



The First Motive to Justice: Hume's Circle Argument Squared

Don Garrett

Hume Studies Volume 33, Number 2, (2007) pp. 257–288.

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

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

The First Motive to Justice: Hume’s Circle Argument Squared

DON GARRETT

“Justice” is Hume’s most common term for respect for property. On Locke’s view, the obligation to respect property is an original moral obligation imposed through a divinely instituted natural law that exists prior to and independent of human conventions. Hume expresses his disagreement with views like Locke’s by calling justice an “artificial virtue”—meaning by this that it is only “by means of an artifice or contrivance” that it produces the moral approbation that constitutes it as a virtue (T 3.2.1.1; SBN 477).¹ In his discussion of justice in *A Treatise of Human Nature* 3.2.1–4, he begins by arguing that justice (which he also calls “equity” and “honesty”) is “artificial” before going on to explain the origin of the convention on which it depends and why adherence to that convention is regarded as virtuous.²

Central to his defense of his claim that justice is an artificial virtue is a line of argument that is known (at least in some circles) as his “Circle Argument.” Hume applies this line of argument in *Treatise* 3.2.1 (“Justice, whether a natural or artificial virtue?”). The argument depends on a core thesis of his virtue-based approach to ethics: that the moral merit of an action is derived entirely from the moral merit of the virtuous “motive” of which it is a sign. (It should be emphasized that the scope of the term “motive” is very broad in Hume’s usage, encompassing character traits, abilities, dispositions, and recurring passions as well as occurrent desires.) The virtue of the motive is, in turn, discerned by the moral sense, by means of sentiments of moral approbation that are typically elicited, in part or in whole, through sympathy

with those who possess or are otherwise affected by the motive [T 3.1.2; SBN 470–6; T 3.3.1; SBN 574–91].) From this “Core Virtue Ethics Thesis,” as we may call it, he derives a general principle that we may call the “First Virtuous Motive Principle”: namely, that for any virtuous action, there must be a “first virtuous motive” that is *other* than a sense of moral duty to the action itself. This principle follows from the Core Virtue Ethics Thesis, he argues, because the only alternative would be a vicious “circle,” in which the moral merit of the action would have to be derived from the antecedent virtue of the motive that produced it, while the virtue of that motive could only be derived, in turn, from the antecedent moral merit of the action expressing it. Duty (i.e., a “regard to the moral merit of the act”) can function as a motive to act, and even (in its way) a morally praiseworthy and so virtuous one, on Hume’s account; but it can do so only when the moral merit of the kind of action in question has *already* been established by its relation to *another* virtuous motive. In acting *from duty*, he holds, we seek either to hide from ourselves our lack of that prior virtuous motive or to inculcate it in ourselves through habit (T 3.2.1.8; SBN 479). Once he has established the First Virtuous Motive Principle, Hume completes the Circle Argument by applying the principle to observations about the motives to acts of justice in order to argue that it must—on pain of circularity—be an artificial virtue.

Hume’s use of this line of argument has seemed notoriously problematic, however, for three main reasons. First, it appears that, despite the requirements of his own First Virtuous Motive Principle, Hume *does not* allow that there is any virtuous motive that can explain the full range of acts of justice other than the sense of duty to perform them. On the contrary, it appears that in at least two passages he explicitly *denies* that there is any such motive.

Second, it appears that, given his survey of possible motives, Hume *should not* allow that there is any virtuous motive for justice other than the sense of duty. He considers three possible motives: public benevolence, private benevolence, and self-interest. But he argues that neither kind of benevolence is sufficient to motivate the full range of just action; and although he does allow an important role to “self-interest” in the origin of justice, he also seems to concede that it, too, cannot ultimately motivate the full range of acts of justice. Even more fundamentally—and independent of worries about its motivational range—self-interest does not appear to be a *virtuous* or morally praiseworthy motive at all.

Third, it appears that Hume *cannot* allow that there is any “first virtuous” motive to justice without contradicting his own conative psychology. For he seems to allow, and even to insist, that one could consistently adhere to the requirements of justice only by applying rules to regulate one’s conduct independently of considerations of pleasure and pain, thereby taking those rules as *authoritative*. Yet his theory of motivation seems to require that only prospective pleasure or pain can ultimately motivate action, and hence the theory seems to provide no way in which to regulate action by an authoritative rule.

All of this seems quite disastrous for Hume. For since he holds that the First Virtuous Motive Principle is an unavoidable consequence of the Core Virtue Ethics Thesis, he cannot allow that there is no motive to justice satisfying the requirements of the Principle without allowing either (i) that the core thesis of his own virtue ethics is false or (ii) that he is wrong to characterize justice as a virtue.³

In the first section of this paper, I will consider whether Hume *does* allow that there is a first virtuous motive to justice other than a sense of duty or whether he denies it. In the second, I will consider whether he *should* allow that there is such a motive in light of his survey of the available options. In the third section, I will consider whether he *can* consistently allow such a motive within the constraints of his conative psychology. My conclusions will be that Hume can, should, and does identify such a motive. Commentators have typically characterized Hume's Circle Argument as involving irreconcilable contradictions or difficulties; but with respect to these three questions, at least, Hume's use of the Circle Argument can be squared.

Hume also holds, against Locke and others, that promise-keeping—which Hume calls *fidelity*—is an artificial virtue, and in *Treatise* 3.2.5 (“Of the obligation of promises”) he employs a version of the Circle Argument to support that claim as well. Three parallel puzzles arise about Hume's application of the argument to the case of fidelity, and parallel solutions apply. The texts giving rise to a puzzle concerning fidelity paralleling the first puzzle concerning justice are discussed in an appendix.

I. Does Hume Allow a Non-Moral Motive to Justice?

In a seminal article, David Gauthier⁴ emphasizes a passage in the *Treatise* in which Hume appears to deny that there is any non-moral motive to justice at all. (By “non-moral motive,” here and in what follows, I shall always mean “a motive that has no regard to the morality of the action” in the way that the motive of duty does have such a regard; to call a motive “non-moral” in this sense does *not* entail that it is not itself a morally *virtuous* motive.) The passage has since been much-cited. In an important subsequent article, Stephen Darwall⁵ also emphasizes a further passage from the essay “Of the Original Contract” in which Hume appears to deny that there is any non-moral motive for *any* artificial virtue. In order to understand the passage emphasized by Gauthier, it will be necessary to examine in some detail the version of the Circle Argument in which it is embedded. In order to understand the passage emphasized by Darwall, in contrast, it will be necessary to examine Hume's theory of obligation.

Does Hume Reject Non-Moral Motives to Justice? The passage cited by Gauthier occurs near the conclusion of Hume's application of the Circle Argument to the case of justice; in common present-day editions of the *Treatise*⁶ it reads thus:

From all this it follows, that we have *naturally* no real or universal motive for observing the laws of equity [i.e., the rules of justice], but the very equity and merit of that observance; and as no action can be equitable or meritorious, where it cannot arise from some separate motive, there is here an evident sophistry and reasoning in a circle. (T 3.2.1.17; SBN 483; emphasis added)

The word “naturally” does not appear in the first-edition version of this passage, which therefore reads simply, “. . . we have no real or universal motive . . .”; Hume added the word to his own printed copy of the *Treatise*, presumably as a correction intended for a possible second edition.⁷ In order to appreciate the significance of its addition and its relation to the argument as a whole, it is necessary to note that, in concluding the section of the *Treatise* that immediately precedes his discussion of justice, Hume distinguishes five senses of the term “natural.” He distinguishes these senses by means of relevant contrasts: what is “natural,” he writes, may be “opposed to miracles,” “opposed to what is rare and unusual,” “opposed to artifice,” “opposed to *civil*” or “opposed . . . to *moral*.” He adds in a footnote that, in what follows, “the opposition will always discover the sense, in which it is taken” (T 3.1.2.7–9; SBN 474–5).⁸ Accordingly, whenever quoting Hume, I try to disambiguate the term (using brackets) wherever it occurs.

Hume begins his Circle Argument about justice with two observations about moral evaluative practice:

- J1. When we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles in the mind and temper.
- J2. When we require any action, or blame a person for not performing it, we always suppose, that one in that situation shou’d be influenc’d by the proper motive of that action, and we esteem it vicious in him to be regardless of it.

He uses these observations to support the Core Virtue Ethics Thesis:

- J3. [Core Virtue Ethics Thesis] All virtuous actions [derive their merit only from virtuous motives, and are consider’d merely as signs of those motives. (from J1 and J2)

In the piece of reasoning that gives the Circle Argument its name, he then makes what he takes to be an uncontroversial logical point and applies it to his Core Virtue Ethics Thesis to deduce the First Virtuous Motive Principle:

- J4. [Logical Point] An action must be virtuous before we can have a regard to its virtue.

- J5. To suppose, that the mere regard to the virtue of the action, may be the first motive, which produc'd the action, and render'd it virtuous, is to reason in a circle. (from J4)
- J6. [First Virtuous Motive Principle] The first virtuous motive, which bestows a merit on any action, can never be a regard to the virtue of that action, but must be some other natural [i.e., non-moral] motive or principle. (from J3 and J5)

This last statement is sometimes read as claiming that the first virtuous motive to any action must be some other *non-artificial* motive, rather than some other *non-moral* motive. Such a reading is implausible, however, for Hume has at this point provided no premises about artificial or non-artificial motives from which such a conclusion could be drawn. Moreover, a restriction to non-artificial motives would have no point, since J6 is used only to derive the next step (J7), which explicitly concerns *non-moral* motives. Hume's use of the phrase "some other natural motive" in J6 suggests that he is contrasting the obviously moral character of "a regard to the virtue of the action" with the non-moral (and hence "natural") character of the "first virtuous motive." Modern punctuation would place a comma between "other" and "natural."

