

1-2006

Hume on Promises and the Peculiar Act of the Mind

Rachel Cohon

University at Albany, State University of New York, rcohon@albany.edu

Follow this and additional works at: http://scholarsarchive.library.albany.edu/cas_philosophy_scholar



Part of the [Philosophy Commons](#)

Recommended Citation

Cohon, Rachel, "Hume on Promises and the Peculiar Act of the Mind" (2006). *Philosophy Faculty Scholarship*. 13.
http://scholarsarchive.library.albany.edu/cas_philosophy_scholar/13

This Article is brought to you for free and open access by the Philosophy at Scholars Archive. It has been accepted for inclusion in Philosophy Faculty Scholarship by an authorized administrator of Scholars Archive. For more information, please contact scholarsarchive@albany.edu.

Hume on Promises and the Peculiar Act of the Mind

RACHEL COHON*

I. HUME'S TWO CLAIMS ABOUT FIDELITY TO PROMISES

PROMISING—WITH ITS MORE FORMAL MANIFESTATION, CONTRACT—is a vital part of social life. Most forms of long-term cooperation, including commerce, depend in some way on the fact that people can bind themselves to others now to perform actions later. One philosophically intriguing question about promises is how this self-binding is possible. Furthermore, there is a special character trait associated with the honoring of promises that calls out for analysis. Intuitively speaking, being a person of your word, someone who is trustworthy about promises, is an admirable trait, one we inculcate in our children and long for in our public figures—in short, a virtue. What does this trait consist of?

In Book 3, Part 2, Section 5 of the *Treatise of Human Nature*, Hume attempts to analyze and explain the virtue of being a person of one's word, a virtue he calls fidelity, and also to give an account of how promising generates obligation.¹ He seeks an explanation of the virtue as we find it in experience. Of course, it is to be an analysis in terms of Hume's own moral theory. In the course of his explanation of fidelity, Hume makes two very puzzling claims. In this paper I offer an interpretation of Hume's account of fidelity that makes good sense of these claims. As a coda I draw a few cautious conclusions about virtue and duty from Hume's insights.

What Hume says about fidelity to promises is laconic, and interpreters have struggled to make sense of it. We need a bit of background to set out the two specific puzzling claims on which I shall focus.

The first claim is that a particular conception of the virtue of fidelity to promises is vulnerable to a *reductio ad absurdum*. The vulnerable view is the view that

¹ David Hume, *A Treatise of Human Nature*. Citations in parentheses in the text will be given to two editions of this work, for example as "(T 3.2.5.1, SBN 516)." Numerals immediately following the 'T' indicate Book, Part, Section, and paragraph of David Hume, *A Treatise of Human Nature*, eds. David Fate Norton and Mary J. Norton (Oxford: Clarendon Press, 2000); numerals following 'SBN' indicate a page in David Hume, *A Treatise of Human Nature*, 2nd ed., eds. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Oxford University Press, 1978).

* **Rachel Cohon** is Associate Professor of Philosophy at the University at Albany, State University of New York.

fidelity is what Hume calls a natural virtue, and the obligation of promises a natural duty, as distinct from an artificial one. For Hume, some virtues, and the moral obligations associated with them, depend for their existence upon the prior presence of a special social custom or convention and could not occur without it. These are the artificial virtues; today we would more naturally call them conventional. Hume argues that fidelity is one of these convention-based virtues. The conception of fidelity that he attacks, then, is that fidelity is a natural, or nonconventional, virtue. On that view, the virtue of fidelity, and the phenomenon of promise-based obligation, do *not* depend for their existence upon any custom or convention, and could occur even in the absence of any jointly-made social rules. Hume thinks there indeed are some nonconventional virtues, among them parental attentiveness and benevolence, but fidelity is not one of them. What is puzzling is Hume's diagnosis of what is wrong with nonconventionalism about fidelity. He claims that nonconventionalists are committed to supposing that there is a "peculiar act of the mind" involved in promising—which turns out to be an act of *willing to be obligated*—that creates obligation when each promise is made, and that motivates virtue-manifesting compliance with the promise. The act in question, Hume argues, is an impossible one. Therefore, he concludes, nonconventionalism about fidelity to promises is false. Thus he says:

If anyone dissent from this [that fidelity is no natural virtue], he must give a regular proof of these two propositions, viz. *that there is a peculiar act of the mind, annexed to promises; and that consequent to this act of the mind, there arises an inclination to perform, distinct from a sense of duty.* I presume, that it is impossible to prove either of these two points; and therefore I venture to conclude, that promises are human inventions, founded on the necessities and interests of society. (T 3.2.5.7, SBN 519)²

Hume does not explain, however, why the nonconventionalist must rely on a special act of the mind to support his position.

The second claim that I shall discuss is even more cryptic. Hume says that even though the obligation of promises and the virtue of fidelity in fact depend upon social convention, and there *is* no peculiar act of the mind, we nonetheless pretend or imagine that whenever someone promises she wills a new obligation into existence. Our common-sense thought about promising "feigns" the occurrence of the peculiar act of the mind, and apparently cannot avoid doing so. Thus he says:

After [the invention of promises] a sentiment of morals concurs with interest, and becomes a new obligation upon mankind . . . The difficulties, that occur to us, in supposing a moral obligation to attend promises, we either surmount or elude . . . Here, therefore, we *feign* a new act of the mind, which we call the *willing* an obligation; and on this we suppose the morality to depend. (T 3.2.5.12, SBN 523)

Some who explicate Hume's theory of the artificial virtues do not touch on these two claims about the peculiar mental act.³ Commentators who do try to account for them have so far been able to do so only by postulating various sorts

² All emphasis is Hume's (as it is here), except where noted to the contrary.

³ For example, there is no discussion of the peculiar act of mind in the otherwise illuminating account of Hume's theory of promise given by Annette Baier in *A Progress of Sentiments [Progress]* (Cambridge, MA: Harvard Press, 1991), ch.10. I have learned a great deal, however, from Baier's detailed and perceptive interpretation.

of errors or weaknesses in Hume's theory. The hope of the present paper is to do better on Hume's behalf.

I will explain why Hume thinks the nonconventionalist about promises is committed to the impossible mental act of willing to be obligated, and how Hume constructs his *reductio* of that supposition. I will argue that Hume is attending to two general questions about promises: (1) What is the virtuous motive at work when a faithful person keeps a promise?, and (2) How do we obligate ourselves merely by speaking? Hume determines that only certain answers to these questions are compatible with our experience of this virtue. If we start with the compatible answers, combine those with Hume's general theory about virtue and duty, and suppose the absence of social convention, the peculiar mental act view more or less follows. I will also explain why Hume thinks that even though nonconventionalism is false and conventionalism is true, we still accept a fiction that there is such a peculiar act of the mind involved in promising. We do this, if my interpretation of Hume is right, because at some level we are all nonconventionalists: our ordinary intuitions about virtue and vice are compatible only with nonconventionalism. We expect all virtues to have the same structure as benevolence and friendship, which are non-conventional, and we accept a fiction in order to make fidelity look as if it has this structure as well.

2. THE TWO QUESTIONS ABOUT PROMISES

At the outset I noted two interesting broad questions about promises. Let me elaborate on them a little.

