



**Artificial Virtues and the Equally Sensible Non-Knaves: A Response to Gauthier**

Annette C. Baier

*Hume Studies* Volume XVIII, Number 2 (November, 1992) 429-440.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at <http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' 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 HUME STUDIES archive only for your personal, non-commercial use.

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

For more information on HUME STUDIES contact [humestudies-info@humesociety.org](mailto:humestudies-info@humesociety.org)

<http://www.humesociety.org/hs/>

# Artificial Virtues and the Equally Sensible Non-Knaves: A Response to Gauthier

Annette C. Baier

Gauthier's splendidly dialectical paper<sup>1</sup> first sets out Hume's official *Treatise* account of how each person has a self-interested motive to curb her natural but socially troublesome self-interest, by agreeing to the adoption of the artifices of private property rights, transfer by consent, and promise (provided others are also agreeing to adopt them), and how the sympathy-dependent moral sentiment approves of such a redirection of "the interested passion." Enlightened self-interest gives us the natural obligation to justice; the fact that it is morally approved turns it into a moral obligation. Self-hatred for any breaches of the moral obligation may come in to bolster the self-interested motivation, and this will be helped along by the extra artifices employed by educators and politicians. So far so good. But now enters the sensible knave, who queries the initial assumption that each will in fact gain more by strict conformity to the rules of the Humean social artifices than by occasional carefully calculated and undetected breaches of them. If the knave is right, then the *Treatise* story of how reasonable self-interested people can get themselves to conscientiously respect property rights (including rights of barter), and to keep promises, is in trouble. Only by an erroneous calculation of their interest could they end with the motivation which Hume attributes to them, and finds to be morally approvable. Gauthier argues that there are indications even in the *Treatise* that Hume is aware of the flawed nature of his official account, since he attributes "sophistry" to some attempted justifications of honest behaviour, and attributes "feigned acts" of impossible willing to some conscientious promise-keepers. Gauthier finds Hume to have, in the *Treatise*, a second, less official story of what it takes to be motivated to be just—that it takes error, and letting oneself be taken in by feignings and sophistry. In the *Enquiry Concerning the Principles of Morals* this submerged account surfaces in the sensible knave's challenge and Hume's admission that his own answer will not be satisfactory to the knave. The charms of official Humean justice, Gauthier concludes, are spurious, since the "anatomy" of the account of the rational motivation to be just is known to be unsound. "The sensible knave's message is that human society ... lacks any moral foundation."

This is a sombre tale. One of its unwelcome implications is that Hume himself, really believing that the knave is right, but unwilling to be seen to have the bad grace to “deliver a theory, however true which ... leads to a practice dangerous and pernicious,” is simply trying to “sink” the knave’s truth in oblivion by creating a smoke-screen of insincere rhetoric about the greater delights of consciousness of integrity over the knave’s satisfaction in his ill-gotten gains. So there really are two sombre tales—one, attributed to Hume, about justice, and one about Hume’s duplicity (or is it mere cynical frivolity?). Gauthier crafts both stories beautifully, directing our attention to bits of Hume’s text in just the right order for his dramatic purposes. But is he giving us truth or fiction?

I completely agree with Gauthier’s version of Hume’s official *Treatise* story, that the moral obligation to justice supervenes on the “natural obligation” for which Hume believes the natural motivation to be collectively redirected self- (or family) interest.<sup>2</sup> I agree that there are one or two passages, especially in the *Treatise* section, “Of the obligation of promises,” that seem to jar with this account, and which much somehow be explained. I agree that Hume believes that his reply to the knave is, from the knave’s point of view, unsatisfactory.