From this principle, Hume infers what he calls "an undoubted maxim":

- J7. [Undoubted Maxim] No action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality. (from J6)

Commentators have not distinguished this Undoubted Maxim (as I shall follow Hume in calling it) from the First Virtuous Motive Principle from which it is derived, and this is not entirely surprising: the Undoubted Maxim is introduced with the words, "In short, it may be established as an undoubted maxim," which leave it ambiguous whether it is intended as a restatement of the First Virtuous Motive Principle or a consequence of it. This is not the only place in his discussion of justice where Hume's language could be less potentially ambiguous—and were there not potential ambiguities, his discussion of justice would not have engendered so much confusion, and so many interpretations, among his readers. But the logical relations involved show clearly that the First Virtuous Motive Principle and the Undoubted Maxim are two different propositions, with the former following from what comes before in the argument, and the latter serving as the premise from which later claims are derived. The difference between them, even though the latter follows from the former, will prove to be important.

Suggesting that there will be "great difficulty" in finding a motive to just actions that will satisfy the Undoubted Maxim, Hume surveys three contenders—self-interest (which he also calls "private interest" or "self-love"), public benevolence, and private benevolence:

- J8. [S]hould we say, that a concern for our private interest or reputation, is the legitimate motive to all honest [i.e., just] actions: it would follow that wherever that concern ceases, honesty can no longer have place.
- J9. Self-love, when it acts at its liberty, instead of engaging us to honest [i.e., just] actions, is the source of all injustice and violence; nor can a man ever correct those vices, without correcting and restraining the *natural* movements [i.e., movements prior to artifice and contrivance] of that appetite.
- J10. Public benevolence . . . or a regard to the interests of mankind, cannot be the original motive to justice.⁹
- J11. Private benevolence, or a regard to the interests of the party concern'd [cannot be] the original motive to justice.¹⁰

From this survey of possible original motives, he concludes:

- J12. We have naturally [i.e., non-artificially] no real or universal motive for observing the laws of equity [i.e., of justice] but the very equity and merit of that observance. (from J8, J9, J10, and J11)

This step, J12, is of course the first clause of the passage emphasized by Gauthier; and the remainder of that passage simply re-iterates that the lack of any motive to justice other than a concern for the “merit of that observance” (i.e., duty) would involve the vicious circularity first identified in J5. On this basis, Hume draws his final conclusion about the artificiality of justice:

- J13. Unless we will allow that nature has established a sophistry, and rendered it necessary and unavoidable, we must allow, that the sense of justice and injustice is not derived from nature, but arises artificially, though necessarily, from education and human conventions. (from J5, J7, and J12)

Although Gauthier reads J12 as denying that we have *any* real motive to justice, he fully allows that it *may* be read instead as denying only that we have any *non-artificial* motive to justice. In fact, however, it can *only* be read in the latter, restricted fashion. First, that is the only reading that makes sense of Hume’s addition of the term “naturally” to J12. Second, it is the only reading that makes sense of J13’s conclusion that (barring a literally impossible “sophistry in nature”) the sense of justice is *artificial*. Third, and just as important, it is also the only reading that allows J12 itself to *follow* from J8–J11, given J9’s explicit limitation to the “natural” (i.e., non-artificial) movements of self-interest or self-love. Finally, to read the passage as denying that there is *any* motive to justice other than duty would conflict directly with Hume’s own subsequent claim that “self-love [i.e., self-interest] . . . produces the rules of justice, and is the first motive of their observance” (T 3.2.8.5; SBN 543).

Does Hume Reject Non-Moral Motives for All Artificial Virtues? Let us turn now to the passage cited by Darwall. It occurs in the essay "Of the Original Contract,"¹¹ where Hume distinguishes two classes of moral duties corresponding to the natural and artificial virtues, respectively:

All *moral* duties may be divided into two kinds: The first are those, to which men are impelled by a natural instinct or immediate propensity, which operates on them, independent of all ideas of obligation. . . . The second kind of moral duties are such as are not supported by any original instinct of nature, but are performed entirely from a sense of obligation. (Hume, "Of the Original Contract," 479–80)

If the "sense of obligation" which "entirely" motivates virtuous acts of this second kind is an exclusively *moral* sense of obligation, as Darwall asserts it is for Hume (438), then these virtuous actions would clearly violate the First Virtuous Motive Principle.

In fact, however, Hume does not say that these virtuous acts are performed entirely from a sense of *moral* obligation, and he gives every indication that they are not. Within the practical (as opposed to the epistemic) realm,¹² he regularly contrasts two species of obligation: (i) moral obligation and (ii) "the natural [i.e., non-moral] obligation of interest" (T 3.2.9.3; SBN 551; and T 3.2.11.4; SBN 556), which he also calls more simply "natural obligation" or "interested obligation."¹³ He explains moral obligation as follows:

When any action or quality of the mind pleases us *after a certain manner* we say it is virtuous; and when the neglect or nonperformance of it displeases us *after a like manner*, we say that we lie under an obligation to perform it. (T 3.2.5.4; SBN 517)

The *manner*, or kind, of displeasure relevant to moral obligation is, of course, the sentiment of moral disapprobation that, along with the sentiment of moral approbation, plays a central role in his moral sense ethical theory. He does not, in contrast, specifically explain the nature of interested obligation. If, however, moral obligation is constituted by displeasure in *one* certain manner at the neglect of an action, then it is reasonable to suppose that natural or interested obligation is constituted by displeasure in *another* manner at the neglect of an action—specifically, displeasure at the realization that the neglect of an act leaves one worse off than one would have been had the act been performed, so that neglect of it harms one's interests.¹⁴

One of the primary aims of the essay "Of the Original Contract" is to refute a central Lockean claim adopted by members of the Whig party—namely, that the

moral duty to *allegiance* (i.e., obedience to government) is derived from the moral duty to *fidelity* (i.e., promise-keeping) through a literal “social contract.” In the passage that Darwall cites, Hume is arguing that allegiance, like justice and fidelity but unlike the natural virtues, is recognized as a moral duty through a process in which a “sense of obligation” results from considering what he there calls “the necessities of human society, and the impossibility of supporting it, if these duties were neglected.”¹⁵ But this consideration is not originally a moral consideration, as his *Treatise* accounts of the origins of these three virtues make especially clear. On the contrary, Hume holds that a “consideration of the necessities of human society” gives rise first to a sense of *natural or interested* obligation to justice, fidelity, and allegiance. For participation in human society is, he argues, essential to the well-being of every human being; and justice, fidelity, and (ultimately) allegiance are essential to the maintenance of society. Human beings are therefore motivated first by an appreciation of the serious harm *to themselves* that would result without the benefits of society that only justice, fidelity, and allegiance can provide. This is then *followed*, on his account, by a sense of *moral* obligation, when moral sentiments arise from reflecting sympathetically on the similar positive effects of justice, fidelity, and allegiance on everyone in society. It is through these (originally self-interested) reflections, Hume remarks, that justice and fidelity “become obligatory, and acquire an authority over mankind” (“Of the Original Contract,” 480).

Accordingly, in the passage that Darwall emphasizes, Hume is claiming only that, unlike actions manifesting natural virtues, adherence to justice, fidelity, and allegiance must be motivated “entirely by a sense of obligation” that is first interested, and then also moral. For lacking natural (i.e., non-artificial) inclinations to adhere to these practices, human beings must instead be motivated first by a sense of the displeasing harm to their own interests, and then also a sense of the displeasing moral disapprobation, that will result from their non-performance.

II. Should Hume Allow a First Virtuous Non-Moral Motive to Justice?

As we have now seen, Hume does not explicitly deny that there is any motive to justice other than the sense of duty. But can he provide a motive for justice that satisfies the First Virtuous Motive Principle, or should he conclude from his survey of available motives that no such non-moral motive exists? In the Circle Argument, as we have seen, he considers three potential motives to justice. He there provides three reasons why public benevolence cannot provide such an original motive (Step J10), and three reasons why private benevolence cannot provide it (Step J11). Both sets of reasons include acts of justice (for example, returning property in a “secret loan” or to a “profligate debauchee,” respectively) that could not be produced by the motive in question. Thus, while not denying that some particular actions

required by justice might happen to be performed out of public or private benevolence, Hume denies that either of these motives alone can explain the full range of just acts that are actually performed—what we might call the full “behavioral profile” of justice. Hence, he concludes that neither can be the original non-moral motive giving rise to just action that is demanded by the Undoubted Maxim. It is a further question, however, whether his survey allows that self-interest, when properly “corrected and restrained,” might be such a non-moral motive.