One is a special question that arises for a moral theory focusing on the virtues and vices. Fidelity to promises, or trustworthiness with respect to one's word, is a character trait we recognize as a virtue. A character trait is a fairly settled disposition to have certain feelings and desires which motivate one to act in characteristic ways in a range of circumstances. It is not just a behavioral disposition; it is an emotional and motivational one. To behave virtuously, one's action must be motivated by one or more of these suitable feelings or desires. This is because to act virtuously or well, it is not enough to do the right thing; one must do it for the right reasons. I don't manifest my benevolence in helping a homeless person if I help her only out of the desire to improve my reputation. What I do is right, but not benevolent. An action that manifests the virtue of benevolence is motivated, at least in part, by concern for the other person's welfare. That is the right reason. Now, what is the right reason for keeping a promise? What motivates a trustworthy person, a person of his word, to fulfill a promise when he does so in a way that manifests his virtue? It should presumably be whatever feelings or desires comprise this good character trait. Now, the appropriate—the virtuous—motive for keeping a promise is not self-interest. If I keep my promise only out of self-interest, I may manifest the virtue of prudence, but not of trustworthiness. A trustworthy person will keep her promise even where it is contrary to her interest; that is often the point of exacting a promise from her. But interestingly enough, the characteristic, virtuous motive for keeping a promise is also not concern for the other person's welfare. A trustworthy person will of course keep promises made to those she hopes to help. But she will also keep promises to people she has no

reason to care about at all, and to those who receive little benefit or even receive harm from her fulfillment of her promise. She will be motivated to fulfill contracts with banks and credit card companies, though she has no reason to be concerned for their well-being. She will be inclined to fulfill her promise to repay a large loan to Hume's "profligate debauchee" (T 3.2.1.13, SBN 482), who will use the money only to harm himself with greater debaucheries. Nor is it a manifestation of trustworthiness about promises to fulfill them out of public spirit. If I keep my promise solely out of public spirit, because it will help my family, my community, or my country, I show my loyalty to the group but not my trustworthiness. (Keeping some promises will not help the group, but a trustworthy person will see reason to keep them nonetheless.) I might perform a public-spirited action anyway, without having promised, out of concern for my group; but if I am a person of my word, then when I promise, this should add an additional reason to do it. In any case, when I meet a colleague for lunch as promised, I am not actually moved to keep my promise by any thought of the public good.

What really shows that someone is a person of his word is that he keeps his promise because it is a *promise*, because it is right to keep one's promises and wrong to break them, and because he wants to be a keeper of promises and not a breaker of them. But what kind of feeling or desire *is* this? If we think of a different virtue, such as benevolence, we can understand the motivating feeling there without any recourse to moral concepts. The motive that manifests the virtue of benevolence is the desire for the happiness of another person; this can be understood entirely apart from the moral concepts of virtue and vice, right and wrong. But the analogous motivating attitude for fidelity to promises, the desire to keep one's promises because it is right to do so and wrong not to, cannot be so understood. It draws upon a prior understanding of right and wrong action, at least, and perhaps also on a prior understanding of the virtue of being a trustworthy or honorable person.

For some theories of virtue ethics, including Hume's but not limited to his, this is problematic. It runs afoul of the doctrine of the primacy of character over action. This doctrine says that right action is to be defined with reference to good character, as whatever the person who possesses the feelings and desires that compose the virtue would be moved to do.

The desire to keep one's promises because it is obligatory to do so also appears to be identical with the motive of duty. This causes some trouble for those virtue theories, both Hume's and others, which deny that genuinely good action, or the best action, can be motivated by the sense of duty.

The second general question about promises that I mentioned arises more or less regardless of what moral theory you hold. In promising, I say certain words, in a certain sort of context, and perhaps with a certain sort of intention, and by doing so I make it morally obligatory for me to do A and morally wrong for me to fail to do A, *ceteris paribus*. How can I obligate myself to act in a certain way just by saying some words, perhaps in a special context and with special intentions? After all, I cannot obligate myself, for example, to touch my ear, just by standing in my office alone and saying "I hereby obligate myself to touch my ear." I can *say* this, but it does not in fact generate obligation. I can, however, obligate myself to pay

someone \$10 by saying “I hereby promise to pay you \$10.” Other moral obligations are not like this. I have general obligations—to tell the truth, to help those in need, not to kill or maim others—but these are not created by me at a stroke; they are always there. And I can incur obligation by some *action* of mine, as I do if I injure you or damage your property and incur a duty to compensate you, or as I typically do if I bring a child into the world. But in promising it seems that I incur obligation just by saying that I do. This power of creating moral obligation just by speaking has struck some philosophers as so singular that they think it can only be explained by a special social rule, something over and above whatever general moral obligations we have, and something over and above the ordinary conventions of language. These are the conventionalists, and not only Hume but Rawls and Anscombe are among them.⁴ Most of Hume’s predecessors, on the other hand, were nonconventionalists; and today the nonconventional view is held, most famously, by Scanlon.⁵

To some extent, I claim, Hume’s puzzling remarks about fidelity are to be explained by his own attempt to answer these two questions, of the motive characteristic of fidelity and the creation of obligation by fiat, in a way faithful to our experience of the virtue.

3. HUME’S SENTIMENTALISM AND HIS VIRTUE ETHICS

Of course, the features of Hume’s over-all moral theory provide important pieces of the puzzle. When Hume takes up the virtue of fidelity to promises in the *Treatise*, he has most of the main planks of his moral theory in place. A character trait, for Hume, is a psychological feature of a person, in particular a motivating sentiment or a disposition to experience a motivating sentiment under certain circumstances. For Hume morals are derived from sentiment and not from reason. This means all moral properties of persons and their actions (goodness, evil, rightness, etc.) are determined by feelings of approval or disapproval on the part of observers. A *virtue* is a character trait (a motivating sentiment) which elicits the sentiment of approval from any observer who contemplates it in a disinterested way and in light of the relevant facts. A *vice* is one that elicits disapproval. Approval (approbation) is a pleasant sentiment of a special sort; disapproval (disapprobation) a painful or uneasy one. Because of the nature of the human psyche, people who possess the same factual knowledge and the same imaginative abilities and who contemplate traits from a common point of view will feel approval and disapproval in response to the same traits, and so will on the whole agree in their moral assessments. Furthermore, Hume is explicitly committed to virtue ethics, includ-

⁴ John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 344–50; G. E. M. Anscombe, “Rules, Rights and Promises” and “On Promising and Its Justice, and Whether It Needs Be Respected *In Foro Interno*,” in her *Collected Philosophical Papers V.III* (Minneapolis: University of Minnesota Press, 1981), 97–103, 10–21. N.B.: Anscombe does not exactly think that the obligation of promises depends upon a rule distinct from the rules of language. She thinks it depends upon a special rule, but it is (roughly speaking) a rule of moral language.

⁵ See T. M. Scanlon, “Promises and Practices,” *Philosophy and Public Affairs* 19 (1990): 199–226, and *What We Owe to Each Other* (Cambridge, MA: Harvard Press, 1998), ch. 7. Another recent author who takes a nonconventionalist view of promising is Michael Robins, *Promising, Intending, and Moral Autonomy* (Cambridge, U.K.: Cambridge University Press, 1984).

ing the primacy of character. For him, traits of character are the primary objects of moral assessment.

Hume takes as his paradigmatic virtues the natural virtues, such as benevolence, parental attentiveness, and gratitude. He thinks people could have these virtues even if there were no organized society that united individuals not bound by ties of affection. Another group of virtues—honesty with respect to property (“justice”), fidelity to promises, allegiance to government, and chastity and modesty—Hume thinks cannot successfully be understood this way, and he argues that they are artificial: their existence depends upon some sort of jointly-devised social convention. He does not assume this, I think; he *concludes* it from the peculiarities of these virtues. Hume’s *reductio* argument against the nonconventionalist view of promises is his second salvo in the struggle to show that some virtues are artificial. (The first was about honesty with respect to property.)