Now for some disagreements. First, about Hume’s reply to the knave.<sup>3</sup> From the viewpoint of the virtuous dues-paying member of “the party of humankind,” Hume’s is a perfectly satisfactory reply. It will be plain truth, for a person with the other Humean virtues (including equity, which is listed as a *natural* virtue),<sup>4</sup> that justice will bring as great or greater good than would judicious injustice. Even if it would be “simply preposterous” to believe that the social order is dangerously threatened by any and every breach of the rules of justice, it is not preposterous to believe that the rule breaker does sacrifice “consciousness of integrity” by breaking rules that she is counting on others to keep. Gauthier suggests that it is the duped or self-deluded non-knave who lacks integrity—he should be able to see that the self-interested foundation for the moral obligation to justice is lacking, so that the obligation is a fake one. I take it that “integrity” is being taken by Gauthier here to mean some sort of wholeness, lack of division, or self-deceit in the soul. And if it is the conformer to the rules who deceives himself, then he cannot have consciousness of integrity. Fair enough. But must he deceive himself?

The Humean challenge with which Gauthier believes the knave can rest her case is that it be shown that the moral duties recommended “are also in the true interest of each individual.” Notice that the phrase is “true interest.” Whose perception of interest is “true,” the knave’s, based as it is on valuing material gains along with consciousness of her superior craftiness and successful duplicity, or the non-knave’s, who

dismisses the material rewards of judicious injustice as "worthless toys and gewgaws" (E 283), and has no wish to take pride in having hoodwinked his cheated fellows? Well, "worthless" is no doubt a bit false. The non-knave may protest too much here. (Wealth is not worthless.) But opting for modest material gains along with consciousness that one's means of getting them will bear public inspection, so that one has no need for concealment, in preference to the wealthy knave's precarious and deceit-dependent satisfaction, does not seem an irrational preference. We may be at a stand-off when we try to decide who, if either, has erred in calculation of her or his true interest, the greed-driven knave, busily accumulating by successful fraud, and sacrificing candour and open dealing for the gains got by deceit and secrecy, or the non-knave who has settled for his different, more sociable, and less secrecy-dependent satisfactions. In a society where both successful fraud and some modest returns for honest dealing are real possibilities, then where our true interest in fact lies will surely depend upon what our ruling passions in fact are. From the knave's point of view, and given her psychology, it may be prudent to continue with knavery. (She does seem mistaken, however, if she thinks that her fidelity to promises will go undetected, except in very unusual circumstances.)<sup>5</sup> From the non-knave's different point of view, with his different psychology, prudence and true interest dictate continued abhorrence of secret dishonesty and other forms of knavery. From the moral point of view, even the knave (if indeed she is able to adopt this viewpoint) will approve and recommend non-knavish prudence over knavish prudence, since she prefers others to have this socially useful trait; indeed she counts on most of them having it. For Hume, the moral point of view is the only general point of view we can appeal to here. There is no Archimedean "rational point of view" from which a judgement could be made between the knave's version and the non-knave's version of self-interest. So Hume's reply to the knave is exactly what it should be: "From our point of view, you are the greater dupe, but you have unfitted yourself for understanding how the rest of us can make this judgement, and find it in our true interest to combine justice (ourselves obeying the rules of co-operative schemes that we count on others obeying) with the other virtues, equity, prudence, kindness, and the rest. For us, there would be an awkward and morally unacceptable split, doubleness, or lack of integrity in our lives if we were to try simultaneously to cultivate generosity along with successful fraud, fidelity to friends along with infidelity to promises, equity along with injustice." (These of course are mostly *not* Hume's words, but a free paraphrase of them.)

So much for my response to Gauthier's sombre story of Hume the cynic. Now for my response to Gauthier's other sombre story,