Self-Interest in the Origin of Justice. In *Treatise* 3.2.2 (“Of the origin of justice and property”), Hume argues that “self-love” (i.e., a regard to self-interest) “is the original motive to the *establishment* of justice” (T 3.2.2.24; SBN 498; see also T 3.2.8.5; SBN 543), for it alone can motivate the adoption of a convention through which individuals allow one another stable possession of material goods.¹⁶ A convention exists among a group of individuals, as Hume explains it, when (i) each has an interest in following a particular course of conduct but only on the condition that the others follow a corresponding course of conduct; (ii) this common interest is mutually expressed and known among the parties; and (iii) this mutual expression and knowledge serve to produce “a suitable resolution and behaviour” (T 3.2.2.10; SBN 490). Conventions thus do not require any promise. The mutual willingness to adhere to the course of conduct specified by rules of justice creates a convention that bestows “stability on the possession of . . . external goods, [and leaves] every one in the peaceable enjoyment of what he may acquire by his fortune and industry” (T 3.2.2.9; SBN 489). The parties to this convention all serve their own interests (along with the valued interests of their families and friends) by “regulating their conduct” in accordance with the “strict and inflexible rules” of property, because such regulation allows the parties to maintain society, and in society each individual enjoys far more in the way of secure material goods than would be possible without this self-restraint. Given the recalcitrance of human passions, consistent adherence to fully worked-out conventions of justice “arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it” (T 3.2.2.10; SBN 490.) Yet the convention of justice is so essential to society that it is needed in some form to stabilize the family itself, for “every parent, in order to preserve peace among his children, must establish it” (T 3.2.2.14; SBN 493). Although the convention originates in self-interest, we soon come to approve morally those who adhere to its rules and disapprove those who do not; for we feel moral sentiments when considering with sympathy the effects of their characters on the members of society who benefit from the conventions.

The increasing size of societies, though in many ways beneficial, creates a further threat to justice and fidelity: the harmful effects of individual acts of injustice and infidelity on one’s own and the public interest become less certain, direct, and obvious, and so the temptations to violate the rules of property and promise-keeping in pursuit of a present desire increase. The solution to this problem

is a new convention involving the institution of a government, a structure that provides some individuals (e.g., magistrates and their deputies) with a more particular interest in protecting and enforcing the property and contract rights of others. Meanwhile, moral approval of justice and fidelity is further encouraged by the praise of politicians, who appreciate the importance of justice to society, and by parents, who see justice as a virtue that will be useful to their children in society as well as conducive to the well-being of society as a whole.¹⁷ As a result, these virtues come to play an especially important role in one's reputation, so that

every one who has any regard to his character, or who intends to live on good terms with mankind, must fix an inviolable law to himself, never, by any temptation, to be induced to violate those principles which are essential to a man of probity and honour. (T 3.2.2.27; SBN 501)

The Range of Self-Interest. Yet, although Hume *specifies* that self-interest is the original motive to justice, commentators have argued on two main grounds that he cannot ultimately regard it as satisfying the First Virtuous Motive Principle. The first of these is that Hume appears to allow, in three different passages, that there are circumstances in which acts of justice will *not* be in one's own interest; hence, it seems, self-interest can no more explain the full behavioral profile of justice than public or private benevolence can. The first of the three passages occurs in Step J8 of the Circle Argument itself:

[J8] Should we say, that a concern for our private interest or reputation, is the legitimate motive to all honest [i.e., just] actions: it would follow that wherever that concern ceases, honesty can no longer have place. (T 3.2.1.10; SBN 480)

Some commentators (e.g., Barry Stroud and Lorraine Besser-Jones)¹⁸ have read this remark as an admission that self-interest cannot be the original motive to justice because it cannot produce the full behavioral profile of justice. In fact, however, the passage only states a condition that self-interest must meet *in order* to provide the original motive—namely, that no actions in the behavioral profile of justice should fall outside its motivational range. Hume's immediately following remark, Step J9, indicates, as we have seen, only that self-interest cannot *naturally* (i.e., without the intervention of artifice and contrivance) meet this condition.¹⁹ This leaves open the possibility that self-interest *can* meet this condition as the result of artifice.

A second passage (also cited by Stroud and Besser-Jones) occurs after Hume has offered his account of the origins of justice:

Nor is every single act of justice, considered apart, more conducive to private interest than to public; and 'tis easily conceived how a man may

impoverish himself by a signal instance of integrity, and have reason to wish, that, with regard to that single act, the laws of justice were for a moment suspended in the universe. (T 3.2.2.22; SBN 497)

The purpose of this remark, however, is only to explain his claim in the same paragraph that

tho' the rules of justice are establish'd merely by interest, their connexion with interest is somewhat singular, and different from what may be observ'd on other occasions. (T 3.2.2.22; SBN 497)

The passage thus reaffirms that self-interest is the original motive that explains the behavioral profile of justice; and Hume then goes on to describe the way in which it does so:

But, however single acts of justice may be contrary, either to public or private interest, 'tis certain that the whole plan or scheme is highly conducive, or indeed absolutely requisite, both to the support of society, and the well-being of every individual. 'Tis impossible to separate the good from the ill. Property must be stable, and must be fix'd by general rules. . . . And even every individual person must find himself a gainer on ballancing the account; 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. When, therefore, men have had experience enough to observe, that, whatever may be the consequence of any single act of justice, performed by a single person, yet the whole system of actions concurr'd in by the whole society, is infinitely advantageous to the whole, and to every part; it is not long before justice and property take place. . . . Taking any single act, my justice may be pernicious in every respect; and it is only upon the supposition that others are to imitate my example, that I can be induc'd to embrace that virtue; since nothing but this combination can render justice advantageous, or afford me any motive to conform my self to its rules. (T 3.2.2.22; SBN 497–8)

That is, self-interest is able to motivate the full range of just acts *not* by favoring the just act in every case-by-case evaluation of interest, but rather by favoring the adoption of a “scheme” or *policy* of adhering uniformly to the rules of justice as long as others do so as well—a policy that *replaces* the default “policy” of trying to assess the outcomes of individual acts. The “infinite” advantages of stable property come only with “general rules” which, because of their generality, cannot

“separate the good from the ill” so far as particular instances are concerned. But considerations of “good and ill” in particular circumstances do not compete with the considerations in favor of the general policy of uniform adherence, on Hume’s account; rather, considerations of the advantages of having the general policy give each individual a motive to *exclude from consideration* the “good and ill” of particular instances—that is, to “fix an inviolable law to himself” (T 3.2.2.27; SBN 501). In short, self-interest “restrains itself” (T 3.2.2.13; SBN 492), as Hume says, by motivating individuals to adopt a policy that precludes acting on the assessment of interest in particular cases in favor of *uniform adherence*, thus producing the full behavioral profile of justice. It is this that benevolence alone, whether public or private, cannot do.

This is by no means the end of the matter, however. For Hume remarks in a famous passage in the concluding section of *An Enquiry concerning the Principles of Morals*:

Treating vice with the greatest candour, and making it all possible concessions, we must acknowledge, that there is not, in any instance, the smallest pretext for giving it the preference above virtue, with a view to self-interest; except, perhaps, in the case of justice, where a man, taking things in a certain light, may often seem to be a loser by his integrity. And though it is allowed, that, without a regard to property, no society could subsist; yet, according to the imperfect way in which human affairs are conducted, a sensible knave, in particular incidents, may think, that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy. That honesty is the best policy, may be a good general rule; but is liable to many exceptions: And he, it may, perhaps, be thought, conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions. (EPM 9.22; SBN 282)

This third passage—emphasized by Marcia Baron,²⁰ Gauthier, Darwall, and many others—raises clearly the threatening prospect of a *different policy*, which we may call the “knaveish” policy: act in accordance with the rules of justice *except* when one can violate the rules to one’s own considerable advantage (typically, without detection).²¹ If self-interest motivates this policy rather than the policy of uniform adherence, then it appears that self-interest cannot, after all, produce the full behavioral profile of justice.

In assessing the implications of this passage, however, it is important to understand what Hume’s search for an original motive requires and what it does not. If self-interest is to be the original non-moral motive to justice, it must *in fact* be capable of causing human beings to will to engage in the full range of actions

contained in the behavioral profile of justice—and to do so prior to any sense of duty to them, so that those actions can be signs of some other motive that will be judged virtuous. It is *not* required, for example, that every individual act of justice should actually maximize the interests of the person who performs it (even after taking into account its positive effects, if any, on the willingness of others to abide by the relevant convention). For a policy may still be the most *advantageous policy to adopt* even if there are some cases in which a violation of it would be more advantageous in its consequences than adherence. Specifically, it may remain the most advantageous policy to adopt as long as there is no other feasible policy that will, when adopted, provide the benefits of those violations without also producing greater harms to one's interest—for example, harms resulting from allowing too many or otherwise undesired exceptions. For these harms are not consequences of the intended violations considered individually, but rather necessary collateral costs of being able to implement the policy that then facilitates the violations. Nor is it even required, in order for self-interest to be the original motive to justice, that uniform adherence to its rules actually *be the best policy to adopt* from the standpoint of self-interest, so long as no other policy *actually under consideration appears* to be a better policy to adopt from that standpoint.