There is evidence in the text that Hume thinks we have a common-sense *conception* of a virtue as having a particular structure (T 3.2.1.2–7, SBN 477–79; T 3.2.1.18, SBN 483–84; and T 3.2.2.8, SBN 488–89). Let us call this the natural conception of virtue, since it coincides with the actual structure of the natural virtues. Human beings are prone to feel love of their offspring, compassion for the suffering, and gratitude to their benefactors, to at least some extent. When these potentially motivating passions are strong and enduring enough in a certain person, we approve them when we contemplate them in the way appropriate for moral assessment. If we contemplate the degree of gratitude a person feels toward her benefactor in light of the facts and in a disinterested way, and find we approve it, then this motivating sentiment is a virtue. We may also contemplate someone’s weak, inadequate, or missing grateful feeling, and disapprove the deficiency. The deficiency of which we thus disapprove is consequently a vice. (Hume may also think there are vices of excess of a motivating sentiment; there again, what makes the trait a vice is that on reflection in the right frame of mind we disapprove of it.) We approve or disapprove of *actions* derivatively, because of the sentiments we take to have motivated them. An action is bad because its motive was bad. If our guess as to what motivated an action proves incorrect, we revise our assessment of the action. So, if we take a man’s neglect of his benefactor to be the result of cold-hearted selfishness, we disapprove of the action. If we find out instead that he had a powerful motive of gratitude but failed to help for some other, good reason, we cease to blame him and instead think his action a good one. Hume says, “The outward performance has no merit. We must look within to find the moral quality” (T 3.2.1.2, SBN 477). So according to the natural conception of virtue and vice, our assessment of the action—our approval or disapproval of it, which is the only sort of moral assessment for which Hume allows—depends entirely on our assessment of its motive. Thus, Hume’s conception of virtue incorporates the doctrine of the primacy of character over action. (The doctrine of the primacy of character over action defines a right or good act as whatever the virtue would dispose one to do. Hume understands a virtue to be a motivating sentiment. Hume’s version of the doctrine, then, is that a right or good act is whatever act that sentiment would motivate one to do.)

It follows from the natural conception of virtue that the motivating sentiment or motive of a virtuous action must be something distinct from the moral senti-

ment. The moral sentiment, or sense of virtue, is approval and disapproval. According to the natural conception of virtue, this is activated only in response to some existing motive, such as gratitude. So the sentiment of gratitude, or whatever the key motivating sentiment is in the case of the virtue in question, cannot be identical to the moral sentiment, Hume says (T 3.2.1.2–7, SBN 477–79; and T 3.2.5.6, SBN 518). There must be a prior, nonmoral motive of action of which we can in turn approve, if any approval is to occur at all, and so if the action is to qualify as virtuous. Approval cannot itself be the motive of which we approve. Or so it would seem.

Hume also gives a brief account of duty (T 3.2.1.8, SBN 479; and T 3.2.5.6, SBN 518). First, he says that it is my duty to have a certain motivating sentiment just in case my having it would elicit approval and my failure to have it, or deficiency of it, would elicit disapproval. Next, an *action* is my duty if it is the sort of action that would be motivated by an obligatory motivating sentiment. Thus, for example, it is my duty to love my children. (This is because we tend to disapprove of the absence or miniscule amount of what is “normal and natural” in human nature, which love of offspring is.) Because it is my duty to love my children, it is my duty to give my children the care and education that such love would motivate. So actions can be duties as well, but again, derivatively. Now if someone is deficient in love of offspring, anyone will feel moral disapproval. If I am the deficient one, I will disapprove of my own temperament and the actions to which it gives rise. This self-disapproval is my *sense* of duty. This moral sentiment, the sense of duty, may be unpleasant enough to move me to try to develop the missing parental affection, and I might try to do this *by* looking after my children. Or alternatively this self-disapproval may move me to look after my children in order to disguise to myself my lack of love for them. So in this indirect way, the sense of duty can be motivating. Thus there is a *motive* of duty, for Hume. But it has two important features, at least as it first appears in his ethical theory. First, the motive of duty can exist only if there is first some *nonmoral* motive common in human nature—that is, some motive distinct from the moral sentiment—the absence or deficiency of which we disapprove. Second, the motive of duty is at most a second-best motive. A parent who tends to her children out of love has the virtue of parental attentiveness; a parent who does so from the motive of duty is deficient, not genuinely virtuous. She executes the “outward performance” for at most a second-best reason.

Hume encounters a general problem about the virtues he labels artificial. (I think he classifies them as artificial largely because he finds them problematic in this way.) He elucidates the problem mainly in his discussion of honesty with respect to property, but he says the same thing occurs with fidelity to promises, so I will use that to illustrate it. When we try to specify the approved motive of an act of promise-keeping—the motive because of which the action exhibits the virtue of fidelity—we find it can be only a *moral* motive, the moral sentiment itself. There is no nonmoral motive of promise-keeping that manifests the virtue of fidelity. Consider the possibilities I briefly suggested before, which are all taken up and rejected by Hume. The faithful person is not the one who keeps promises out of self-interest, or from benevolence. For these do not specifically motivate us to keep promises; indeed, these sentiments can move us to break promises just as

well. The faithful person is not even one who keeps promises out of concern for the public good, for as Hume says, “men, in the ordinary conduct of life, look not so far as the public interest, when they . . . perform their promises . . . That is a motive too remote and too sublime to affect the generality of mankind” (T 3.2.1.11, SBN 481). The motive that really exhibits a person’s status as a person of his word, the one that elicits our approval of his promise-keeping action when we contemplate it appropriately, is his desire to keep the promise because he thinks he *should* keep it and not break it. For Hume this is just his desire to keep it because he approves of keeping promises and disapproves of breaking them. Hume says, “If we thought, that promises had no moral obligation, we never shou’d feel any inclination to observe them” (T 3.2.5.6, SBN 518). But if the natural conception of virtuous motives and virtuous action that Hume attributes to us is correct, such a motive is impossible. In the absence of a *nonmoral* virtuous motive of promise-keeping that can stimulate our approval, we can *have* no approval of acts of promise-keeping. The outward performance alone has no merit or demerit; the approval of the acts must be derived from approval of a nonmoral motive. There is no nonmoral motive. So we have no approval of the acts. If approval is what motivates the acts, then we have nothing that can motivate such acts. That is the general problem.

4. HUME’S SPECIAL PROBLEMS WITH FIDELITY TO PROMISES

But that is not the only problem with fidelity to promises that concerns Hume. That problem, which plagues the other artificial virtues as well, does not lead him to expect any peculiar act of the mind associated with the others. The “peculiar act” accusation arises because he is also concerned with a problem unique to fidelity. Promising creates a new moral obligation. Of course it is possible to promise to do something that is already obligatory for other reasons, but a great many of the actions I promise to perform are not obligatory until I promise to do them, and become so only *because* I promise (for example, to read the work of thirty undergraduates or to send money to a certain mortgage company). It is not some intrinsic property of the actions that makes them obligatory. Somehow either my saying “I promise” in the right way or the right place creates the new duty, or something else that comes into being simultaneously with that utterance creates the new duty. It cannot be something that was present before the utterance, or I would have been obligated before, which I was not.