concerning the baselessness of the moral obligation to justice, the supposed error in the official *Treatise* story. I have already hinted at some of my disagreements here, by pointing out that, for Hume, "equity" is not, or not only, an artificial virtue, that "true interest" can take different forms for different people, and that from a moral point of view some of these forms are preferable to others. Not merely is true interest, in the *Enquiry's* sense, relative to the values of the person whose interest it is, but there will also be historical shifts in what values the average person will be likely to have, and which of these will then be morally preferred. In the *Treatise* story, there is a shift from the form that the "interested passion" takes in those who invent property rights to the form it takes in the promise-inventors and the magistrate-inventors. "The interested passion," which is redirected by the convention of property, is glossed by Hume as, "This avidity alone, of acquiring goods and possessions for ourselves and our nearest friends" (T 491-92), where the goods in question have, a few pages earlier, been described as goods that "may be transferr'd without suffering any loss or alteration" (T 487-88). These clearly are the main sort of goods that the sensible knave wants. But by the time that government is invented, the interested passion seems to include concern for those non-transferable goods, "life and limbs" (T 540), and for public goods such as "order in society" (T 535). The "contradiction to their known interest" which those who see the need for government aim to prevent is not merely action leading to their own impoverishment, but also that leading to "dangerous and uncertain" commerce between people, and a violent life-endangering and anxiety-generating climate of life (T 535). The conditions of life brought about by the use and the limitations of the pre-government artifices, especially contract, have apparently altered people's perception of what sort of goods it has become most important for them to try to increase by co-operative means. Security of life, equity, and public order have become the scarce and valued goods.

It is ironical that Gauthier, whose own account is so nicely dialectical, should not notice, or should choose to ignore, the dialectical social development that Hume describes in the *Treatise*. There we find not one static thing, "self-interest," redirected time after time in four successive artifices. As I read Hume's story,<sup>6</sup> there are four successive versions of the interested passion (the passion for increase) which are redirected, and the change in the form the passion takes can be seen to be due to the changes in what the salient scarce goods are, at each point. From concern for security of possession of external transferable goods, at the time of the establishment of the first artifice, we move to concern for security in transfers, in the establishment of the second, on to concern for security in non-simultaneous transfers, in the third. In the

*Treatise's* last artificial redirection of natural passions, that of setting up government, we have moved on to a much more generalized concern about all sorts of human goods, from life and limb and security of possession to non-dangerous commerce and peace of mind.

Early on in setting up the problem he is concerned with, Gauthier directs our attention to passages from the *Treatise* section, "Of the obligation of promises." It is therefore important to see what Hume takes to be the precise form of the interested passion that is redirected by that particular artifice, and to see if those passages really do hint at the "error theory" that Gauthier is claiming to be present as a subtext in the *Treatise*. The sort of people whom Hume postulates as the ones there would be, if property and barter, but not yet contract, had been established, are people accustomed to mutually advantageous simultaneous exchange, who are looking for a way to extend this safely to non-simultaneous exchange, and to exchange of measured goods. They care about getting (and getting in the future) exactly what has been agreed to be what is due to them, in return for what they are now giving up. Exactitude has come to matter, and it is words, weights and measures that enable it to matter. The device that is invented is itself a form of words. What is now seen as in everyone's interest is that people "be true to their word" (T 522). What more natural version of the interested passion is redirected into this concern with people's truth to their words? Hume mentions lack of gratitude, and lack of friendly feelings, along with the need for the help of others, and concern for the future. His famous harvest example makes it plausible enough to say that the driving passion here is a wish to fix the future, to have some security for oneself against future disaster or injury. It is both goods and services for tomorrow that these apparently selfish people want to be sure of. They want what Hobbes termed "power," a present means to a future good. The "natural" way to satisfy this wish is to accumulate now—to con one's neighbour into helping one get one's harvest in, so that one will be fed for the winter. The redirection of this natural selfish desire for a secure future is the desire for security for an even longer future, promised by the reputation of being a reliable promise-keeper, one who can therefore count on her neighbour's return help next year as well as this year.

Reputation is a non-transferable good, and this is what people must see to be in their interest (in addition to whatever else is), if the artifice of promise as Hume describes it is to work. So the "interest" seen at this stage in Hume's dialectic has been extended from transferable material possessions to non-transferable immaterial ones, but not yet, as when government is later invented, to public goods such as social order and a climate of fairly secure public confidence that

official law-enforcers are trying to catch and punish thieves and those guilty of force, breach of contract, and fraud.