Individuals faced for the first time with the question of whether to adopt a policy with respect to property are very unlikely to adopt the knavish policy for two reasons. First, the knavish policy is unlikely to occur to them. Such individuals will necessarily be socially inexperienced and so lacking in social cunning. (Indeed, since initiation into the conventions of justice typically occurs in the family, such individuals will usually be children.)²² And for obvious self-interested reasons, if for no other, no one is likely to suggest the knavish policy to them. Second, they will be in circumstances in which adopting (i.e., attempting to implement) the knavish policy would not, in fact, be or appear to be superior to uniform adherence from the point of view of self-interest. The parties to the convention will initially constitute a very small society, consisting mostly of other individuals seen every day; and the opportunities for successful violations will be correspondingly very few, while the negative consequences for discovered violations may be great.²³ In addition, these individuals will typically be bad judges of *opportunities* for successful violation; and they will lack skill in concealment. (For example, although small children naturally experiment with thievery and cheating, they are also—fortunately for the cause of virtue—outlandishly bad at it.) Thus, when such individuals do violate the rules—more likely from strong temporary temptation than from any proposed knavish *policy*—they will undergo what Hume calls “repeated experience of the inconveniences of transgressing” them (T 3.2.2.10; SBN 490), even in cases where successful violation seemed possible. For all of these reasons, the knavish policy will neither be nor seem preferable to such individuals. Accordingly, the possibility of that policy need not prevent self-interest from motivating adoption of a policy of universal adherence to justice.

It is important to recognize that self-interest can thus constitute the original motive to justice even if later developments offer better opportunities for both the formulation of the knavish policy and the undetected rule-violations that it recommends. For once individuals motivated by self-interest *have* regulated their conduct by means of the rules of justice, prior to a sense of duty, the requirement of an *original* non-moral motive has already been met, regardless of the effects of later developments.

In fact, however, Hume's account also offers many reasons why universal adherence to justice remains stable even in the face of such opportunities. First, the full determination to *regulate* one's conduct by its rules is, *ipso facto*, a commitment to try to avoid *allowing* contrary considerations to weigh in one's deliberations in cases covered by the rules; hence later alterations in circumstances may not be salient. Second, for what it is worth, the performance of any kind of action of moderate difficulty that becomes habitual—including, presumably, adherence to these rules—may become pleasurable through what Hume calls the mechanism of "facility" (T 2.3.5.1–5; SBN 422–4). Third, the more one participates in society, the more likely one is to take greater pleasure, through sympathy, with all those who do or would benefit from one's own acts of justice. Fourth, as societies become larger, governments are instituted by convention to provide additional sanctions against violators. Fifth, as Hume remarks in considering the "sensible knave," the psychological and practical difficulties of trying to implement a knavish policy without detection and the potentially catastrophic consequences of failure may still render it an inferior policy to adopt even from the perspective of a narrowly material conception of self-interest.²⁴ Sixth, as he remarks in the same passage, the awareness of instances of such catastrophes for violators can be a source of "satisfaction" to those who do adhere uniformly to the rules. Seventh, what he calls a "reputation" for justice—something best obtainable by the inadvertent revelation that one *has* in fact acted justly when one might instead have violated the rules in secrecy—may be more valuable with a view to long-term self-interest, and also more immediately enjoyable through sympathy with others, than the material benefits of violating the rules. This will be especially true once *moral approbation* for justice arises, along with its resulting love and admiration for just individuals—a development that follows very soon after the origination of the conventions themselves. Finally, however, and most importantly for Hume, once moral approbation for existing justice does arise (and is then encouraged by "politicians and parents"), a new source of "*interested obligation*" to justice is created in the "inward peace of mind, consciousness of integrity, [and] satisfactory review of our own conduct" (EPM 9.23; SBN 283) that comes with moral approbation of, and pride in, one's own virtuous character. This steady approbation and pride, and the avoidance of the corresponding disapprobation and humility, is far more important to long-term happiness, Hume claims, than are

mere material goods, which he calls “worthless toys and gewgaws” (EPM 9.25; SBN 283) in comparison.²⁵

The Moral Merit of Self-Interest. Thus, self-interest can and does, on Hume's view, constitute the original non-moral motive to justice. It may still be objected, however, that self-interest cannot be the “first virtuous motive, which bestows a merit on” (see [J6]) acts of justice, as required by the First Virtuous Motive Principle. For unrestrained self-interest, at least, is not a virtue at all—on, the contrary, it is, as Hume observes, the primary source of violence and injustice. And he seems at best diffident about whether self-interest in general may be considered a virtue.²⁶

Because he regards unrestrained self-interest as the primary cause of violence and injustice, it is sometimes supposed that the first virtuous motive bestowing merit on acts of justice is, for Hume, “enlightened self-interest”—meaning, presumably, a combination of the motive of self-interest with accurate knowledge about one's own real interests, including how the conventions bear on them. Yet it seems implausible to suppose that even this restricted form of self-interest could be the “first virtuous motive” responsible for the merit of acts of justice. For, as Rachel Cohon has observed, someone who performed a just act solely as a result of calculating that that individual action would best serve his or her own interests might be praised for being *prudent*, but not for having the distinctive motive constituting the virtue of justice.²⁷ Conversely, if someone could discover through a careful and accurate analysis that a particular act of secret *injustice* would in fact be in his or her own interest, all things considered, such a violation of the convention would then be a behavioral sign of “enlightened self-interest” as just defined; yet that motive could not bestow any moral merit on the act. Moreover, it is possible for a *sense of duty*, in Hume's sense, to motivate acts of justice that one *believes* are contrary to one's enlightened self-interest; but this would be impossible if the “missing virtuous motive” one felt to be lacking and wished to develop in acting from duty were just “enlightened self-interest” itself. It seems, then, that although self-interest *is*, on Hume's view, the original motive in human nature to justice—thereby satisfying the Undoubted Maxim—it *cannot be*, even when “enlightened,” the “first virtuous” motive that bestows merit on acts of justice.

But there is no inconsistency here; for although the Undoubted Maxim is derived from the First Virtuous Motive Principle, they are not equivalent. The latter requires that, for every virtuous action, there be some first *virtuous* motive other than duty that bestows merit on the actions. The former requires that, for every virtuous action, there be *some* original motive, other than duty, *in human nature* to motivate its performance. The motive satisfying the Undoubted Maxim need not itself *be* a virtuous motive bestowing merit on the actions; for it may instead help to *create* that first virtuous motive, through the creation of a convention. The motive satisfying the Undoubted Maxim need only be, as Hume says, original “in human nature” since it should serve to explain the origins of any relevant conventions or

contrivances. The motive satisfying the First Virtuous Motive Principle, in contrast, must be *virtuous* in order to bestow merit on actions, but it need not be *original in human nature*, since it may arise instead from convention.

That is precisely what occurs in the case of justice. Self-interest motivates the invention of the relevant convention and the adoption of the policy of regulating one's conduct in accordance with its rules; but the motive that we find beneficial to its possessors and others, and so morally approve through sympathy, is the resulting artificial disposition to *regulate one's conduct by the rules of justice*, a disposition that is grounded in the new convention-dependent *desire* to regulate one's conduct by those rules.²⁸ As Hume remarks in connection with fidelity, "the conventions of men . . . create a new motive" (T 3.2.5.10; SBN 522). A particularly beneficial aspect of this regulating disposition is the disposition to follow the rules in a way that is independent of what might appear to be self-interested, or even publicly-interested, reasons to violate them—for the willingness to weigh costs and benefits of particular acts at the expense of the policy of uniform adherence will rightly seem dangerous to society and often to individuals as well.²⁹

III. Can Hume Allow a First Virtuous Non-Moral Motive to Justice?

As we have now seen, the virtuous non-moral motive originally constituting the virtue of justice is, for Hume, the disposition to be just, or to respect property—that is, to regulate one's conduct by the relevant rules. Indeed, Darwall proposes, quite rightly though on other grounds, that Hume *should* regard such a disposition as the original virtuous motive to justice. Yet he also holds that Hume does not and cannot consistently do so, for two reasons. First, citing the passages from the *Treatise* and "Of the Original Contract" discussed previously, he interprets Hume as denying that there is *any* first non-moral motive to justice. If what I have argued is correct, these passages do not contain such a denial. Second, however, citing passages from the *Treatise* concerning the motivational role of pleasure and pain, he also argues that Hume cannot consistently regard the first virtuous motive to justice as a disposition to regulate one's conduct by rules, on the grounds that Hume's own conative psychology cannot accommodate the needed regulation of conduct by rules that motivate through their own authority independent of prospective pleasure or pain.

Prospects of Pleasure and Pain. In support of this further claim, Darwall cites three passages in which Hume comments on the motivational power of pleasure and pain:

- [1] There is implanted in the human mind a perception of pain and pleasure as the chief spring and moving principle of all its actions. (T 1.3.10.2; SBN 118)

- [2] 'Tis from the prospect of pain or pleasure that the aversion or propensity arises towards any object: And these emotions extend themselves to the causes and effects of that object, as they are pointed out to us by reason and experience. (T 2.3.3.3; SBN 414)
- [3] DESIRE arises from good [i.e., pleasure] consider'd simply; and AVERSION is deriv'd from evil [i.e., pain]. The WILL exerts itself, when either the good or the absence of the evil may be attain'd by any action of the mind or body. (T 2.3.9.7; SBN 439)

As Darwall interprets these passages, they show that “Hume’s theory of action thus not only employs the traditional idea that the will invariably aims at the good . . . [but] interprets that idea hedonistically and egoistically” (422). He notes quite properly that “just persons would appear to regard the fact that something belongs to another, or that the rule of property requires forbearance, as a *motive*—i.e., a ground or a reason—for not taking it” independent of the prospect of pleasure or pain in doing so. He concludes that “this motive [i.e., requirement by a rule] can find no place in Hume’s theories of action and will as advanced in the *Treatise*” (422).