Now, if we reflect on what counts and what does not count as a promise in daily life, we cannot help but notice that it is not merely my saying the words that creates the duty, or even my saying them to someone who hears them. For if I don’t understand the words I utter, or say them as an obvious joke, they create no obligation. As Hume observes, “one who shou’d make use of any expression [of promise], of which he knows not the meaning, and which he uses without any intention of binding himself, wou’d not certainly be bound by it” (T 3.2.5.13, SBN 523). So merely saying the words is not sufficient to obligate. Now, nothing else occurs in the world outside of me when I promise, besides my speaking and being heard (or so it certainly seems). And my understanding or my intention to bind myself, or both, are relevant. So presumably what differentiates binding promising from

merely saying the words without incurring obligation is something that happens in my mind. There must be something specially suited to promising that occurs in my mind whenever I make a binding (that is, an actual) promise, in virtue of which I am obligated—something peculiar to the creation of the duty of promises. What could this be?⁶

Now we have the materials at hand to set up Hume's *reductio* of the non-conventionalist view of promises. He assumes the nonconventionalist will be committed to the natural conceptions of virtue and duty. On these conceptions, in order for an action to be my duty, there must be some motive for it that is found in human nature, and this motive must be distinct from the moral sentiment, and so distinct from the sense of duty. Here we should notice something about natural motivating sentiments and the actions they motivate. The natural sentiments of "humanity" (concern for other people) and love of offspring, for example, conduce to actions that share a certain feature: the actions serve the goals at which those passions aim: relief of suffering in the one case, benefit to children in the other.⁷ But a huge variety of actions can be made obligatory by promising, and these actions need not have any characteristic in common that makes them the object of a nonmoral motivating sentiment. In particular, promised acts don't all tend to fulfill any particular goal of agents. Some of them aid the agent, some aid the agent's business partners, customers, students, or political constituents; some do not aid anyone, such as actions promised to those now dead, or to a profligate debauchee. Now for Hume, all motivating passions are goal-directed. What motive could there be that could move us to any sort of action just in virtue of an event that occurs in the agent's mind at the time she says "I promise?"⁸ We need a *generic* motive, one not directed at any inherent features of the actions it motivates, so it could be directed to any action. A reasonable candidate for this is a motive whose object is, not some outcome or goal, but the action itself. And we

⁶ It bears emphasis that "the intention of binding himself" here is not the intention to *keep* the promise. Hume is well aware that we can make binding promises we do not intend to keep, and acknowledges that this forms a part of our common sense view (T 3.2.5.13).

However, there must be something (else) that distinguishes genuine promises from those utterances of 'I promise . . .' that do not bind. And as we have seen, given the mutually-acknowledged facts, it is natural to suppose (incorrectly, as Hume will ultimately argue) that the distinguishing phenomenon is something in the mind. What must go on in the person's mind might only be the awareness as he speaks that he is creating a duty by saying these words—that this is what the words mean. But awareness of meaning is also something in the mind. So this suggestion does nothing to block our natural tendency to think that the distinguishing item is mental. I am grateful to an anonymous referee for pressing me to clarify these points.

⁷ Strictly speaking, for Hume it is not the love of one's children that motivates acts of caring for them, but the benevolence toward them that invariably accompanies such love (T 2.2.6, SBN 366–68).

⁸ It is easy to miss the link between the peculiar act which creates the obligation of promises and the motive to fulfill the promise. But Hume says that nonconventionalists, to establish their position, must prove "that there is a peculiar act of the mind, annexed to promises and that consequent to this act of the mind, there arises an inclination to perform, distinct from a sense of duty" (T 3.2.5.7, SBN 519). So the motive to fulfill the promise must result from the peculiar act of the mind itself. This is so because the motive alone is the object of approval, on the nonconventional view; the approval of performance and disapproval of nonperformance—the duty to perform—is merely derived from the obligatoriness of the motive to do so. Since the special mental act creates the duty to perform the promised action (which was not a duty before the promise was made), that act of the mind must either be or produce the motive to do what is promised as well.

need a *new* motive, one that was not present ahead of time. For these reasons, Hume proposes that there are only four candidates for the right sort of motive: the resolution to perform the action itself, the desire to do so, the willing of the action, and the willing of the (promise-based) obligation to perform the act (T 3.2.5.3; SBN 516). Although Hume does not spell this out, the other sorts of desires and aversions that he acknowledges in his moral psychology do not satisfy the prerequisites for a nonmoral motive of fidelity. They are not motives that both are generated anew at the moment of promising, and also are capable of motivating us to action of any type rather than just to those that serve a particular end. (For Hume the only motives we have in life are (1) the instincts, which always reside in the mind, (2) the desires, aversions, and related passions that have the agent's pleasure or pain as their goal and so are not sufficiently generic, and perhaps (3) resolutions and volitions, and perhaps also desires, to perform particular actions. The instincts are not suitable, and only certain motives of the other two kinds could qualify.) This leaves him with the four candidates I mentioned.

Hume points out how consideration of the actual conditions under which binding promises are made quickly eliminates three of his four candidates. First, a resolution to act is not sufficient to obligate me. I can resolve or decide to do A (say, attend a film), change my mind and resolve not to do A, refrain from A'ing, and violate no duty. The resolution to A, then, is not what is peculiar to promising. A desire to do the thing is not necessary for making a binding promise; we can make perfectly binding promises to do things we have no desire to do and even feel averse to doing. And willing to do the thing does not actually occur at the time we make the promise, Hume thinks, since we promise now to perform later.⁹ This leaves only "the *willing* of that *obligation*, which arises from the promise" (ibid.), in Hume's words, a mental act we can indeed perform at the moment when we promise. And this is indeed what we commonly think: that in promising we choose or consent to be obligated, and that is why we are. We are bound by our own will.

It is, according to Hume, a "manifest absurdity" (T 3.2.5.3, SBN 517) that there should be such a special act of the mind, the willing of an obligation to A, which would make it the case that I am obligated to A. On Hume's sentiment-based moral theory, for someone to have a moral obligation to perform some action is for his nonperformance of that action to elicit our disapproval. Thus, to create a new obligation to act in a way not formerly obligatory, as promising does, is to change our sentiments: to cause a new sentiment of disapproval to arise in anyone who contemplates the omission in a disinterested way. But we cannot change people's sentiments, making something disapproved that was formerly approved, simply by willing that it be disapproved.¹⁰ Therefore it would be impos-

⁹ Hume apparently has the view that a volition occurs only at the time of action. See T 1.1.4.5, SBN 12; T 2.3.1.2, SBN 399. Even without this doctrine, there are other reasons to reject the volition to act as a necessary condition for promise-based obligation. It would make lying promises (where one has no intention whatsoever to act) nonbinding.

¹⁰ Of course, I can make an action disagreeable that was otherwise agreeable by performing some other action. If I ordinarily enjoy playing the piano, I can will to make it unpleasant, and carry out my volition by pouring a disgusting liquid on the keys. I take it Hume is not denying this. Rather, he is denying that I can make a formerly approved action disapproved merely by willing that it be disapproved.

sible to create a new obligation by willing it, and it would be absurd to try. In a footnote Hume adds an argument to show that even if morality consisted in relations of objects discoverable by reason, it would be impossible to obligate oneself by the act of willing oneself to be under an obligation, for this would require creating a new relation of objects simply by willing that the objects shall stand in this new relation, which is also impossible. So not only sentimentalists such as Hume, but also moral rationalists, must give up the peculiar act of the mind. “A promise, therefore, is *naturally* something altogether unintelligible,” he concludes, “nor is there any act of the mind belonging to it” (ibid.).¹¹

But, Hume says, “suppos[e] the mind could fall into the absurdity of willing [an] obligation” (T 3.2.5.5, SBN 518). I suppose he means here that I could *try* to generate obligation by willing, just as I could try to move a matchbox on the other side of the room by willing, if I thought I had telekinetic powers. Such an act of willing to be obligated could not naturally *produce* any obligation, for the reasons already given. A new obligation involves the development of new sentiments, and one cannot create new sentiments of approval and disapproval simply by willing them to occur.

Thus, Hume claims, the nonconventionalist view of promising is refuted.

5. THE ORIGINS OF THE OBLIGATION OF PROMISES AND THE VIRTUE OF FIDELITY

The truth, according to Hume, is that “promises are human inventions, founded on the necessities and interests of society” (T 3.2.5.5, SBN 518). He gives a genetic account of the act of promising, the motive to conform, and the virtue of fidelity, which is supposed to explain the virtue as we find it. Prior to forming a convention of promising, our selfishness, confined generosity, and imperfect gratitude impose limits on our usefulness to one another; promises help us overcome these limits. Without the institution of promise, human beings could invent ownership of property, Hume thinks, and so stabilize possession and make possible very limited transfer of material goods by consent. They could simultaneously swap concrete goods that were within their view. But they would be unable to exchange services performed at different times or goods at a distance or goods described generically, so they could not profit from a personal surplus or over-

To will to be obligated would be, according to Hume’s metaethics, to will that a formerly approved or neutral action shall be disapproved by anyone who contemplates it from a universally-accessible, disinterested point of view. To obligate oneself by so willing would be to cause the action to be disapproved in this manner solely by willing it to be so.

