemistry not only has disvalue as a breaking of a promise, but has disvalue because it angers Jones and leads to widespread distrust of the dean. We can represent all this as follows:

Recruiting	
Moral values	Underlying values
P(A) = 11	A [if P(A)] = 5
not-P(A) = 2	A [if not-P(A)] = 4
K(P) = 0	not-A [if P(A)] = 6
B(P) = -10	not-A [if not-P(A)] = 9

Values of options as of t1		
1. P(A) [11]	+ A [5]	= 16*
2. P(A) [11]	+ not-A [6 - 10]	= 7
3. not-P(A) [2]	+ A [4]	= 6
4. not-P(A) [2]	+ not-A [9]	= 11

Values of options as of t2			
If P(A) at t1		If not P(A) at t1	
1. A [5]	= 5*	1. A [4]	= 4
2. not-A [6 - 10]	= -4	2. not-A [9]	= 9*

³³Other, still more complex, cases are ones in which the making of the promise affects which acts are available at all to the agent at t2.

Here we get a prescription as of t_1 for Dean Allen to make the promise to Jones and carry it out, since these acts are members of the best sequence, sequence 1, available to him as of t_1 . If he makes the promise as required, he then acquires an obligation as of t_2 to carry it out, since this is his best option. We can now see clearly that there is no bootstrapping problem here, since the sequence of acts the dean performs if he does what is required as of t_1 and then as of t_2 (namely, $P(A)$ and then A) has an underlying value of 16, while the best available alternative sequence not involving a promise (namely $\text{not-}P(A)$ and then $\text{not-}A$) has an inferior underlying value of only 11. Hence the prescriptions in this case satisfy the test of Principle II. These are the right prescriptions for the dean, and the ones we intuitively believe are correct.

6. Further Reflections on the Moral Costs Solution

It appears, then, that the Moral Costs solution enables us to avoid the bootstrapping problem in simple problematic cases, and at the same time enables us to derive intuitively correct prescriptions in more complicated cases, such as *Recruiting* in which the promised act, although it would be wrong apart from the promise, nonetheless forms part of a morally preferable sequence of acts, and so should be performed. But is the Moral Costs solution, which prohibits us from ascribing positive moral value to the keeping of promises, an inherently plausible view? Or does it trample too heavily on convictions that there is genuine moral value to carrying out one's duties?

I believe that common moral thought does tend to ascribe positive moral value to the keeping of promises. When asked, many people readily agree that the act of doing A , when one has promised to do it, has a different and superior moral quality to the act of doing A when no such promise has been made.³⁴ When pressed about why this should be, people note that keeping one's word is in itself an important accomplishment, often requiring significant resistance to temptation; that keeping one's promises leads other people to trust one; that keeping one's promises enables one to

³⁴In this and the succeeding paragraphs I am leaving out of consideration cases in which the promise should never have been made, or in which it is actually worse to keep the promise than to break it.

respect oneself; and that a general practice of promise making and promise keeping fosters beneficial social trust and cooperation.³⁵

But, with the possible exception of the first phenomenon, these are not benefits that are, so to speak, intrinsic to keeping one's promise per se; rather they are *consequences* of keeping one's promise that only arise in certain (very common) social and personal contexts. The man who wants to establish a good credit rating borrows money, promising to repay it, and then keeps his promise. His promise keeping has value because it causes others to trust him.³⁶ But the proverbial ship-wrecked traveler stranded on a desert island who makes a promise to his dying lone companion, and who finds no particular source of self-respect in carrying it out, would realize none of these values in fulfilling his promise.³⁷ Hence we cannot accept these benefits as part of the moral value of keeping promises per se, a deontological value that is present in all promise keepings apart from their context-dependent consequences. Since these effects do not always occur, they cannot be part of the explanation for the alleged fact that every act in which a promise is kept, insofar as it is a promise keeping, is better than that same act would have been if no promise had been made. Even the first phenomenon, the fact that keeping one's word is a sig-

³⁵These points were all made by participants at a University of Arizona colloquium discussion of this paper; Christopher Griffin in particular emphasized the importance of using promise making and promise keeping to create trust in oneself.

³⁶I owe this example to a referee for the *Philosophical Review*.