In fact, however, although Hume clearly holds that prospective pleasure or pain can be sufficient to move the will, he never claims that it is necessary. (Nor does he treat motivation by the prospect of pleasure and pain egoistically—Appendix 2 of *An Enquiry concerning the Principles of Morals*, entitled “Of Self-Love,” argues directly against psychological egoism). In the first passage just cited, he claims only that pleasure and pain are the “chief spring and moving principle” of human actions, not that they are the only spring or principle. The second passage cited may appear to be more restrictive, but Hume begins the paragraph in question by limiting his discussion to cases of motivation that arise “when we have the prospect of pain or pleasure from any object.” His point is only that, *in such cases*, it is that very prospect of pleasure or pain—rather than reason itself—that first produces desire or aversion directed toward an object and then extends an additional desire or aversion to the inferred means to it. The third cited passage again states only that prospective pleasure or pain can be sufficient for motivation, not that prospective pleasure or pain is necessary. On the contrary, immediately after remarking that the will is moved by prospective pleasure or pain, he emphasizes that there are other passions that move it *without* any such prospect:

Beside good and evil or, in other words, pain and pleasure, direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. (T 2.3.8; SBN 439)³⁰

More importantly, however, and regardless of his view of the doctrine that only prospective pleasures and pains can motivate, Hume need have no difficulty in explaining how human beings can be motivated to act in accordance with a rule while discounting or even ignoring the prospective pleasure or pain *offered by particular actions*. For on his account, one can readily come to believe that *adopting the policy* of uniform adherence is better, even from the limited point of view of self-interest, than adopting any other course of conduct one is considering—including the default policy of simply weighing the consequences of each individual act and acting accordingly.³¹ This belief, combined with what he calls the “calm desire” for pleasure and avoidance of pain, naturally creates a desire to adhere uniformly—a desire that is thus caused directly by prospects of pleasure and pain. This desire may, in turn and if all goes well, motivate one to act consistently in accordance with those rules despite the potential motivating force of prospective pleasures and pains associated with particular actions. Moreover, as one realizes that the blandishments of particular pleasures and pains offered by particular violations of the rules may cause deviations from this desirable and desired policy, the desire to adhere uniformly and so not to deviate will produce a further desire to discount or ignore those blandishments. Yet it is still the prospective pleasures and pains offered by the possible *policies* that motivate one to try to ignore these blandishments.

To be sure, these blandishments may be difficult to ignore, and they may themselves produce a *competing* desire to ignore the advantages of the policy of uniform adherence. The recognition of such conflicting desires about what considerations to ignore may give rise to a second policy question: namely, what policy to adopt about ignoring considerations in cases covered by the rules of justice. A consideration of prospective pleasures and pains will again tend to favor the policy of discounting individual blandishments over the policy of discounting the advantages of uniform adherence; and this may give rise to a further desire to implement the policy of discounting blandishments. Even so, uniform adherence may be psychologically difficult: it may require outside encouragement, voluntarily directed attention, habituation, repeated reinforcement (both positive and negative), and what Hume calls “strength of mind” (i.e., the tendency to be motivated by “calm” rather than “violent” passions; see T 2.3.4.10; SBN 418). But forces suitable for motivating uniform adherence are available in his conative psychology, and there is nothing in that psychology dictating that they cannot be effective—nor, on the other hand, that they must be effective. That is as it should be, for they are not always effective.

Obligation and Authority. Even with this schematic account of the *motivation* to adopt and follow the rules of justice, however, one might still wonder whether Hume can explain the nature of the obligating authority that one must recognize in such rules in order truly to *regulate* one’s actions by *reasons* that invoke them, so

as to “fix an inviolable law” for oneself concerning them. In Darwall’s interpretation, it is the tensions produced by just such an explanatory inability that lead Hume (1) sometimes to deny that there is any non-moral motive to justice at all, and (2) surreptitiously and somewhat confusedly to introduce a third species of *obligation* (in addition to moral and interested obligation), which Darwall calls “rule-obligation.”

We have already seen, however, that Hume does not deny that there are non-moral motives to justice; he denies only that there are such motives naturally—i.e., without a convention. The suggestion that Hume introduces a third species of obligation derives from the following passage:

After this convention, concerning abstinence from the possessions of others, is entered into . . . there immediately arise the ideas of justice and injustice; as also those of property, right, and obligation. The latter are altogether unintelligible, without first understanding the former. (T 2.1.1.22; SBN 490)

Darwall rightly emphasizes that neither moral nor interested obligation, as Hume understands them, is unintelligible without a convention; hence, he proposes, Hume here recognizes a third species of obligation that must be understood in terms of the authority of rules. If so, however, Hume’s remark would still be odd—for the rules of justice are only *some* of the rules that might have such authority, and Hume gives little reason to think that justice must necessarily be the first convention into which one can enter. Nor would understanding this third kind of obligation be a precondition for understanding “obligation” generally. Fortunately, however, such a revisionary interpretation is not required to make sense of the quoted passage. For as Jason Baldwin has recently argued, the “idea of obligation” that becomes intelligible with the convention of justice need not be that of a new *species* of obligation in addition to moral and natural obligation; it may instead be, and presumably is, simply that of a new *obligation* (or alternatively, a new set of obligations)—that is, the obligation (or obligations) to act justly.³² The obligation to act justly is, as we have seen, both interested and moral, but is unintelligible without conventions of property.

In fact, Hume has no need to introduce a third species of obligation, for he can readily explain both the authority of the rules of justice and their ability to generate reasons to act without it. “Authority,” for Hume, is simply reflectively approvable power, just as “reasons” are reflectively approvable considerations, and virtues are reflectively approvable character traits. The authority of the rules of justice originates in their power to inspire the desire to act in accordance with them—a power that is then regularly approved, first from an interested and then also from a moral standpoint. This approval involves a sense of natural and moral

obligation—that is, awareness that we will feel disapproval of ourselves, as self-harming and as immoral, respectively, for neglecting adherence to these rules. Since we approve of ourselves, in both an interested and a moral way, for regulating our action by these rules, we recognize them as reasons, both interested and moral, to act. The power and authority of these rules is particularly evident in their naturally and morally approved ability to motivate us to exclude other considerations, such as the prospects of pleasure or pain offered by particular acts, from weighing in the determination of our wills. In this way, the rules constitute what Joseph Raz has called “exclusionary reasons”—that is, reasons to exclude other potential reasons from weighing at all in the determination of one’s action.³³ Indeed, Raz explains the *authoritative* character of a directive precisely as a matter of its providing a positive reason to act while excluding other reasons:

The fact that an authority requires performance of an action is reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of them (46).

That is precisely how Hume understands the approved positive and exclusionary force of the rules of justice once they have, as he says, “become obligatory, and acquired an authority over mankind” (“Of the Original Contract,” 480)—an authority that they begin to acquire when we are moved by the consideration of prospective pleasure and pain to adopt uniform adherence to these rules as a scheme or policy.

IV. Conclusion

I have argued that, contrary to common perception, Hume does, should, and can recognize a motive to justice that satisfies the requirements of the First Virtuous Motive Principle. This motive is a disposition grounded in a desire to regulate one’s actions by the rules of justice. Because it refers to conventional rules, it is a “new motive” not original in human nature; it is instead produced by self-interest, which is the original motive in human nature that satisfies Hume’s Undoubted Maxim. In creating this new motive, self-interest authoritatively restrains itself, just as Hume says it does.

Another popular perception is that, on a Humean account, justice can only be maintained by erroneous judgment—what Baron has called a “noble lie.” While the question cannot be pursued further here, we have found little reason to think that any such errors are essential for Hume. Uniform adherence to rules of justice, both in small societies and in modern larger ones, may, as we have seen, rightly be judged to be the best policy to adopt from the point of view of *self-interest* even if

some individual acts of adherence fail to maximize self-interest. Similarly, uniform adherence to rules of justice may also properly win Humean *moral* approval, at every stage of society, without reliance on any errors about their consequences for public utility; for the Humean moral sense may approve of those who adhere to the rules of property and disapprove of those who would readily violate them for public benefit, even if some individual acts of adherence fail to maximize public interest. This is easy enough to understand: a likely result of the existence of public-spirited violators will be that others will feel forced to take greater precautions to protect their property than they otherwise would, so that no greater sum of public benefits will actually be produced and everyone will instead simply be inconvenienced by the need for costly additional precautions. Justice may indeed be judged, without error, to be the best policy³⁴—which is not, of course, any guarantee that we will not make many mistakes about it anyway.

Appendix: Fidelity and the Circle Argument

Just as Hume gives a version of the Circle Argument to establish that justice is an artificial virtue, so too he gives a version of it to establish that fidelity (promise-keeping) is an artificial virtue. And just as it may be asked whether Hume can, should, and does allow a first virtuous non-moral motive to justice, so too it may be asked, on parallel grounds, whether he can, should, and does allow a first virtuous non-moral motive to fidelity. The answers to all three questions are in both cases positive, and for similar reasons. Just as his cognitive psychology can accommodate the regulation of one's conduct by authoritative rules in the way required for justice (Section III above), so too it can accommodate the regulation of one's conduct by authoritative rules in the way required for fidelity. Just as self-interest serves as the original motive *in human nature* to the practice of justice while the disposition that it creates to regulate one's conduct by those rules serves as the first *virtuous* motive to justice (Section II), so too self-interest serves as the original motive in human nature to the practice of fidelity (T 3.2.5.11; SBN 523) while the disposition that it creates to regulate one's conduct by those rules serves as the first *virtuous* motive to fidelity. A close reading of the passage Darwall cites from "Of the Original Contract" (Section I) shows not only that it allows a non-moral motive to justice but also a non-moral motive to fidelity, and for the same reasons.