¹¹ Hume’s argument that even the rationalist cannot appeal to the willing of an obligation claims that such a volition would involve a vicious circularity. Suppose an obligation is a relation discoverable by reason, and suppose a new (promissory) obligation is created by a volition to create a new obligation. Therefore a new relation is created by a (mere) volition to create a new relation. Either this is absurd, because willing to create a relation does not create one, or willing does create a new relation, in that the volition itself is one of the relata. Suppose the latter: that the new relation is created by a volition V which is a volition to create a relation one of whose relata is V itself. Therefore V is (in part) a willing of V. But a volition must have an object: one wills something, or else one is not willing at all. (The object of one volition could presumably be *another* volition, but this must be individuated from the first by having a distinct object.) If the object of V is V itself (that very volition), then V has no object. Therefore V (the volition to create a promise-based obligation) is impossible. This circle may not be vicious after all, but I will not investigate that here.

come a personal shortage of many goods. Without promises, such exchanges would depend upon one party performing first and the second performing in his turn out of gratitude alone; but that motive cannot generally be counted on in self-interested transactions, so people would not enter into transactions of these kinds. Moralists and politicians cannot alter human selfishness and ingratitude, but they can teach us better to satisfy our appetites “in an oblique and artificial manner” (T 3.2.5.9, SBN 521). This happens in two stages. First, people can easily recognize that additional kinds of mutual exchanges would serve their interests. All they have to do is express this interest to one another in order to motivate everyone to invent and to keep such agreements. They invent a form of words to distinguish these exchanges from the generous reciprocal acts of friendship and gratitude. When someone utters this form of words, he is understood to express a resolution to do the action in question, and he “subjects himself to the penalty of never being trusted again in case of failure” (T 3.2.5.10, SBN 522), Hume says. This “concert or convention” (ibid.) creates a new motive to act. One is motivated by self-interest to give the promising sign (in order to obtain the other party’s cooperation), and once one has given it, *self-interest* demands that one do what one promised to do so as to insure that people will exchange promises with one in the future. In making a promise, then, one in effect leaves one’s eligibility to be trusted as a security deposit that one forfeits on failure to perform.

Is a somewhat enlightened self-interest, then, the only motive for keeping one’s promise, once the practice of promising has been created? Some interpreters say that this is Hume’s view.¹² But a careful look at the text shows otherwise. This is indeed how things work in the early days of the practice, in small, face-to-face communities that offer few temptations to be a free-rider. But, Hume says, the sentiment of morals comes to play the same role in promise-keeping that it does in the development of honesty with respect to property (T 3.2.5.12, SBN 523). In that account, self-interested compliance with the conventional rules can be expected to break down in larger societies with more anonymous transactions (T 3.2.2.24, SBN 499). So we need a second stage. We need to develop a *virtue* of fidelity in people, to insure that the benefits of cooperation persist even when people know they can sometimes break a promise and still find others naïve enough to trade with them.

The next stage is that “a sentiment of morals concurs with interest and becomes a new obligation” (T 3.2.5.12, SBN 523). This sentiment of approval and disapproval arises in response to public interest, education, and the artifices of politicians, just as the like sentiment arises toward honesty with respect to property, Hume says.¹³ Presumably, then, it works this way in the case of fidelity to

¹² See Baier, *Progress*. Haakonssen and Gauthier, discussed in detail below, consider enlightened self-interest the principal motive of promise-keeping, at least in the *Treatise*, but acknowledge that a motive of duty (understood as self-hatred or guilt for failing to have an imagined but nonexistent further motive) sometimes serves as an auxiliary motive should self-interest fail.

¹³ Hume’s account of that process for honesty (at T 3.2.2.23–28, SBN 498–501; and T 3.2.6.11, SBN 533–34, with some elaboration at T 3.3.1.9, SBN 577) is not perfectly straightforward. It seems that we come to approve of a system of rules of ownership by spontaneous sympathy with the victims of dishonesty and perhaps with society as a whole. This general approval is reshaped by a new artifice of politicians and parents. These influential people manipulate the public, and children from their earliest infancy, to feel that every instance of honest behavior is honorable and worthy, and every instance

promises: sympathy with the victims of broken promises, and perhaps with society as a whole, generates approval for rules of promise and for general compliance with them, and disapproval for violations. (For Hume, sympathy is the capacity we have to share the emotions we believe to be experienced by others.) Parents, politicians, and others with influence over people's feelings then make use of this approval and disapproval. In a "second artifice," they use psychological manipulation (for good purposes) to produce a strong feeling of moral approval for every act of promise-keeping, and a strong feeling of moral disapproval for every act of promise-breaking, especially the agent's own, by associating the moral approval and disapproval with feelings of pride and shame.

As a result, these feelings become strong enough to *motivate* promise-keeping by subjects and children, at least some of the time, regardless of whether they see advantage in it.¹⁴ As Hume says about the virtue of honesty (T 3.2.2.25, SBN 500–501), "politicians . . . in order to govern men more easily . . . have endeavour'd to produce an esteem for justice, and an abhorrence of injustice." Since they do this in order to govern men more easily, and it "must have its effect," then clearly it influences motivation: the esteem and abhorrence give motivating power to the moral approval and disapproval that arose from sympathy. Similarly he says in the next paragraph that "parents . . . inculcate in their children, from their earliest infancy, the principles of probity, and teach them to regard the observance of those rules, by which society is maintain'd, as worthy and honorable, and their

of dishonesty is "base and infamous" (T 3.2.2.26, SBN 500–501), so that these moral attitudes become well entrenched and come to be directed to each and every relevant action. The approval and disapproval are so greatly strengthened by being linked with pride and humility, love and hatred, that they acquire motivating force. See Rachel Cohon, "Hume's Difficulty with the Virtue of Honesty," *Hume Studies* 23 (1997): 91–112. In the present text I spell out the parallel process for promise-keeping.

¹⁴ Hume is explicit that sympathy is the source of our moral *approval* of promise-keeping and *disapproval* of promise-breaking, so most interpreters would agree with the previous paragraph. I diverge from several of them in my further claim that the approval and disapproval so generated themselves become *motives* to keep promises distinct from self-interest, and, as I go on to argue, the sole virtuous motives at that. For Baier, for example (*Progress*, 242), the motive of compliance with promises is "enlightened self-interest, awareness of her own share" in the public goods of order and safe commerce that the cooperative scheme makes possible; "but when it comes to approving such motivated acts, her own and others', it is the public interest as such that becomes the relevant concern, not just any one person's share in it." And here sympathy comes into play.

The text does not make it completely obvious that Hume thinks moral approval itself becomes a motive of promise-keeping, although I go on to argue that the pieces fit together to entail this. The further obstacle to so reading him, of course, is that he has already argued that a moral motive is impossible. That difficulty is tackled next.

One interpreter who reads this text as I do, as saying that moral approval and disapproval strengthened by "propaganda" provide a principal motive of just action, is Páll S. Árdal, *Passion and Value in Hume's Treatise* (Edinburgh: Edinburgh University Press, 1966), 186–87 (where he sees the resultant problem of circularity as basically insoluble). His "Convention and Value" (in *David Hume: Bicentenary Papers*, ed. G. P. Morice [Austin: University of Texas Press, 1977], 64), also makes reference to the motive of duty, enhanced by education, as one of the motives of the just man.