³⁷Perhaps a Kantian would say that this desert-islander, in keeping his promise, *makes himself* a person worthy of respect, whether or not he actually respects himself for this action. Since this would seem to be a feature of every act of keeping a promise, such a Kantian would then be in a position to assert that this is a kind of inherent value in all promise keeping. It appears this Kantian would also have to say that one should continually make and keep promises, as frequently as possible, so as to sustain one's respect-worthiness (otherwise he would have to admit that only a few promise keepings would have this value; the other promise keepings, coming too fast on the tail of the "effective" ones, would lack the value since they would do nothing to add to the agent's existing respect-worthiness). But this seems absurd. It would be better to develop this general line of thought by saying that anyone who, at a given time, is capable of making and keeping promises, is worthy of respect. Whether or not he actually engages in this practice is irrelevant, and should be governed by his circumstances, not by a need to establish in some metaphysical sense that he is worthy of respect.

nificant accomplishment, is true only when one must resist temptation to stay the course. An individual who promises to do what he wants to do anyway, and who never has the slightest reason to deviate from this promise, has hardly achieved a significant moral victory, or any victory at all. These considerations, then, do not establish that there is any deontological value to keeping promises *per se*.

Another common line of thought supporting the idea that promise keeping *per se* has positive value arises from the fact that we standardly have more reason to perform an act that has been promised than we would have had if the promise had never been made. It is easy to explain this fact by assuming that making a promise gives one an additional moral reason to perform the promised act. And, if we are trying to represent these ideas by ascribing numerical values to acts, it is natural to suppose that an act that has been promised—an act that one has an additional moral reason to perform—should be ascribed a higher moral value than that same act when it has not been promised.

But closer examination of this very natural train of thought tends to undermine its appeal. Consider the fact that, when faced with the question, most of us do *not* think that it is a reason for someone to *make a promise* that she would be able to keep it later. Keeping a promise does not seem to add any moral value to the world that must be taken into account when deciding whether to make that promise. People do not argue that one ought to make a promise now for the sole reason that one then will be able to do the right thing later on by keeping it. If I am already planning to do my grocery shopping on Saturday, I don't do anything better by promising my secretary I will do the shopping on Saturday and then fulfilling that promise. In view of this it begins to be tempting to conclude that making a promise in fact adds no moral value to the act of keeping that promise. We can notice further that although people do not think it is a reason for someone to make a promise that she would be able to fulfill it later on, people do argue that they should *not* make a promise now because there is some chance they would later break it. This asymmetry in the factors that we take into account in deciding whether to make a promise can readily be explained by the Moral Costs view, since it ascribes no positive value to keeping the promise (hence gives a person no reason in advance to make such a promise, even when she will

carry it out), but ascribes negative value to breaking a promise (and hence gives a person a clear reason in advance not to make it if she might not carry it out). If the Moral Costs view is correct, we can fully explain the natural view that making a promise gives one an additional moral reason to perform the promised act by noting that once the promise is made, *failing to keep it acquires disvalue that the act otherwise would not have had*. Hence one has reason to keep the promise, not because keeping it has greater value, but rather because breaking it would have less value, and so would be wrong. From this point of view, a person making a promise has everything to lose and nothing to gain.

In light of these considerations we can now see that my statement at the beginning that “a system of moral rules avoids this problem only if it ascribes no moral value at all to the keeping of promises” is, in a certain sense, ambiguous. It could be interpreted as a sweeping statement that there is no positive value in keeping promises nor any negative value in breaking them. But I meant the statement only in a different, narrower, sense as the restricted claim that there is no positive value in keeping promises. I do not endorse the broader statement, since we need to ascribe disvalue to breaking promises as an important part of the explanation for why it is wrong to break one’s word.³⁸

Further support for the Moral Costs view arises from the fact that when we survey an agent’s entire course of action from a timeless point of view, we do not ascribe any more moral value to a course of action that involves making and keeping a promise than we do to an otherwise identical course of action that does not involve promising. This is possible only if keeping a promise does not, per se, have positive moral value. Consider two agents, Pete and Tom, each of whom is asked to pledge a donation to his alma mater. Pete pledges \$25 and subsequently donates that amount. (Assume that the fundraising officials place no reliance on Pete’s pledge until they actually see his check.) Tom declines to make any pledge, but subsequently donates \$25. Has Pete, as compared to Tom, performed a morally better or preferable action in do-

³⁸One might ask whether a similar solution to the bootstrapping problem could be achieved by the reverse technique of maintaining positive value to keeping promises, but assigning no negative value to breaking promises. The answer is no.