However, in addition to citing the specific *Treatise* passage that we have already considered (Section I), in which Hume appears to deny that there is any motive to *justice* other than a sense of morality or duty (T 3.2.1.17; SBN 483), Gauthier also calls attention to a parallel passage from the *Treatise* in which Hume appears to deny that there is any motive to *fidelity* other than the sense of duty:

Now, 'tis evident we have no motive leading us to the performance of promises, distinct from a sense of duty. 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 518–9)

This passage occurs near the conclusion of Hume's extension of the Circle Argument to fidelity. In order to determine whether it indeed denies that there is any non-moral motive to fidelity, it is necessary to understand both the context and the content of that extension of the argument.

Hume begins the section "Of the obligation of promises" (T 3.2.5.1; SBN 516) by describing his strategy for demonstrating the artificiality of fidelity:

That the rule of morality, which enjoins the performance of promises, is not natural [i.e., is not non-artificial], will sufficiently appear from these two propositions, which I proceed to prove, viz. *that a promise would not*

be intelligible before human conventions had established it; and that even if it were intelligible [before human conventions had established it], it would not be attended with any moral obligation. (T 3.2.5.1; SBN 516)

Throughout the execution of this strategy, Hume emphasizes the limitation of the circumstances he is considering to “natural” ones, not involving human conventions. He argues for the first proposition—“that a promise would not be intelligible before human conventions had established it”—by surveying various acts of the mind that do not involve conventions and showing that none of them is equivalent to promising; these acts include *resolving*, *desiring*, and *willing* the promised action. This result suggests that if promising is to be any natural (i.e., non-artificial) act of the mind, it must somehow be, not the willing of the promised action, but the willing of the obligation itself. Yet obligations result, Hume claims, only from agreeable or disagreeable sentiments with respect to an action, and we cannot “naturally” (i.e., directly, without some artifice and contrivance) change our sentiments by an act of will. Hume later refers back to this conclusion—in a passage that Gauthier also cites—when he remarks that “we have prov’d already, that there is no such act of the mind [as naturally willing an obligation in a promise], and consequently that promises impose no natural obligation” (T 3.2.5.12; SBN 523). There is no violation of the First Virtuous Motive Principle in this passage, since “natural obligation” again means “non-artificial obligation,” and an account of the obligation to fidelity as artificial does not require that promising be an act of willing an obligation.

Hume then offers two arguments for his second proposition—namely, that even if promises were intelligible prior to human conventions, such pre-conventional “promises” could not “naturally [i.e., non-artificially] be attended with any moral obligation” (T 3.2.5.5; SBN 518). The first argument for this proposition appeals to the same set of considerations that ruled out the suggestion that promising could be the natural willing of an obligation:

This appears evidently from the foregoing reasoning. A promise creates a new obligation. A new obligation supposes new sentiments to arise. The will never creates new sentiments. There could not naturally [i.e., non-artificially], therefore, arise any obligation from a promise, even supposing the mind could fall into the absurdity of willing that obligation. (T 3.2.5.5; SBN 518)

Hume’s extension of the Circle Argument to fidelity constitutes the second argument for the second proposition. He introduces it by remarking that “the same truth” just established—namely, that “there could not naturally [i.e., non-artificially] . . . arise any obligation from a promise” (T 3.2.5.5; SBN 518)—“may

be prov'd still more evidently by that reasoning, which prov'd justice in general to be an artificial virtue" (T 3.2.5.6; SBN 518). Thus, by the time he comes to apply the Circle Argument to fidelity, he can assume that the restriction of his negative arguments about fidelity to circumstances that are "natural," in the sense of being "non-artificial," is well-understood, and he makes the restriction explicit only intermittently—though still reasonably often—in the course of the new version of the Circle Argument itself.³⁵

Hume begins this version of the Circle Argument with a new formulation of the Undoubted Maxim. This is, of course, the same maxim that he has already derived from his Core Virtue Ethics Principle Thesis by means of the First Virtuous Motive Principle in his discussion of justice:

- F1. No action can be requir'd of us as our duty, unless there be implanted in human nature some actuating passion or motive, capable of producing that action. (from J3)

He then offers an analogue of his earlier logical point (J4), phrased this time in terms of duty and obligation:

- F2. A sense of duty supposes an antecedent obligation. (from J4)

Next, he gives a three-step argument to show that a "natural" obligation to an action requires a "natural" passion to motivate it:

- F3. Where an action is not requir'd by any natural [i.e., non-artificial] passion . . . it may be omitted without proving any [non-artificial] defect or imperfection in the mind or temper.
- F4. Where an action is not requir'd by any natural [i.e., non-artificial] passion it may be omitted without proving any . . . [non-artificial] vice. (from F3)
- F5. Where an action is not requir'd by any natural [i.e., non-artificial] passion it cannot be required by any natural [i.e., non-artificial] obligation. (from F4)³⁶

In this argument, the unstated implicit limitation of steps F3 and F4 to *non-artificial* defects and vices (i.e., to conditions that could be defects or vices independent of artifice) is required in two ways. First, it is required for plausibility. For Hume allows that conventions can give rise to "new motives"³⁷ or passions—such as desires concerning the rules of those conventions—and he has done nothing to argue that the lack of such motives or passions, in the presence of a convention, could not constitute a new defect or vice. Second, the limitation is required in order to explain the *explicit* limitation to "natural" (i.e., non-artificial) obligations in the conclusion at step F5.

From this conclusion and the previous logical point, he infers that it must be some motive other than duty itself that makes possible a natural (i.e., non-artificial) duty:

- F6. The motive [implanted in human nature through which an action can be requir'd of us as our {non-artificial} duty] cannot be the sense of duty. (from F2 and F5)

The restriction in the scope of step F6 to non-artificial duties is required by the explicit restriction to *natural* obligations in step F5, from which it is derived.

Next, Hume observes:

- F7. If we thought, that promises had no moral obligation, we shou'd never feel any [non-artificial] inclination to observe them.

From this, he infers:

- F8. We have no [non-artificial] motive leading us to the performance of promises, distinct from a sense of duty. (from F7)

And from this, in turn, he infers:

- F9. There is naturally [i.e., non-artificially] no inclination to observe promises, distinct from a sense of their obligation. (from F8)

Finally, from this plus his Undoubted Maxim and his elimination of duty itself as a possible original motive to acts required by natural duties he concludes:

- F10. Fidelity is no natural [i.e., no non-artificial] virtue, and . . . promises have no force, antecedent to human conventions. (from F1, F6, and F9)

Steps F7 and F8 together constitute Gauthier's second passage. Their implicit restriction to non-artificial motives and inclinations is *justified* by the general and often-expressed restriction of the entire argument (and of the larger strategy of which it is a part) to the case of non-artificial circumstances; but it is also *needed* in order to make sense of the *explicit* restrictions to "natural inclinations," "natural virtues," and circumstances "antecedent to human conventions" that occur in steps F9 and F10, which are derived from Steps F7 and F8. Thus, Hume is not denying that there is *any* motive to fidelity other than a sense of duty, but only that there is no such motive independent of artifice and convention. On the contrary, he immediately goes on to argue that self-interest leads individuals to adopt a convention through which they *do* have a "new motive" to keep their promises, concluding that "interest is the *first* obligation to the performance of promises" (T 3.2.5.11; SBN 523).

NOTES

I have benefited greatly from discussions of Hume's Circle Argument with Kate Abramson, Geoffrey Sayre-McCord, Gerald Postema, Lorraine Besser-Jones, David Owen, Tito Magri, and Colin Howson, as well as from written comments from Annette Baier. I have also benefited from comments and questions from audiences at the Thirty-Second Hume Conference at the University of Toronto (2005); the Philosophy in Assos Conference, in Assos, Turkey; Dartmouth College; the University of Massachusetts, Amherst; and Princeton University. Finally, I would like to thank Michael Gill and two anonymous *Hume Studies* referees for helpful comments and suggestions.

1 References to the *Treatise* (abbreviated "T") by Book, part, section, and paragraph are to Hume, *A Treatise of Human Nature*, The Clarendon Editions of the Texts of David Hume, ed. David Fate Norton and Mary J. Norton, 2 vols. (Oxford: Clarendon Press, 2007); the page reference following "SBN" is to Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, revised by P. H. Nidditch (Oxford: Clarendon Press, 1978). As soon emerges, artifice is essential to the "approbation" of justice by playing an essential role in the very origin of justice itself. Artifice does, however, often play a role in the encouragement of virtue and the heightening of approbation for it as well, as Hume also makes clear.

2 Hume also sometimes uses the term "justice" somewhat more broadly, to include "fidelity"—i.e., promise-keeping—as well as respect for property, but this usage is only sporadic. Throughout T 3.2 ("Of justice and injustice"), he uses the term "equity" interchangeably with "justice" as a term for accordance with the rules of property. He also uses the term "equity" in a list of *natural* virtues (at T 3.3.1.11; SBN 578), but there it clearly has the quite different sense of "evenhanded mildness, especially in pressing claims." The explanation of Hume's sometimes surprising terminological choices—including his very narrow sense of "justice," which is uncharacteristic even for the period—is beyond the scope of this paper.