J. L. Mackie, in *Hume's Moral Theory* (London: Routledge, 1980), 100–101, thinks moral sentiment plays some role in motivation. Self-interest is the principal motive, but our interests will be well served only if people develop stable dispositions to keep promises even where it does not serve their interest to do so. We fulfill the occasional disadvantageous promise because the human psyche is inflexible: "dispositions . . . cannot be quickly switched on and off." (So, no further motive would seem to be involved.) But the development of the full disposition to keep promises depends upon moral feeling: "the very fact that I have promised to do something should be felt to give me a reason for doing it."

violation as base and infamous.” This *moral motivating sentiment* is one that observers will in turn approve, because of the psychological mechanisms involved in the generation of Humean approval.¹⁵ So moral approval and disapproval have become motives which in their turn are subject to moral approval. A motivating sentiment that we approve is a virtue, according to Hume’s account of what qualifies as a virtue. Thus we create, by means of two artifices, the virtue of fidelity.

6. WHY WE STILL “FEIGN” THAT A
PROMISOR WILLS AN OBLIGATION

There seems to be no further reason to appeal to any special act of the mind. Yet Hume cautions that there remain difficulties “in supposing a moral obligation to attend promises” (T 3.2.5.12, SBN 523). To elude these difficulties “we *feign* a new act of the mind, which we call the willing an obligation” (ibid.), and we take the moral obligation to depend upon this act of mind; but of course it has been shown that no act of the sort is possible. Our practice of promising, seen as imposing a new obligation on ourselves by an act of will, resembles transubstantiation and the giving of holy orders, “mysterious and incomprehensible operations” (T 3.2.5.14, SBN 524) in which a certain form of words together with a certain intention change a thing’s very nature.¹⁶ Hume clearly means that in our practice of promising, as we standardly interpret it, we invoke a fictitious mental power as part of a social scheme of deception.

But why? Why do we not simply recognize that the moral obligation of promises arises from an extremely useful social convention, and dispense with any pretense that we obligate ourselves by a special act of the mind?

Other interpreters have offered answers that shed light but do not fully fit the text. For example, Knud Haakonssen says we feign a special mental act to conceal from ourselves the fact that the only available motive for fulfilling promises is self-interest.¹⁷ We do not approve this real but morally neutral motive, and so fidelity is in fact no virtue. However, the promise-keeping acts done from this motive are obviously highly beneficial. As a result we imagine there must be a motive in others directing them toward these good consequences (although we never find it in ourselves), and we naturally come to approve the imaginary motive, creating the illusion that the actions are the products of a virtue. We consequently come to

¹⁵ According to Hume, the moral sentiment is generated as follows. We observe or imagine the effects of a character trait on those individuals on whom it has an impact for good or ill (the person herself and her intimates, usually). The psychological mechanism of communication of emotion, which Hume calls sympathy, transfers their pleasure or pain to us, and apparently associates that pleasure or pain with the trait. This resulting feeling on our part is our moral sentiment of approval or disapproval of the trait. Thus, if some motivating sentiment in the faithful person has a beneficial impact on others in society (or even can be expected to have one), observers will come to approve that sentiment as a matter of course.

¹⁶ Our practice of promising, however, involves even more “contradictions” than do the two religious procedures, because unlike transubstantiation and ordination, which (Hume thinks) are useless, the practice of promising has been warped in various ways to serve the good of society.

¹⁷ Knud Haakonssen, *The Science of a Legislator: The Natural Jurisprudence of David Hume and Adam Smith* (Cambridge, U.K.: Cambridge University Press, 1981), ch. 2 (esp. Part 8). Haakonssen’s effort to account for the peculiar mental act is the most perceptive that I have seen to date.

hate ourselves for lacking this approved (though imaginary) motive; and this self-hatred becomes a motive of duty for keeping our promises. We are indeed moved by this to keep promises, but since there is no natural sentiment on which duty may be based, it is not in fact our duty to do so.

There are three textual difficulties with this. First, Hume is adamant that fidelity is no less a virtue for being artificial, and that the obligation of promise is genuine. Second, if I am right, self-interest is not the sole motive to promise-keeping, and not the one characteristic of the virtue; there is another (albeit moral) motive that *is* approved. Third, the problem of motivation arises for the artificial virtues of justice (honesty with respect to property) and allegiance to government as well as for fidelity. To make the interpretation coherent, then, Haakonssen applies it to all the artificial virtues, postulating a feigned virtuous motive in each case. But it is only with respect to fidelity that Hume ever mentions any feigned mental act.

We have now set out the tools for a better answer. As I interpret him, Hume thinks that the *common-sense concept* of a virtue, any virtue, is the concept of a trait of character consisting of a particular laudable, non-moral, motivational disposition. That is, he thinks that what we ordinarily have in mind when we think of a virtue is the natural conception of virtue. We all have this, and we apply it (mistakenly) to all the virtues that we recognize. We have this conception because of the way our psyches function. First, we are not naturally impartial creatures; we have a natural bias in favor of ourselves and those near and dear to us. Partiality to ourselves and those dear to us is part of “the original frame of our mind.” Secondly, this partiality has been taken up into our conception of virtue and vice. We do not think we *should* have impartial sentiments.

This partiality . . . must not only have an influence on our behaviour and conduct in society, but even on our *ideas of vice and virtue*, so as to make us regard any remarkable transgression of such a degree of partiality, either by too great an enlargement, or contraction of the affections, as vicious and immoral. . . . [O]ur *natural uncultivated ideas of morality*, instead of providing a remedy for the partiality of our affections, do rather conform themselves to that partiality. (T 3.2.2.8, SBN 488–89, emphasis mine)

The naturally virtuous sentiments such as compassion and love of one’s children exhibit a partiality of which we approve, and which is plainly natural and suitable to such creatures as we are. Consequently this shapes our very concept of a virtue. In order that we may cooperate for mutual gain with those outside our circle of friends, our “judgment and understanding” supplies a remedy for what is “irregular and incommodious in the affections” (T 3.2.2.9, SBN 489), in the form of a trait whose operation is *impartial*. We need that impartiality in promise-keeping if a mere party to a contract is to have any security that I will do my share in the face of the countervailing interests of my intimates and myself. But we still think it virtuous in me to *care* more about my intimates than my trading partner, to sympathize more with them and be more readily moved to act by the feelings they communicate to me. The partiality of sentiment continues to elicit our approval. A thoroughly impartial temperament would never look as attractive. We retain the natural concept of virtue. Thus we must employ a small fantasy in order to squeeze fidelity under it.

According to the natural conception of virtue, which is the one we have, fidelity to promises is not a virtue, and the obligation of promises is not a duty. For fidelity is not a non-moral motivating sentiment whose deficiency we disapprove. But we think of fidelity as a virtue anyway. The situation is similar with the other artificial virtues, which also lack non-moral motivating sentiments. In their case we do not notice the mismatch between our natural conception of virtue and the traits we actually approve. However, each act of promising draws our attention to the formation of new obligation, and so presents a challenge to the natural conception of virtue and of the obligation to which it gives rise. The obligation of promise is not like the steady obligations to refrain from the property of others or to obey our rulers, which only puzzle us upon philosophical reflection. It comes into conflict with our assumptions about virtue in daily practice, as we make promises and enter into contracts. Thus we must hit upon a way, within the framework of our preconceived notions of virtue and duty, to think about this sudden appearance of obligation. We do this by accepting a fiction and not noticing the contradictions to which it gives rise. We pretend that the obligation of a promise is created by a peculiar act of the mind, the willing of an obligation, and that this is the approved motive of promise-keeping. This keeps all the features of the natural conceptions of virtue and duty in place. But it leaves us burdened with a fiction.¹⁸

So if I am right, Hume thinks that the artificial virtues are virtues in the sense in which an artificial leg is a leg. On the natural conception of virtue, they are not virtues, just as an artificial leg is not a natural leg. But according to Hume's *own* account of virtue, any motivating sentiment that elicits approval is a virtue. He does not limit this official description of virtue to the motivating sentiments that are non-moral. And in the case of fidelity, there is eventually a motivating sentiment that elicits our approval of promise-keeping acts and our disapproval of their omission, namely, the enhanced moral sentiment. Hence fidelity *is* a virtue on Hume's definition. It is functionally a virtue, and that is what counts for Hume.