nating his \$25, simply because Pete is fulfilling a promise while Tom is not? Is Pete's entire course of action better than Tom's? It doesn't seem so. It is true, of course, that Pete's act of promising places him under moral constraints that do not apply to Tom: Pete has an obligation that he would violate by failing to donate his \$25, while Tom would merely fail to perform a supererogatory act. But it appears we can fully capture this difference by noting that Pete's act of failing to donate the \$25 has a negative value arising from his failure to keep his promise, while a similar failure on Tom's part has no such negative value, since he made no promise. This fully accounts for our belief that Pete has more reason to donate the \$25 than Tom does. We need not ascribe any positive value to Pete's keeping his promise *per se*.³⁹ Thus we can explain the moral

³⁹This is a case in which it is not clear that simply ascribing single-dimensional moral "values" to acts can adequately reflect their moral status, where the types of moral status in question include being supererogatory as well as being obligatory, permissible, or wrong.

To see this, consider how we might try to describe this case by assigning values to the two agents' various options. Apparently the acts of giving money have some positive value because of their good consequences for the agents' alma maters. Let us say this consequential value is +10 in each case. Pete's act of failing to donate the money has a negative value of -10 because it is a breaking of a promise, while Tom's act of failing to donate money has no such disvalue. Let us assume that the acts of failing to donate money each have a consequential value of 0. We can then represent the agents' options as follows:

	PETE	TOM
1. Donate money	10	10
2. Not donate money	$0 - 10 = -10$	$0 - 0 = 0$

On this analysis we can say that Pete has more reason to donate the money than Tom does, because Pete's alternative of not donating the money would have far more negative value (-10) than Tom's would (0). Still this analysis does not adequately capture what we want to say about these two agents: it appears to imply that Tom, like Pete, ought to donate the money, since Tom's donating the money has greater moral value than its alternative, and so should be the prescribed act. But if we think that Tom's act is merely supererogatory rather than something he ought to do, this analysis does not accurately capture our view of the two cases.

Within the context of a theory that admits no obligation whatsoever to be benevolent, and only recognizes benevolent acts as supererogatory, we can represent which acts are required or not only by not assigning any positive moral value to any act merely because of its good consequences. Hence a more accurate analysis would appear as follows:

difference between their two acts of donating the money without invoking any positive value to keeping a promise.

Or consider another case. Imagine two possible worlds in which Sarah is a lifeguard at the beach. In both worlds a child is seen drowning, and Sarah has the duty of attempting to rescue the child. In World 1, the child's frantic mother simply watches the rescue operation. In World 2, the child's mother asks Sarah to promise that she will do everything possible to save the child's life, and Sarah makes this promise. In both worlds Sarah does everything possible to save the child's life. If keeping a promise receives positive moral value, then Sarah in World 2 does something with higher moral value than the act performed by Sarah in World 1. And yet this hardly seems right. If, on the other hand, both Sarahs abandon the rescue effort prematurely, it seems reasonable to judge not only that both Sarahs have done something wrong, but also that Sarah in World 2 has done something worse, since she has violated two duties (her official duty to rescue, and her promise to the mother), not just one.⁴⁰

Many people believe, on initial reflection, that an act which in-

	PETE	TOM
1. Donate money	0	0
2. Not donate money	$0 - 10 = -10$	$0 - 0 = 0$

Here Tom's two options appear as equal-valued and hence as both morally permissible. Pete's acts by contrast have different values: the one with the lower value is wrong, while the one with the higher value is morally required. The supererogatory character of Tom's act of donating money would have to be represented in some other way, perhaps by a numerical assignment on a different and noncomparable dimension.

⁴⁰Keith Lehrer has noted that under a rule-utilitarian scheme, it is plausible that promising would be recognized as a valuable utility-producing institution, and hence that keeping promises would be required and breaking them would be forbidden. Under such a scheme it appears coherent to hold that any act complying with the rules should receive positive value, and hence that keeping a promise should be accorded positive value, contrary to the Moral Costs solution. However, if a practice of promising incorporating this element is assessed by rule-utilitarian standards, it will be found to produce less utility than an alternative practice of promising that incorporates only negative values to breaking promises, because the former practice will be subject to utility-lowering bootstrapping cases, while the latter will not. Thus there does not appear to be a successful rule-utilitarian argument for according positive value to keeping promises.

