



Motive and Obligation in Hume's Ethics

Stephen Darwall

Noûs, Vol. 27, No. 4 (Dec., 1993), 415-448.

Stable URL:

<http://links.jstor.org/sici?sici=0029-4624%28199312%2927%3A4%3C415%3AMAOIHE%3E2.0.CO%3B2-Y>

Noûs is currently published by Blackwell Publishing.

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

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

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

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

*Motive and Obligation in Hume's Ethics*¹

STEPHEN DARWALL
University of Michigan

My object in this paper is to locate Hume in a fundamental debate in the seventeenth and eighteenth centuries concerning morality's normativity—its power to obligate. This is something Hume addresses only indirectly. Indeed, he belongs to a group of thinkers, which includes Shaftesbury and Hutcheson, who were sharply critical of the modern natural law tradition's focus on obligation. Nevertheless, what he does say on this topic places him in a very interesting position in the debate, both in relation to other modern ethics of virtue and to the problematic of the modern natural lawyers. Moreover, since Kant's revolutionary theory of moral obligation is perhaps best appreciated as an attempt to work out problems inherited from these two traditions, understanding Hume's relation to them should improve our understanding of Hume's relation to Kant.²

I. Setting the Stage

Guided by a broad-gauged philosophical naturalism, or, increasingly, by the thought that obligation can only arise through *autonomous* rational agency, or by both, it was virtual orthodoxy among British moralists at the end of the seventeenth century to view obligation as a *practical* condition—to understand normative force as the weight of rationally conclusive motive from an agent's deliberative point of view. Thinkers as diverse as Hobbes, Locke, Cumberland, Cudworth, and Shaftesbury took versions of this line.³ By 1699, Shaftesbury must have thought the idea sufficiently familiar that he could simply assume it, asking in his *Inquiry*, “[w]hat Obligation there is *to virtue*; or *what Reason* to embrace it.”⁴ Like many thinkers of the time, Shaftesbury assumed that an agent's single ultimately rational end is his own interest, and so argued that, since this is best realized in a virtuous life, there is indeed an obligation to virtue.

On its surface, this latter thought may seem more ancient, or perhaps Thomist, than modern. For the classical natural lawyers from St. Thomas on, morality was a formulation of God's eternal law: the harmonious ordering of final

causes in nature as this relates specifically to human beings. A harmony of individuals' interests was thus nothing morality is required to effect; that was already guaranteed by natural teleology. Following natural law leads each to his real good naturally harmonized with those of others.

The seventeenth-century natural lawyers rejected the teleological metaphysics of classical natural law, however, and with it any metaphysically guaranteed harmony of interests. For Grotius and Pufendorf, not to mention Hobbes and Locke, the lack of such a guarantee is what gives morality its very point.⁵ Conflict may result from genuinely conflicting interests, not just from individuals' failures in effectively pursuing their respective real goods. A stable harmony of human interests can be achieved only through a natural law (morality) that governs or redirects the pursuit of individual good. But precisely because morality was now thought to involve a form of guidance distinct from prudence, it became thinkable that individuals might do better by not being moral, whether occasionally or throughout a life. Thus the modern problem: how does morality obligate?⁶

Shaftesbury and Hutcheson rejected much of the modern natural law tradition—especially divine or secular voluntarisms like Locke's or Hobbes's that modeled morality on positive law and held moral obligation to depend on avoidance of immorality's extrinsic evils, rather than enjoyment of the moral life's intrinsic goods.⁷ But they agreed with the modern natural lawyers that morality is necessary to harmonize interests. And unlike ancient virtue theorists, they held that what makes a trait or motive a virtue is its being approvable from the moral point of view, by moral sense, not its contribution to a flourishing life for the agent. Both did believe that the virtuous life actually is the (nonmorally) best life for the agent, and, as I mentioned above, Shaftesbury stressed this as the source of its obligatory force. But that is because they thought that an agent's disinterested contemplation of her own virtue (through moral sense) makes a decisive contribution to the value of her life.