Do we have to retain our pre-reflective, natural conception of virtue, for Hume? He does not say, but the answer seems to be no. We can be faithful to promises without it, since our actual motivation to fulfill promises does not depend for its genesis on this belief. We can become more philosophically consistent on this subject without undermining our fidelity. But it is a strain to give up this way of thinking of virtue, since it fits so comfortably with our emotional make-up. In relinquishing it, we make it clearer than ever that fidelity is a virtue, but we have to acknowledge that it is a prosthetic one.

This interpretation allows us to explain Hume's remarks about the peculiar act of the mind without postulating that one or both claims is a symptom of Hume's confusion or even of general failure in Hume's theory of promises, as other interpretations do. J. L. Mackie and Francis Snare, for example, each offer instructive explanations of why a mental-act account of promise-based obligation might be a

¹⁸ An anonymous referee wonders how common sense could possibly suppose that the peculiar mental act which occurs at the time of promising could motivate promise-keeping much later (even years later). Presumably common sense confusedly thinks of the willing of an obligation as a motive like those that constitute the natural virtues such as parental love and humanity—motives that persist for long periods and cause actions at various times.

natural alternative to a conventional explanation. Both, however, take Hume to be confused or mistaken in seeing reason to claim that, even given the convention, we nonetheless always feign such an act when we promise.¹⁹

David Gauthier's finely-honed analysis offers a more complex but even more dire diagnosis.²⁰ Gauthier construes Hume's treatment of fidelity and the other artificial virtues differently in the *Treatise* and in the *Enquiry concerning the Principles of Morals*, but in each interpretation he leaves Hume with error or failure.

Gauthier agrees with Haakonssen that interest is the actual motive of conformity to the rules of promise; but on Gauthier's view this enlightened self-interest, redirected by the convention so as to motivate promise-keeping acts instead of acts of selfish acquisition, is a virtuous motive, one of which we approve. It is artificial in the sense that it takes the form it does only in response to convention, but natural in that it is non-moral. This motive is theoretically sufficient to account for the virtue of fidelity, for Gauthier, and needs only to be reinforced by moral motivation because in practice, in complex societies, we too easily overlook our own interest.²¹ Gauthier claims that in the *Treatise* Hume does not doubt that under conventional circumstances all individual acts of promise-keeping really are to the agent's advantage. Agents who recognize this, then, will be reliably moved; and the resulting benefits to society will cause this motive to be approved.²²

This reading of the *Treatise* is very resourceful, but leaves Hume with no reason to say that, after the moralization of the conventional practice of fidelity, we find ourselves having "difficulties . . . in supposing a moral obligation to attend promises," and so we feign the willing of an obligation (T 3.2.5.12, SBN 523). Perhaps Gauthier regards Hume's remarks about this, and his comparison with transub-

¹⁹ J. L. Mackie, "Hume's Moral Theory," 96–104, esp. 103; Francis Snare, *Morals, Motivation and Convention* (Cambridge, UK: Cambridge University Press, 1991), Part II, esp. chs. 7 and 9 (see, e.g., 265–69).

²⁰ David Gauthier, "Artificial Virtues and the Sensible Knave" ["Artificial Virtues"], *Hume Studies* 18 (1992): 401–27.

²¹ So for Gauthier, in the *Treatise* Hume successfully finds the requisite non-moral motive of promise-keeping that he seeks. There is much to be said for this reading. However, it must extrapolate somewhat from the text, since Hume never says that interest, even redirected by the convention of promises, is a motive of which the moral sentiment approves, and he explicitly denies that we are motivated either by interest or concern for society when we conform to the requirements of the artificial virtues (T 3.2.1.11–140, SBN 480–82). (Hume does of course say that we approve of justice and fidelity. Gauthier construes this as approval of redirected interest by claiming that Hume identifies redirected interest with the sense of justice, so that to approve of one is to approve of the other. But it strains the text to understand the "sense of justice" to be a non-moral motive.) In the present paper, I argue instead that Hume never finds any non-moral motive to the keeping of promises in established societies such as ours, and that he appeals to the redirection of interest only to explain a) the motive to create the convention of promising and b) the *initial* incentive to keep promises in a small, new community.

²² To some extent, and in different ways, both Gauthier and Haakonssen are indebted to Marcia Baron's account of Hume's artificial virtues as involving a "noble lie" we all must believe in order to possess fidelity and honesty: the lie that every act of compliance serves the agent's own interest. See "Hume's Noble Lie: an Account of his Artificial Virtues," *Canadian Journal of Philosophy* 12 (1982): 539–55, repr. in *Hume: Moral and Political Philosophy*, ed. Rachel Cohen (Aldershot: Ashgate/Dartmouth, 2001). Gauthier thinks that the Hume of the *Treatise* does not see it as a lie—he believes it himself, but the Hume of the moral *Enquiry* realizes its falsehood.

stantiation, as a slip.²³ This reading has a further important weakness. It depends upon the claim that throughout the *Treatise* Hume held the naïve belief that all individual acts of promise-keeping (within society) redound to the net advantage of the agent, and furthermore that this enabled Hume to identify a reliable motive of promise-keeping in human nature. The textual evidence that Hume held the naïve belief is equivocal.²⁴ But even had he believed this, Hume explicitly recognized that human nature is such that (rightly or wrongly) we will not be able to foresee long-range net benefit in every case, and even when we do, we will not always be moved by it, because we are short-sighted, as Gauthier notes.²⁵ Thus, given human nature, in large societies we can be expected to lack a reliable non-moral motive to keep promises. Whatever Hume thought about the true wages of virtue, he saw the problem of motivation to promise-keeping as theoretical as well as practical, since it is rooted in human nature itself.

In the *Enquiry concerning the Principles of Morals*, according to Gauthier, Hume first confronts the fact that some acts of promise-keeping do not serve the agent's interest (even including her interest in preserving the practice), because in some situations breaking a promise would be highly profitable without posing any significant threat to the social confederacy. In people such as the sensible knave who reflect on this, "the initial self-interested motive fails to survive critical reflection."²⁶ Here Gauthier proposes that Hume's only available response is a developmental

²³ On Gauthier's interpretation of the *Treatise*, the peculiar act of the mind may be mistakenly postulated by nonconventionalists (although he does not say why they would postulate the willing of an obligation rather than just some non-moral motive or other), and perhaps an illusion of it resides in the popular imagination (though he does not say why it would). But Hume himself "has no need to burden his own account with such an act" (Gauthier, "Artificial Virtues," 416). However, alas for Hume, he does it anyway.

²⁴ For example, Gauthier quotes a passage in which Hume says that an act of justice may be contrary to the agent's interest or the public interest: "Tho' in one instance the public be a sufferer, this momentary ill is amply compensated by the steady prosecution of the rule, and by the peace and order, which it establishes in society. And even every individual person must find himself a gainer, on ballancing the account" (T 3.2.2.22, SBN 497). This certainly seems to support Gauthier's reading. But note that Hume does not say the individual is a gainer (at the end of his life, presumably) by the very act that produces a "momentary ill," provided merely it is joined by the just acts of others. Hume may simply mean that while a single act may be harmful to the agent and to society (as in returning a fortune to a seditious bigot), perhaps forever reducing the agent's total welfare, the condition of general conformity of *all* to the scheme of justice, as opposed to the condition where there is no rule-governed society, is a net benefit to the individual. For Hume goes on to say, in the sentence about the individual being a net gainer, "since, without justice, society must immediately dissolve, and every one must fall into that savage and solitary condition, which is infinitely worse than the worst situation that can possibly be suppos'd in society."