volves keeping a promise has more value than that same act would have had if the promise had not been made. In addition, the fact that we have more reason to perform a promised act than we have to perform the same act unpromised seems, at first glance, to be best explained by ascribing positive moral value to promise keeping per se. If these natural ideas were correct, the Moral Costs solution to the bootstrapping problem would have to be rejected, and “common moral consciousness” would be inconsistent, since it would be committed by virtue of these ideas to moral obligations in bootstrapping cases that it intuitively rejects. But we have seen that these natural ideas tend to dissolve under closer scrutiny. We can say everything we need to say about the obligations generated by making promises without ascribing any positive moral value to the keeping of promises as such—we need only to ascribe negative moral value to the breaking of promises. We can conclude, then, that the Moral Costs solution may be adopted as our remedy for the bootstrapping problem.⁴¹

⁴¹The counterintuitive prescription in *Nepotism* for Dean Baker to promise the adjunct line to Classics (unless the Moral Costs solution is adopted) depends partly on the assumed fact that Dean Baker is not in a position to make any alternative promise (such as a promise to allocate the adjunct line to the German Department). Lack of such an option might occur in this case if, for example, the Chair of the German Department cannot be contacted at the time the promise would be made. If instead Baker has the option of making a promise to the Chair of German, as well as to the Chair of Classics, then the moral dynamics of the case change. To see this, consider *Nepotism III*, which includes such an option. To simplify the case I assume that Dean Baker must give the funds to either Classics or German.

Nepotism III

A = allocating funds to Classics; B = allocating funds to German; P(A) = promising to do A; P(B) = promising to do B; not-P = not promising to do anything; K(P) = keeping a promise; B(P) = breaking a promise.

Moral values	Underlying values
P(A) = 0	A = 2
P(B) = 0	B = 8
not-P = 0	
K(P) = 10	
B(P) = -10	

As the reader can work out, in this version of *Nepotism* we derive a prescription for Baker to make a promise to German, and then to give the

This discussion illuminates why the Moral Costs solution works to defuse the bootstrapping problem. If the promise to do A and the subsequent act of keeping that promise are assessed as a single sequence of actions, as I have argued they must be, then ascribing positive value to promise keeping per se creates a moral reason for the agent to make the promise initially. We can now understand this alleged reason to make the promise as a mistaken one that distorts the true value of the entire sequence, and so results in the objectionable prescriptions we have seen in the *Nepotism* cases. Once the Moral Costs solution is adopted, and the purported positive value to keeping the promise per se is eliminated, that value can no longer operate to give the agent a false reason for promising to do an act that would do more harm than good.

One important question that needs to be asked here is how far to *generalize* the Moral Costs view. I have argued that keeping a promise, in itself, has no moral value per se, although breaking a promise has negative value per se. Should parallel things be said for other kinds of special obligations incurred by one's own actions, such as the obligation to compensate those whom one has harmed, or to support one's child? And what about other special obligations that are not the result of one's own action, such as the obligation to support one's elderly parents?⁴² I shall not attempt

funds to the German Department. These prescriptions do not lead Baker to do anything that we intuitively judge to be *wrong*, but they do seem strange, since they mandate that she make a promise that intuitively does not seem to be required: nothing is genuinely gained by Baker's making this promise and then carrying it out, as compared to her dispensing with promises, and simply giving the money to German. This version of the case (which is similar to the lifeguard case and others in the main text) provides additional but perhaps more subtle evidence that keeping a promise should be accorded no moral value in itself, since it is only this value (10) that leads to the counterintuitive mandate that Baker make the promise to German. Under the Moral Costs solution, Baker is required simply to give the funds to German.

⁴²If I rear-end your car, doing \$250 worth of damage to your bumper, I owe you \$250 in compensation. Does my act of paying you the \$250 in compensation have no more moral value than my act of giving you \$250 if I don't owe it as compensation—if I had never rear-ended your car? Clearly my rear-ending your car, and coming under an obligation to compensate you, changes the moral character of my act of paying you \$250. Under these conditions the payment is obligatory, whereas if no accident had taken place, the payment would be supererogatory. But it's not at all clear that we want to say that the act of compensating you with \$250 is a better act, or has more moral value, than the act of simply giving you \$250.

to answer this broader question here, although I suspect that in these cases, too, fulfilling these obligations has no intrinsic moral value, although violating them has negative value. If so, then the bootstrapping problem may be eliminated by the Moral Costs solution for such cases as well. But fully addressing this question lies outside the scope of this paper.