3 Francis Snare, *Morals, Motivation, and Convention* (Cambridge: Cambridge University Press, 1991) and Rachel Cohon, "Hume's Difficulty with the Virtue of Honesty," *Hume Studies* 23.1 (1997): 91–112, have proposed that Hume should be interpreted as making the former allowance; Knud Haakonssen, *The Science of a Legislator: The Natural Jurisprudence of David Hume and Adam Smith* (Cambridge: Cambridge University Press, 1981) and David Gauthier, "Artificial Virtue and the Sensible Knave," *Hume Studies* 18.2 (1992): 401–28, have suggested that Hume may (ultimately, at least) be interpreted as making the latter allowance. Both of these lines of interpretation are desperate remedies, however. Contrary to the former, Hume must rely on the Core Virtue Ethics Thesis and the First Virtuous Motive Principle that he derives from it not only as key premises in his argument that justice is artificial but again to support his later argument that fidelity is artificial, which occurs only after his discussion of justice is quite concluded and beyond revision; and indeed he never issues anything like a recantation of either the thesis or the principle. Contrary to the latter, his lists of the virtues consistently include justice and fidelity, and at no point does he show any inclination to revise his classification of them; indeed, one of the final appendices to *An Enquiry concerning the Principles of Morals*, ed. Tom L. Beauchamp [Oxford: Oxford University Press, 1998]—Hume's last exclusively moral work—is devoted to them

specifically *as virtues*. It should be noted that, according to Haakonssen's proposal, Hume allows that we originally regard acts of justice and fidelity as meritorious as a result of the *misapprehension* that they indicate the presence of a praiseworthy motive (which is in fact merely feigned); and we are then moved to perform those acts not by this (feigned and illusory) motive but by the further virtuous motive of duty, out of a regard for their supposed merit. This might be interpreted as allowing Hume to retain both his view of justice and fidelity as real virtues *and* his virtue ethics by denying that the First Virtuous Motive Principle really follows from his Core Virtue Ethics Thesis. But while Haakonssen's proposal, so understood, suggests an interesting potential objection to Hume's derivation of the First Virtuous Motive Principle—namely, that his virtue ethics requires only that for any virtuous action we at least *suppose* that there is a first virtuous non-moral motive—Hume himself never expresses any doubts about that principle.

4 David Gauthier, "Artificial Virtue and the Sensible Knave."

5 Stephen Darwall, "Motive and Obligation in Hume's Ethics," *Noûs* 27.4 (1993): 415–48. See also his *The British Moralists and the Internal 'Ought': 1640–1740* (Cambridge: Cambridge University Press, 1995), chap. 10 ("Hume: Norms and the Obligation to Be Just"), which is based on this article.

6 Specifically, in *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, revised by P. H. Nidditch (Oxford: Clarendon Press, 1978) and *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton, 2 vols. (Oxford: Clarendon Press, 2000/2007).

7 Unfortunately, he never supervised a second edition, as the first edition did not sell out during his lifetime. For this reason, the correction remained absent from subsequent editions of the *Treatise* as well prior to the Selby-Bigge/Nidditch 1978 edition. (Remarkably, Nidditch makes the addition to the 1978 edition without any editorial comment.) In consequence of its history of omission, commentators sometimes quote the passage without the qualifying amendment. For example, two of the three quotations of the passage in Darwall, "Motive and Obligation in Hume's Ethics," lack it, as does Cohon's quotation of it in "Hume's Difficulty with the Virtue of Honesty." Cohon does add a footnote mentioning, though not defending, the omission.

8 In his primary discussions of justice and fidelity (*Treatise* 3.2.1–5), he uses the term "natural" in no fewer than three of these five senses—occasionally to mean "not unusual" and sometimes to mean "non-moral" (i.e., not morality-regarding), but most often to mean "non-artificial"—often relying on context alone, as promised, to indicate the intended meaning.

9 Hume provides three premises in support of this claim: "(i) Public interest is not naturally [i.e., non-artificially] attach'd to the observation of the rules of justice, but is only connected with it, after an artificial convention for the establishment of these rules. (ii) If we suppose, that [a] loan was secret, and that it is necessary for the interest of the person, that the money be restor'd in the same manner (as when the lender wou'd conceal his riches), in that case the example ceases, and the public is no longer interested in the actions of the borrower; tho' . . . there is no moralist, who will affirm, that the duty and obligation ceases. (iii) [M]en, in the ordinary conduct of life, look not so far as the public interest, when they . . . abstain from theft, and robbery, and injustice of every kind" (T 3.2.1.11; SBN 480–81).

10 Hume provides three premises in support of this claim: “(i) [If] the party concern’d . . . be my enemy, and has given me just cause to hate him . . . [or] he be a vicious man, and deserves the hatred of all mankind . . . [or] he be a miser, and can make no use of what I wou’d deprive him of . . . [or] he be a profligate debauchee, and wou’d rather receive harm than benefit from large possessions . . . [or] or I be in necessity, and have urgent motives to acquire something to my family . . . in all these cases, the original motive to justice wou’d fail. (ii) Were private benevolence the original motive to justice, a man wou’d not be oblig’d to leave others in possession of more than he is oblig’d to give them [or] at least the difference wou’d be very inconsiderable. (iii) A man’s property is suppos’d to be fenc’d against every mortal, in every possible case. But private benevolence towards the proprietor is, and ought to be, weaker in some persons, than in others: And in many, or indeed in most persons, must absolutely fail” (T 3.2.1.13–16; SBN 482–83).

11 Hume, “Of the Original Contract,” in *Essays Moral, Political, and Literary*, ed. Eugene F. Miller (Indianapolis: Liberty Fund, Inc., 1978): 465–87.

12 In epistemic contexts, Hume often describes steps in reasoning or argumentation (for example, taking particular points into consideration, challenging opponents to provide counterexamples, and drawing particular conclusions) as things we are “obliged” to do.

13 In keeping with his general policy on the use of the term “natural,” Hume also sometimes uses the phrase “natural obligation” to mean “non-artificial obligation” where the context makes this clear, as in his claims that “where an action is not required by any natural passion, it cannot be required by any natural obligation” (T 3.2.5.6; SBN 518) and that “promises impose no natural obligation” (T 3.2.5.12; SBN 523); see note 9.

14 This point was originally suggested to me in conversation with Daniel Amoni. The two kinds of practical obligation thus correspond to the two different kinds of sentiments that, Hume holds, one can feel in *considering a character*—namely, those “from interest and morals” (T 3.1.2.4; SBN 472). In the case of “sentiments from interest,” one considers only the effects of a character on one’s own interest; whereas in the case of “sentiments from morals,” one considers its effects in general, without special regard to how it will affect one’s own particular interest. It is reasonable to conclude, then, that on Hume’s view one has a *natural*, or interested, obligation to perform an act when not performing it would leave one worse off with respect to one’s interests in such a way that one would be *personally* displeased with oneself for not having performed it; and when not performing an act would indicate a character trait of which one would *morally* disapprove, in contrast, one has a *moral* obligation to perform it. Gauthier proposes that, while moral obligation is created by sentiments of moral approbation, “the motive associated with a redirected passion constitutes a natural obligation. The redirection curbs the natural force of the passion, restraining it so that it may be ‘better satisfy’d’” (410). Annette Baier proposes that “anything others count on or expect of us counts as an ‘obligation,’ a tie to them. So . . . self-interest may, if socially coordinated, generate obligations” (Baier, “Artificial Virtues and Equally Sensible Non-Knaves: A Response to Gauthier,” *Hume Studies* 18.2 [1992]: 439). But neither of these interpretations accords with all of Hume’s examples of natural or self-interested obligation. He describes, for example “very strong motives of interest or safety to forbear any action”

as, in general, constituting an "obligation" (T 2.1.10.6; SBN 312), without any mention either of motives restraining themselves for their own better satisfaction or of social coordination. Similarly, near the end of *An Enquiry concerning the Principles of Morals*, Hume remarks: "Having explained the moral approbation attending merit or virtue, there remains nothing, but briefly to consider our interested obligation to it, and to enquire, whether every man, who has any regard to his own happiness and welfare, will not best find his account in the practice of every moral duty" (EPM 9.1; SBN 278). (The reference is to Hume, *An Enquiry concerning the Principles of Morals* [EPM], ed. Tom L. Beauchamp [Oxford: Oxford University Press, 1998], cited by section and paragraph, and to Hume, *Enquiries concerning Human Understanding and the Principles of Morals* [SBN], ed. L. A. Selby-Bigge, revised by P. H. Nidditch [Oxford: Clarendon Press, 1975], cited by page). Here he seems to treat the question of whether we have an interested obligation to act as our moral duties requires as depending entirely on whether such action is most conducive to one's own "happiness and welfare." The implication is that we may have an interested obligation to *all* virtuous action—even though acting as required by natural virtues need not involve any redirected passions, on Hume's account. And although even such self-benefiting virtues as prudence may well create social expectations, Hume's defense of the claim that there is an interested obligation to virtue does not mention such expectations.