But how does this help with the problem passage at T 523? Why should there be any temptation to “feign” the private willing of an obligation, a mental act that Hume has previously shown to be impossible? It happens, Hume suggests, to shore up the belief that an “obligation ... attend[s] promises” (T 523). But what difficulties were there to be “surmount[ed] or elude[d]” by this suspect bit of fancy mental footwork? There seems to have been no trouble in seeing that there is an interested obligation to keep promises—as long as reputation is valued and future contractual exchanges are desired, each person will have and perceive this obligation. It is the *moral* obligation that seems to raise some difficulties. Why? Well, Hume may just be of the opinion that this device is not, from a moral viewpoint, as obviously desirable as were the previous ones. He derides promise in ways in which he does not deride the first or the fourth artifice, likening it to transubstantiation. It certainly is an accepted custom, but so, in Hume’s time, was slavery, which could be, and was, seen to arise from the slave’s alienation to the slave trader of his alienable right to his liberty, in return for life—that is, it was seen to arise from a special case of reciprocal exchange. Both promise and slavery could count as “mere artificial contrivances for the convenience ... of society” (T 525). But the moral sentiment does not blindly approve of any and every scheme of rights and duties that the socially inventive human mind has come up with, nor of all schemes that have some social utility. I believe that Hume was of two minds about just how valuable contract has proved to be, when we consider what it has done and is doing in human society. His description of what property in absent and general goods does to human society, in the section, “Of the origin of government” and, “Of the source of allegiance,” is far from a panegyric. It is “encrease of riches and possessions” that he sees to lead to the abuses that “disturb men in the enjoyment of peace and concord” (T 541) to such an extent that magistrates are seen as a necessary invention. “[V]iolations of equity,” and “licentiousness” (T 535) follow the institution of contract. We then have a much worse state of affairs, from Hume’s description of it, than the pre-property “natural” state. The latter had the “inconvenience” (T 488) of insecurity of possession, but was not portrayed as violent, nor as nearly as intolerable as that state which drives people to invent magistrates and submit to their yoke.

But apart from that somewhat speculative hypothesis, there is another less speculative explanation of the likelihood of myths about how promises morally oblige us (supposing that they really do). On Hume’s official account, they are, after all, a social device to enable

individuals to fix an otherwise worryingly unfixed future, to fix it in a way that does not unduly threaten their fellows. Once promises are invented, an individual person's resolution, when expressed in the appropriate way, can indeed change her obligations. It is not so surprising an error if some people suppose that it cannot be just the words that are so powerful, but rather the act of mind that they express. This would be to try to appropriate, as a wholly private power, what on Hume's account is a *socially accorded* and *conditional* private power. But it is the sort of mistake that it is quite easy to make, especially for selfish individualists, apt to exaggerate their own single-handed powers. Such a mistake was made by theorists such as Hobbes, Grotius and Locke, who see there to be a basic moral rule telling us to keep promise, as if the giving (and so the keeping) of promises were not, for its very possibility, dependent on human invention and convention.

After the paragraph about the feigning of the willing of a new obligation, Hume goes on to "confirm this" (T 523) by discussing the role of the spoken word in valid promises, and to raise the question of whether a lying promise binds. That it does bind confirms the finding that it is not the individual inner act of will, or resolving, that creates a promissory obligation, but the knowing and known public act of invoking a publicly recognized form of speech act, in which to voice one's solemn resolution.