We can say more general things, however, about the implications of the Moral Costs solution in complex—as opposed to simple—problematic cases involving promising. We have proved that the Moral Costs solution generates prescriptions satisfying Principle II in all simple cases of promising, and we can demonstrate this for more complex cases as well. One type of complex case is a kind of case in which the act of making a promise itself has a greater underlying value than the act of not making a promise (this is a feature of the *Recruiting* case, where the dean's making the promise to Jones will result in her accepting the chairmanship). Another common type of complex case is one in which the making of the promise itself makes a difference to the values of the promised act and its alternative (this is also a feature of the *Recruiting* case, and will be true in many promise making cases). It can be demonstrated in both these types of cases that any prescription for a sequence of acts involving making a promise satisfies Principle II. (For informal proofs of these claims, see the Appendix).

In section 3 I said that adopting the controversial assumption that we could assign acts numerical moral values could be construed as a temporary expedient to help us understand and devise a solution for the problem confronting us. We have now accomplished these aims, and it appears we can indeed set this temporary assumption behind us if we are worried about the cogency of numerical representations of moral values. All we need to say is that we have discovered we must reject any normative theory which holds that an act of keeping a promise is morally better than the same act would be if the promise had not been made. Such theories will prescribe wrongful acts, such as Dean Baker's allocating funds to Classics. These unwanted results can be avoided if we instead deny that acts of keeping a promise are better than they

Here again, it appears that supererogation must be calculated on a different value dimension than are obligations and prohibitions (see also note 39).

would have been if no promise had been made, but confirm that acts of breaking a promise are indeed worse than they would have been if no promise had been made. Such statements do not require that we be able to assign numerical values to the acts under consideration.⁴³

7. Conclusion

In this paper I have examined a little-noted problem with the moral power of promising: the availability of this power, as usually interpreted, creates situations in which an agent's making a promise can convert an otherwise wrong act into one that is morally required. Some such situations, for example, that in the *Recruiting* case, are intuitively acceptable; but others, such as that in the *Nepotism* case, are not. In these latter cases agents are actually required by morality to make abusive promises. I have argued that these problems can be eliminated entirely if we reinterpret the power of promising. We must determine whether to make a promise, or to keep it, by considering these acts, not as isolated events to be assessed independently, but rather as elements in a linked sequence to be evaluated in its entirety. And we must jettison the common assumption that keeping a promise has positive moral value, realizing that the act which keeps a promise has no more moral worth than it would have had if no promise had been made. We need to retain only our assumption that breaking a promise has negative moral value. A moral theory incorporating these recommendations can readily and accurately distinguish between the cases in which one genuinely ought to promise to perform an otherwise immoral act and those in which one must not so promise.

University of Arizona

⁴³Our theory must still enable us to judge which sequence of acts among those available to the agent at a given time is best. In principle it may be possible to make such judgments in a theoretically satisfactory way without assigning numerical values to the sequences or to their component acts. However, how this might be done is a substantial task confronting those who hold that assigning such numerical values is incoherent.

As Keith Lehrer and John Pollock have pointed out, if one retains the idea that the intrinsic value of acts may be measured on an *interval* scale, one can implement the Moral Costs solution by simply stipulating that keeping a promise to do A has the same value as simply doing A, whatever interval scale is chosen.

Appendix

This appendix presents an informal proof that applying the Moral Costs solution to certain complex cases of promising generates prescriptions that satisfy the test of Principle II. To satisfy Principle II it must be the case that an agent who is prescribed or permitted (by the Moral Costs solution) to make a promise, who makes this promise, and who does what is prescribed or permitted as of each time, will perform a course of actions that has no less underlying value than the course he would perform if he did what was prescribed or permitted at each time if his options were identical except that he did not have the option of making a promise. The kind of complex case in question has the same abstract features as those present in *Nepotism*: not performing the promised act has greater underlying value than performing the promised act, and breaking the promise has disvalue. In addition, the case may have the further features present in *Recruiting*: making the promise itself has greater value than not making the promise; and the underlying value of the promised act (or its nonperformance) varies depending on whether the promise is made. The proof will focus only on cases in which the agent must choose between two acts (A and not-A) at t2, and promising or not promising at t1 to perform A at time t2.