Moreover, Baier's interpretation cannot readily explain why there seem to be only two species of obligation—moral and interested—on Hume's account, since, given human foolishness, other people may well "count on" one regularly doing certain things that are neither morally required nor in one's own interest. And Gauthier's interpretation cannot readily explain why Hume regards both moral obligation and natural obligation as members of a *common species*, since the redirection of passions and the feeling of moral approval or disapproval have little obvious in common.

It is worth emphasizing that not every case of preference or choice is a case in which one has a natural obligation to choose one of the alternatives; for although one might happen to *enjoy* one alternative more than another, it does not follow that either alternative would leave one sufficiently worse off in the aftermath that one would be displeased with oneself for having chosen it. And this corresponds to Hume's usage, for he treats only choices in which one's ongoing *interests* may be harmed as involving a natural obligation. It is, of course, quite possible that one should have both a natural and a moral obligation to perform the same action.

15 The passage reads in full: "The second kind of moral duties are such as are not supported by any original instinct of nature, but are performed entirely from a sense of obligation, when we consider the necessities of human society, and the impossibility of supporting it, if these duties were neglected. It is thus justice or a regard to the property of others, fidelity or the observance of promises, become obligatory, and acquire an authority over mankind. For as it is evident, that every man loves himself better than any other person, he is naturally impelled to extend his acquisitions as much as possible; and nothing can restrain him in this propensity, but reflection and experience, by which he learns the pernicious effects of that licence, and the total dissolution of society which must ensue from it. His original inclination, therefore, or instinct, is here checked and restrained by a subsequent judgment or observation. The case is precisely the same with the political or civil duty of allegiance, as with the natural (i.e., non-civil) duties of justice and fidelity. Our primary instincts lead us, either to indulge ourselves in unlimited freedom, or to seek dominion over others: And it is reflection only, which

engages us to sacrifice such strong passions to the interests of peace and public order. A small degree of experience and observation suffices to teach us, that society cannot possibly be maintained without the authority of magistrates, and that this authority must soon fall into contempt, where exact obedience is not payed to it. The observation of these general and obvious interests is the source of all allegiance, and of that moral obligation, which we attribute to it” (Hume, “Of the Original Contract,” 480).

16 Such a convention is valuable, because—given the “selfishness and limited generosity” of humans and the relative “scarcity and easy change” of possession of material goods—it is essential to the maintenance of human society; and human society, in turn, is essential to satisfying human beings’ basic needs. Society is essential to human well-being, Hume observes, because it alone allows human beings to satisfy their needs by (i) augmenting their force through combining their powers, (ii) augmenting their abilities through the specialization permitted by a division of labor, and (iii) augmenting their security through rendering mutual aid in time of need. Human beings could not be expected to foresee these advantages prior to any experience with society; but fortunately, they are naturally induced to enter into society, first by the “appetite between the sexes” and then by the “natural affection” of parents for their children. Hume offers this explanation in T 3.2.2.2–4 (SBN 484–5).

17 Thus, while self-interest is the only motive that could motivate the original adoption of the conventions of property that make justice possible, such additional and potentially non-moral motives as the desire to please one’s parents may be among the contributing motives that first motivate adherence to rules of property within a society already practicing such a convention.

18 Barry Stroud, *Hume* (London: Routledge & Kegan Paul, 1977); and Lorraine Besser-Jones, “Living on Good Terms with Mankind: Hume’s Circle Argument” (conference paper, Pacific Division Meetings of the American Philosophical Association, Pasadena, CA, March 2004). See also Besser-Jones’s *Personal Integrity and the Sense of Duty: A Humean Theory of Justice and Moral Obligation* (unpublished Ph.D. dissertation, University of North Carolina at Chapel Hill, 2003).

19 Similarly, Hume’s reference to the inability of public benevolence to “operate with any force in actions so contrary to private interest as are frequently those of justice and common honesty” (T 3.2.1.10; SBN 481), also refers implicitly to a pre-conventional consideration of the benefits of just action.

20 Marcia Baron, “Hume’s Noble Lie: An Account of His Artificial Virtues,” *Canadian Journal of Philosophy* 12.3 (1982): 539–55.

21 Although the passage begins by mentioning only justice, Hume is here using the term “justice” in the broader sense that includes fidelity, as the remainder of the passage makes clear.

22 It would presumably be difficult at the least, on Hume’s view, for any human being to come to maturity without the nurturance of a family that inculcated the rudiments of justice. However, to deny that any human being could ever come to maturity without growing up in a family whose older members had themselves grown up in other families would be to court an infinite regress. As his silence on this question suggests, Hume has no answer to the question of the origins of humanity itself.

23 These negative consequences may include hatred and loss of material benefits to violators if their identity is discovered, and harm to the potentially fragile but essential convention of justice even when the occurrence of a violation is discovered without the identity of the violator. In the case of children, the negative consequences will likely also include loss of desired parental approval and protection when violations are discovered.

24 Thus he writes, “[W]hile they purpose to cheat with moderation and secrecy, a tempting incident occurs, nature is frail, and they give into the snare; whence they can never extricate themselves, without a total loss of reputation, and the forfeiture of all future trust and confidence with mankind” (EPM 9.24; SBN 283).

25 Jason Baldwin, “Hume’s Knave and the Interests of Justice,” *Journal of the History of Philosophy* 42.3 (2004): 277–96, in the course of arguing that the “sensible knave” passage poses no serious problem to Hume’s accounts of justice and fidelity, also mentions the effects of sympathy, government, and (to some extent) reputation in providing motives to the maintenance of justice and fidelity.

26 Thus he writes, “[W]hether the passion of self-interest be esteemed vicious or virtuous, it is all a case, since itself alone restrains it; so that if it be virtuous, men become social by their virtue; if vicious, their vice has the same effect” (T 3.2.2.13; SBN 492).

27 Rachel Cohon, “Hume’s Difficulty with the Virtue of Honesty,” *Hume Studies* 23.1 (1997): 91–112; see pages 104–5.

28 Because this new motive is artificial, it is essential to note again that, as previously argued, the term “natural” in the First Virtuous Motive Principle clearly means “non-moral,” not “non-artificial.”

29 Although the disposition to regulate one’s actions by the rules of justice is the first virtuous motive bestowing merit on acts of justice, we may well think that a fully civilized individual in modern society should also be motivated, in part, by a sense of duty in performing such actions. In the case of a natural virtue, such as affection for children, we would not, Hume holds, think less of someone who acted directly from that inclination without stopping to think of its moral character. But someone who faced a decision regarding justice without some awareness of its moral character might well be thought to be lacking in an appropriate, and important, use of the moral sense. Because the disposition to regulate our actions by rules of justice is artificial and requires that we discount other reasons for acting that might be attractive, we strongly approve the desire to attempt to re-inculcate and sustain in ourselves by practice—and this is a kind of motive of duty. (Similar remarks apply to fidelity.) Kate Abramson has helped me to appreciate the force of this point. Given this, it might be argued, one can even have a motive of meta-duty: regretting that one lacks a sufficiently strong first virtuous motive, *and* regretting that one doesn’t feel much present desire or ability to inculcate it in oneself by practice, either one then tries to make oneself more like the kind of person who *would* try to inculcate it by practice, or one tries to hide one’s lack of a sense of first-order duty from oneself.

30 Darwall notes this passage, of course, but interprets it as further evidence of Hume’s inconsistency (“Motive and Obligation in Hume’s Ethics,” 422–3; *The British Moralists*: 293–4).

31 Of course, self-interest is *not* the only motivating point of view for Hume, since there is also genuine benevolence; but his account of the origin of the rules of justice does ascribe their initial motivating force primarily to self-interest. As already indicated, the virtuous motive to justice is, for Hume, the firm disposition to follow the rules of property; that the acquisition or development of this disposition may itself be the result of self-interest does not undermine the virtue of the disposition itself. At most, it illustrates the fact that a genuinely virtuous character can be acquired through a non-virtuous motive—a state of affairs that would not disturb Hume.

32 Baldwin, “Hume’s Knave and the Interests of Justice,” 286n8.

33 Joseph Raz, *The Morality of Freedom* (Oxford: Oxford University Press, 1988), 124.

34 This is not to say, however, that Hume would necessarily disapprove morally of *every* violation of the rules of justice. For while it is a genuine virtue, it is not the only virtue; and someone who adhered to the rules of property even when a small violation of them was the only way to preserve life on the planet would obviously be taking the disposition of uniform adherence too far. Hume notes that the number of circumstances excusing violations of justice and fidelity in the case of “princes” is “greater” than that in the case of private individuals—which implies that there can be such circumstances even in the case of private individuals.

35 Baldwin also mentions, more generally, that the context of the argument suggests a limitation of the passage Gauthier cites to natural (i.e., non-artificial) motives (288–9). As with the parallel case of justice, Gauthier rather remarkably allows that this interpretation of the passage is possible but rejects it anyway.

36 The line of argument in steps F3–F5 parallels a line of argument that Hume has already offered—also limited, implicitly but clearly, to non-artificial circumstances (i.e., “in uncultivated nature”)—with respect to justice (T 3.2.2.8; SBN 488).

37 “They are the conventions of men, which create a new motive, when experience has taught us that human affairs would be conducted much more for mutual advantage, were there certain symbols or signs instituted, by which we might give each other security of our conduct in any particular incident. After these signs are instituted, whoever uses them is immediately bound by his interest to execute his engagements, and must never expect to be trusted any more if he refuse to perform what he promised” (T 3.2.5.10; SBN 522).