²⁵ Gauthier draws support from Hume's later discussion of the need to create government. There Hume sounds as if he blames most common violations of the laws of justice (here including the rules of promise-keeping) on the irresistible temptation to choose near rewards at the cost of greater future benefits. Thus Gauthier quotes this passage (T 3.2.7.3, SBN 535): "The consequences of every breach of equity seem to lie very remote, and are not able to counterballance any immediate advantage, that may be reap'd from it. They are, however, nonetheless real for being remote." So Hume seems to think conformity on every occasion has real net benefits to the agent. However, since the tendency to prefer the near to the remote is one of the "natural principles and propensities of human nature" (T 3.2.7.4, SBN 536), that cannot be changed but only circumvented by further artifice, it is clear that redirected interest does not have the power to motivate compliance in any society where the benefits of compliance are sufficiently long-range.

²⁶ Gauthier, "Artificial Virtues," 420.

version of the error theory that Haakonssen finds in the *Treatise*. Redirected interest cannot move us to keep promises under some circumstances; therefore, in those cases we have no natural motive of fidelity after all. We continue to approve of general conformity to the rules of promise even so, because of its social benefits. So we then imagine a different non-moral motive on the part of others, the (nonexistent) willing of an obligation. We direct our approval to that and hate ourselves for lacking it, and this self-hatred becomes a sense of duty that motivates compliance. This motive of duty can come into being only if we accept the fiction. However, this Humean strategy cannot succeed, Gauthier argues. The fiction does not stand up to reflection, for it is impossible to will an obligation. When we realize this, Gauthier says, we will see that fidelity and the other artificial virtues are not our duty, and therefore “human society, which depends on these dispositions, lacks any moral foundation.”²⁷ Hume’s theory of promise collapses under this pressure.²⁸

The present interpretation leaves Hume in a far stronger position. It does not depend on the general, enduring efficacy of a self-interested motive to promise-keeping. It identifies a persistent (though moral) motive for promise-keeping even where the faithful person does not see such behavior as advantageous. It locates a morally-approved and real motive, so human society does not lack moral foundation. It leaves Hume with explanations of the obligation of promises and of the motivation for promise-keeping that are not undermined when people realize that it is not to their net advantage to fulfill certain promises and that there is no mental act of willing an obligation. And it finds a coherent reason within Hume’s understanding of our common-sense conception of virtue for us still to feign a special mental act after the convention is in place. We do not feign the peculiar act in order to preserve our motivation to fulfill promises, but rather in order to preserve our assumptions about the structure of virtue.

7. TWO QUICK LESSONS FOR PRESENT-DAY VIRTUE ETHICS

I do not think it follows from Hume’s discussion of fidelity that present-day virtue theorists must regard fidelity as conventional rather than natural. The lessons I wish to draw are more subtle than this, but follow more directly from Hume’s observations about fidelity.

Some present-day virtue theorists are committed to the primacy of character over action. This apparently entails that we must have a prior, independent concept of what it is to have a certain virtue (an admirable set of psychological dispositions), and then use that to determine which actions are required of us. (John McDowell and Gary Watson say things like this.²⁹) The Humean version of this

²⁷ *Ibid.*, 422.

²⁸ I have not been able to ascertain why, on Gauthier’s reading, feigning an obligation is introduced in the *Treatise* to solve a problem that Hume is unaware of until the moral *Enquiry*. If instead what Gauthier means to do here is use some materials from the *Treatise* to attempt a solution that Hume himself never offers, then he leaves us with no account of why the feigning is mentioned in the *Treatise*.

²⁹ See G. Watson, “On The Primacy of Character,” in *Identity, Character and Morality: Essays in Moral Psychology*, eds. O. Flanagan and A. O. Rorty (Cambridge, MA: MIT Press, 1990), 449–69; and J. McDowell, “Virtue and Reason,” *Monist* 62 (1979), 331–50.

was the pivotal thesis in the natural conception of virtue that made fidelity (and the rest of the artificial virtues) impossible. If I have interpreted him correctly, Hume ultimately treats this thesis as an over-general common-sense assumption. On his own considered, philosophical view he weakens it in order to account for the actual approved motive of promise-keeping, which is moral. He allows that morally required actions are not all defined by their motivating sentiments; acts of promise-keeping are initially defined by a social convention, and only later become the object of a characteristic motive, a moral one. This may offer a lesson for present-day virtue theorists. The person of her word in fact fulfills promises because she thinks it obligatory to fulfill promises. Hume seems to be right about this. An accurate account of the character of the person of her word, then, makes essential reference to obligatory action. Hence the motives and feelings that make up that character trait cannot be our sole criterion for obligatoriness of actions. When it comes to fidelity and the other virtues that Hume calls artificial, character cannot determine *what* we should do; that must be settled in another way. It does not follow, of course, that it must be settled by conventional rules. But it must be settled by something other than agents' attitudes, at least initially. This runs counter to the doctrine of the primacy of character.³⁰

The second lesson for virtue ethics is the *nature* of the true motive of trustworthy actions. Present-day virtue theorists tend to denigrate the motive of duty, to some extent understandably. If I visit a friend in the hospital, in Michael Stocker's famous example, for this to manifest the virtue of friendship I must do it out of affection for my friend and a desire to cheer him up.³¹ If a husband rescues his wife from mortal danger in preference to a stranger, in Bernard Williams's famous example, for this to be a good action, one that manifests the husband's virtue of love, it should be motivated by his love of his wife and sensitivity to what she means to him.³² If I visit my hospitalized friend merely because it is my duty, I am not a thoroughly good friend. If the husband rescues his wife in preference to a stranger in part because it is permissible (not contrary to duty) for husbands to rescue their wives first, this is "one thought too many," in Williams's phrase, and the wife will have grounds for disappointment. These actions fail to manifest the relevant virtues because they are not done solely from love or concern, but rather (in whole or in part) from duty, revealing a deficiency in the agent's attachment. Many virtue theorists, then, regard the motive of duty as second-best, as a kind of

³⁰ Some virtue theorists will be perfectly happy with this result. See Rosalind Hursthouse, "Applying Virtue Ethics," in *Virtues and Reasons: Philippa Foot and Moral Theory*, eds. R. Hursthouse, G. Lawrence and W.S. Quinn (Oxford: Oxford University Press, 1995), 57–75. But note important differences in her view in *On Virtue Ethics* (Oxford: Oxford University Press, 1999).

³¹ Michael Stocker, "The Schizophrenia of Modern Moral Theories" ["Schizophrenia"], *Journal of Philosophy* 73 (1976): 453–66.

³² Bernard Williams, "Persons, Character and Morality," in his *Moral Luck* (Cambridge: Cambridge University Press, 1981), 18. (Williams's actual example is more complex than this, and is aimed to make a subtler point, in criticism of Kantian ethics. He thinks that should the man reason that saving his wife in preference to a stranger is permissible and not unfair to the stranger, this would be "one thought too many" and incompatible with genuine love, which would motivate him to save her solely because it is she who needs saving.)

fallback motive in case one's virtue fails.³³ But if I am a person with the virtue of fidelity, what moves me to fulfill an inconvenient promise is my feeling or thought that I should do it, that it would be wrong (indeed, a betrayal) not to, and I don't want to betray my word. That is the motive of duty, and here it is the right motive. For fidelity to promises, love is not enough. But neither is self-interest or desire for the public good. We need the motive of duty or we do not get the virtue of fidelity. Consequently, we had better not be too dismissive of the motive of duty; it needs an important place in our virtue ethics.³⁴

³³ Stocker says, "Duty seems relevant in our relations with our loved ones and friends, only when our love, friendship, and affection lapse." Stocker, "Schizophrenia," n. 8.

³⁴ I am grateful to audiences at the University at Albany, S.U.N.Y., Virginia Polytechnic Institute, Washington University, and Union College for helpful comments on earlier versions of this paper.