The values in this class of cases may be represented as follows:

$$\begin{aligned}
 \text{Value of P(A)} &= X \\
 \text{Value of not-P(A)} &= (X - Y) \\
 X &> (X - Y) \\
 \text{Value of A [if P(A)]} &= Z \\
 \text{Value of A [if not-P(A)]} &= Q \\
 \text{Value of not-A [if P(A)]} &= (Z + N) \\
 Z &< (Z + N) \\
 \text{Value of not-A [if not-P(A)]} &= (Q + M) \\
 Q &< (Q + M) \\
 \text{Value of B(P)} &= -W
 \end{aligned}$$

Then the values of the agent's options as of t1 are as follows:

1. $V[P(A) \ \& \ A]$ = 1. $X + Z$
2. $V[P(A) \ \& \ \text{not-}A \ \& \ B(P)]$ = 2. $X + Z + N - W$
3. $V[\text{not-}P(A) \ \& \ A]$ = 3. $(X - Y) + Q$
4. $V[\text{not-}P(A) \ \& \ \text{not-}A]$ = 4. $(X - Y) + (Q + M)$

There are four possible ways in which the agent might, without doing wrong, make a promise: either by being required or permitted to perform sequence 1, or by being required or permitted to perform sequence 2. We can examine these as two separate cases.

Case 1. Let us first consider the case in which the agent is either required or permitted to perform sequence 1.

For this prescription or this permission to obtain, it must be the case that sequence 1 is equal to or better than each of sequences 2, 3, and 4, that is,

$$(i) \quad 1. \ V[P(A) \ \& \ A] \geq 2. \ V[P(A) \ \& \ \text{not-}A \ \& \ B(P)]$$

$$X + Z \geq X + Z + N - W$$

$$W \geq N$$

$$(ii) \quad 1. \ V[P(A) \ \& \ A] \geq 3. \ V[\text{not-}P(A) \ \& \ A]$$

$$X + Z \geq (X - Y) + Q$$

$$Z \geq Q - Y$$

$$(iii) \quad 1. \ V[P(A) \ \& \ A] \geq 4. \ V[\text{not-}P(A) \ \& \ \text{not-}A]$$

$$X + Z \geq (X - Y) + (Q + M)$$

$$Z \geq Q - Y + M$$

If sequence 1 is prescribed or permitted, then an agent who does what is required or permitted as of each time may, or possibly must, perform $P(A)$ at $t1$. Even if it is permissible for him to perform $\text{not-}P(A)$, the case in which he actually performs $P(A)$ is the one in which we are interested, since only a sequence involving promising can give rise to an abuse of this moral power. If the agent performs $P(A)$ at $t1$, he then must choose between A and $\text{not-}A$ at $t2$.

Act A will be prescribed as of $t2$ only if the value to be obtained

by performing A, namely Z, is greater than the values to be obtained by performing not-A in this context, namely $(Z + N) - W$.

$$\begin{aligned} V[A] &> V[\text{not-A}] \\ Z &> (Z + N) - W \\ 0 &> N - W \\ W &> N \end{aligned}$$

Thus A will be prescribed only if $W > N$; if $N > W$, not-A will be prescribed. If $W = N$, then either A or not-A will be permitted as of t2.

Thus, if sequence 1 is prescribed or permitted, and the agent does what is required or permitted as of each time, he will do P(A) at t1 (in the interesting case) and either A or not-A at t2. In order for these prescriptions to satisfy Principle II, it must be the case that the underlying values of the following two courses of action, (a) and (b), are each (if permissible) no less than the underlying value of the course of action that would be prescribed if making a promise were not a possibility.

- (a) P(A) and A
- (b) P(A) and not-A

If making a promise were not a possibility, the agent would have only sequences 3 and 4 as options. Since by hypothesis the value of not-A (in sequence 4) is greater than the value of A (in sequence 3), and the sequences are otherwise identical in underlying value, the agent would be required as of t1 to perform sequence 4. If the agent performed the first act of sequence 4 at t1, he would then be required as of t2 to do not-A at t2, because the value of not-A is greater than the value of A. Hence the course of action he would perform, if he did as he was required as of each time, would consist in:

- (c) not-P(A) and then not-A.