Another passage in the *Treatise* account, which Gauthier cites as suggesting that Hume's theory is really an error theory, is that early one at T 483 where Hume has been asking what the morally approved motive to honest acts can be, as a preliminary to presenting us with his version of honesty in matters of property as an artificial virtue. In this difficult and much discussed paragraph, he reasserts his Ciceronian "undoubted maxim," that "no action can be equitable or meritorious, where it cannot arise from some separate motive," and yet finds that the motive for observing "the laws of equity" seems to depend upon the conviction that property rights are inviolable (T 483). This gives us not a "separate motive," but one already involving justice and property, so "there is here an evident sophistry and reasoning in a circle." And what does Hume conclude from this exposure of a sophistry? "Unless, therefore, we will allow, that nature has establish'd a sophistry, and render'd it necessary and unavoidable, we must allow, that the sense of justice and injustice is not deriv'd from nature, but arises artificially, tho' necessarily from education, and human conventions" (T 483). Gauthier seems to take this to mean that educators and other sources of social pressure somehow do "establish a sophistry," since it is unavoidable that people be led into error if they are to be got to be just. But Hume seems to be saying that his conclusion, the artificiality and convention-dependence of justice, is an *alternative* to the establishment

of a sophistry. The reference to an established, necessary and unavoidable sophistry is surely ironic. The whole section is trying to show us that justice and the moral obligation to it are unintelligible until we appeal to conventions which intervene between purely natural and moral motivation. The sophistry and circular reasoning arise from an acceptance of the false dichotomy: "spontaneous natural motive *or* sense of moral duty/obligation." What the account of artifice offers us is a third middle way "*redirected* natural motive, a sense of *natural* obligation."<sup>7</sup> Hume is setting things up for his own account. For it is only when the natural motive of avidity is redirected, by the assurance that others are similarly redirecting their avidity, that there comes to be any approvable and effective natural motive to leave others with their possessions. And the knowledge that others are counting on us to go along with the co-operative scheme that establishes the redirection does give rise to a sense of duty to them, albeit not a fully moral one until the scheme gets moral approval.

The more puzzling passage, for my interpretation, is the one Gauthier cites at the very start of his paper, where Hume appears to repeat the false dichotomy that he had begun with: "either fully natural *or* moral motivation," this time in connection with promises. Surely by now the reader knows that these are not the only alternatives, that there is also convention-redirected natural motivation to consider. So why does Hume write as if we are beginning again, from scratch? He did not do that in establishing the second "law of nature," transfer by consent. Why, here when the third artifice is presented, must we be reminded of the difference that convention can make to human motivation? Well, Hume knows and we all know that there seems to be an especially great temptation to think that promising is a natural activity. The whole earlier moral tradition took it to be, in Hume's terms, "naturally intelligible." Hobbes, who spends a lot of time spelling out the intricacies of covenant, does not see covenants to be of human invention, an artificial thing, as government is for him. The claim that promise is an artifice is, perhaps, the most original claim in Hume's ethics,<sup>8</sup> so of course he wants to rub it in. And so, we get the same story as before, of how we must invoke the redirection of a more spontaneous motive by a convention before we can locate the motive that is approved when promise-keeping is approved. As earlier, in the account of the approved motive for respect for property, we are here offered "a new direction to those natural passions," whereby we satisfy them in "an oblique and artificial manner," rather than by "their headlong and impetuous movement" (T 521). And what self-interested passions are redirected at this point? The concern for our own future, with its tendency to lead us into "selfishness and ingratitude" (T 521). Future-concerned selfishness and ingratitude are allowed to remain in

force, but are now channelled into the desire to make advantageous contracts, followed by the willingness to keep them. This will not be out of gratitude to the first performer, but merely because otherwise one would be disabled from making further advantageous deals. So, immediately following the sentence pointing out the need for and possibility of a redirection of this rather nasty version of self-interest, Hume writes: "Hence I learn to do a service to another, without bearing him any real kindness; because I foresee, that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me or with others" (T 521).

For Hume, prudence is a virtue, but some forms of it are troublesome and vicious. Avidity is a vice, but its self-corrected redirected form becomes a virtue. Ingratitude and selfishness are vices, but by promise they are redirected into the virtue of fidelity to promises. (Other sorts of fidelity need not have such ugly natural sources.) In both cases, it is convention or artifice, giving rise to an "interested obligation," that achieves the redirection of troublesome self-interest, and makes possible a sort of motivation that can secure moral approval. But if it really is redirected ingratitude that we are approving, when we approve of contract-keeping, then this fact supports my earlier suggestion that Hume has his doubts about whether we really have good enough reason to approve of it. For Humean morality is in the business of encouraging gratitude and kindness, not in the business of making the social world commodious for those lacking kindness and gratitude.