There are then two sub-cases to examine, depending on whether the agent permissibly does A or does not-A at t2 after making the promise at t1. In one sub-case we must determine whether the underlying value of course (a) is no less than the underlying value of course (c). In the second sub-case we must determine whether

the underlying value of course (b) is no less than the underlying value of course (c). The underlying values of the three courses of action are as follows:

- | | |
|-------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| <p>(a) $V[P(A) \text{ and } A]$
 $X + Z$</p> | <p>(c) $V[\text{not-}P(A) \text{ and not-}A]$
 $X - Y + Q + M$</p> |
| <p>(b) $V[P(A) \text{ and not-}A]$
 $X + Z + N$</p> | |

First sub-case: The agent permissibly performs course (a). This permission satisfies Principle II only if the underlying value of course (a) is no less than the underlying value of the best non-promising alternative, namely course (c). But by hypothesis this is true, since in case 1, sequence 1 \geq sequence 4, and the underlying values of these two sequences are identical with their overall values. Hence a prescription or permission to perform sequence 1, followed by a prescription or permission to perform A, satisfies Principle II.

Second sub-case: The agent permissibly performs course (b). This permission satisfies Principle II only if the underlying value of course (b) is no less than the underlying value of the best non-promising alternative, namely course (c). Thus the question is whether $X + Z + N \geq X - Y + Q + M$, or, more simply, whether $Z + N \geq Q - Y + M$. We know by hypothesis (see (iii) above) in case 1 that $Z \geq Q > Y + M$. We also know that $Z < Z + N$. Hence $Z + N$ must be $> Q - Y + M$. Thus a prescription or permission to perform sequence 1, followed by a prescription or permission to perform not-A, also satisfies Principle II.

Thus we have shown that in the case where the agent is either permitted or required to perform sequence 1, the courses of action he might follow if he made the promise and always did what was required or permitted as of each time would have underlying values no less than those of the best course of action not involving a promise.

Case 2. Let us next consider the case in which the agent is either required or permitted to perform sequence 2. For this prescription

or permission to obtain, it must be the case that sequence 2 is better than, or equal to, each of sequences 1, 3, and 4, that is,

$$2. \quad V[P(A) \ \& \ \text{not-}A \ \& \ B(P)] \geq 1. \quad V[P(A) \ \& \ A]$$

$$X + Z + N - W \geq X + Z$$

$$N \geq W$$

$$2. \quad V[P(A) \ \& \ \text{not-}A \ \& \ B(P)] \geq 3. \quad V[\text{not-}P(A) \ \& \ A]$$

$$X + Z + N - W \geq (X - Y) + Q$$

$$Z + N - W \geq Q - Y$$

$$2. \quad V[P(A) \ \& \ \text{not-}A \ \& \ B(P)] \geq 4. \quad V[\text{not-}P(A) \ \& \ \text{not-}A]$$

$$X + Z + N - W \geq (X - Y) + (Q + M)$$

$$Z + N - W \geq Q - Y + M$$

If sequence 2 is prescribed or permitted, then an agent who does what is required as of each time may, or possibly must, perform $P(A)$ at t_1 . As before, the interesting case for us is the one in which he does perform $P(A)$. If he performs $P(A)$ at t_1 , he then must choose between A and $\text{not-}A$ at t_2 .

As we saw under case I, A will be prescribed only if $W > N$, $\text{not-}A$ will be prescribed only if $N > W$, and either A or $\text{not-}A$ will be permitted if $N = W$.

Thus, if sequence 2 is required or permitted, the agent carries this out by performing $P(A)$ at t_1 , and then does what is required or permitted as of t_2 , he will do $P(A)$ at t_1 and either A or $\text{not-}A$ at t_2 . In order for these prescriptions to satisfy Principle II, it must be the case that the underlying value of course (a) and the underlying value of course (b) are each no less than the underlying value of the course of action that would be prescribed if making a promise were not a possibility. As we saw under case 1, if making a promise were not a possibility, the agent would be required to perform course (c). So in ascertaining the acceptability of prescriptions arising out of the requirement or permission to perform sequence 2, we need to compare the very same three courses of action we compared in examining sequence 1. We saw in that case already that the underlying values of courses (a) and (b) were no less than those of course (c), and the same result of course obtains here.

Hence we can straightway conclude that in the case where the agent is either permitted or required to perform sequence 2, the courses of action he might follow if he made the promise and always did what was required or permitted as of each time would have underlying values no less than those of the best course of action not involving a promise. Here, too, carrying out a set of prescriptions involving making a promise would not violate Principle II.