I have considered and rejected Gauthier's suggestion (similar to one made earlier by Marcia Baron),<sup>9</sup> that Hume, while really holding an error theory of justice, finds it socially desirable to keep it submerged under a more upbeat but fallacious official story. The passages that seemed to support Gauthier's interpretation can be equally well explained as part of Hume's own dialectical presentation of his official upbeat account, or as pauses allowing him to address the real errors of other more contractarian theorists, such as Hobbes (and latter-day Hobbists such as Gauthier himself). The upbeat official theory is not fallacious, but it cannot be grasped in its full power and glory unless attention is paid to the historically shifting versions of "the interested passion" that Hume describes, as he gives us his account of each version's self-correction through a new artifice. Nor can one grasp or assess Hume's account of the motivation to justice if, like Gauthier, one "relegates" Hume's theory of morality "to a subordinate position." Gauthier writes that the sensible knave's message is that human society, which depends upon dispositions to justice, fidelity and allegiance, "lacks any moral foundation." He finds Hume not only to have no answer that will convince or convert the knave, but to have no

reason to reject the knave's message himself. But Hume, unlike Gauthier, does not find human society to depend so uniquely upon fidelity to contract, nor indeed upon any artifice. There are plenty more frequently found virtues in its moral foundation than justice, fidelity, and allegiance. "Meekness, beneficence, charity, generosity, clemency, moderation, equity" (T 578) are among the natural social virtues, and those special virtues that arise from "mere human contrivances for the interest of society" (T 577), namely justice, allegiance, respect for the law of nations, modesty, and good manners, will not be understood unless we suppose, with Hume, that the people who are motivated to cultivate them are also cultivating all the other, more "natural" Humean virtues. Artifices presuppose nature, and Hume's account of the artificial virtues presupposes that he has an account of human nature and of the natural social virtues. To separate the account of justice from that embedding framework, as Gauthier does, is indeed to undermine its foundations.

*University of Pittsburgh*

1. David Gauthier, "Artificial Virtues and the Sensible Knave," this volume.
2. See Baier, *A Progress of Sentiments: Reflections on Hume's "Treatise"* (Cambridge, 1991), chap. 10, for my similar account. It is, however, a concession on my part to allow that self-hatred for non-conforming acts may bolster the natural motivation that enlightened long-term self-interest could provide for conformity to the rules of justice. It was Rob Shaver who persuaded me on this point.
3. In what follows I am in agreement with Gerald Postema, "Hume's Reply to the Sensible Knave," *History of Philosophy Quarterly* 5 (1988): 23-40.
4. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2d ed., rev., ed. P. H. Nidditch (Oxford, 1975), 578 (hereafter cited as "T").
5. See Baier, *Progress of Sentiments* (above, n. 2), 245.
6. *Ibid.*, chaps. 10 and 11.
7. In recent correspondence Rob Shaver has raised the question of why what mere self-interest gives rise to should merit the name "obligation." Here it is not mere, but self-reflective and collectively co-ordinated self-interest, that is seen to produce the "obligation," and others are at risk if one neglects to follow this form of self-interest. My view is that anything that others count on or expect of us counts as an "obligation," a tie to them, whether or not

their expectations are morally reasonable. So there can be conventional obligations as well as moral ones, and self-interest may, if socially co-ordinated, generate obligations.

8. See Baier, "Hume's Account of Social Artifice—Its Origins and Originality," *Ethics* 98 (July 1988): 757-78
9. Marcia Baron, "Hume's Noble Lie: An Account of His Artificial Virtues," *Canadian Journal of Philosophy* 12 (September 1982): 539-55.