It is well known that Hume's fundamental approach in moral philosophy is more in sympathy with Shaftesbury and (especially) Hutcheson than the seventeenth-century natural lawyers. His meta-ethic is a kind of sentimentalism, whether cognitivist or non-, and he holds the object of moral sentiment always to be states internal to the moral agent—motives and character—rather than external conduct.⁸ Indeed, Hume's virtue ethics go Hutcheson's one better. Whereas Hutcheson maintained that in the context of *choice* (an agent's) moral sense approves derivatively of external actions that the motive most highly approved by (an observer's) moral sense would lead her to choose, Hume's ethics feature no such evaluations of acts from the agent's point of view. Moral sentiment always has motive and character as object; it is invariably an observer- rather than agent-phenomenon: "when we praise any actions, we regard only the motives that produced them, The external performance has no merit. We must look within to find the moral quality." (T.477). And Humean moral obligation, I shall argue,

derives from an observer's response to contemplated character, not in anything (to use Cumberland's words) that "can superinduce a *Necessity* of doing or forbearing any thing, upon a Human Mind deliberating upon a thing future."⁹

As always with Hume, however, things are not so simple. When he comes to discuss justice and the obligation to be just, the richness and complexity of his thought outrun his usual categories, giving him a much more interesting view and a distinctive place in early modern thinking about moral obligation. Although consistently with his brand of virtue ethics, Hume must continue to hold that the moral obligation to justice fundamentally concerns character—the *virtue* of justice rather than just acts—this involves a motivational state quite unlike any other Humean virtue: the agent's regulating her conduct by rules she regards as authoritative.

Before we can begin to consider this complexity, we need one last bit of stage setting. Pufendorf influentially defined obligation as "that whereby one is required under moral necessity to do, or admit, or suffer something."¹⁰ But whereas it became quite common (under the pressure of philosophical naturalism) for the necessity involved in obligation to be understood *practically*, as generated by an instrumentally necessary means to a naturally necessary end, Pufendorf explicitly rejects this approach. He contrasts moral necessities (and what he calls "moral entities" more generally) with those he considers merely natural.¹¹ The former, he argues, can exist only by way of *rules* or *norms* prescribed to guide the will. And their authority derives from the superiority of prescriber to subject. Most important for our purposes, Pufendorf holds that the authority of such a norm cannot be reduced to any goods the agent might realize by following it. Sanctions may be required to create motives to conform, but norms obligate, not through these, but through "fear mingled with reverence," owing to subjects' knowledge that sanctions will be *justified*—in the case of God's sanctions, by His goodness towards us.¹²

Now this appears to create a split between obligation and rational motive. But while he normally accepts the classical theory of will as necessarily aiming at the good (saying, for example, that "nothing can constrain the human mind, as it deliberates on the future, to do or to avoid anything, except reflections on the good and evil which will befall others and ourselves from what we do"), Pufendorf also says that, unlike action to avoid an evil, "whatever we do from an obligation is understood to come from an intrinsic impulse of the mind."¹³ Roughly seventy years later, we find a similar theme in Crusius. Prudence, the "choosing...good means to one's ends," contrasts with "true obligation," the latter requiring *recognition of a law* (again, with Crusius, the prescription of a superior). Obligatory obedience is properly motivated, not by a desire for the good or an aversion to evil, but (somehow) by respect, through "the drive of conscience" acknowledging superior authority.¹⁴

What we have here is a fundamental interaction in the seventeenth and eighteenth centuries between a question about the nature of moral obligation and one

concerning the nature of the will. So long as the rational will is held necessarily to aim only at good and the avoidance of evil, then either obligation must be understood in some such terms, or it must be disconnected from rational motive. Kant, of course, will later attempt to synthesize these elements with an account of free rational will as involving self-given laws, but in the middle of the eighteenth century no such thought was clearly in sight.

Hume follows orthodoxy in his official remarks about the will, which, he says, “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.439, see also T.399) But while he generally tries to accommodate his moral psychological claims within this framework, what he says about the virtue of justice and the obligation to be just simply does not fit. What is distinctive about just persons, he tells us, is not their seeking some particular good or avoiding some particular evil “by any action of the mind or body,” but that they *regulate* themselves by *rules* (of property, transfer, and promise) that they regard as “sacred and inviolable.” (T.533) Since Hume is committed to holding the moral obligation to justice most fundamentally to concern the *virtue* of justice (“when we require any action...we always suppose, that one in that situation shou’d be influenc’d by the proper motive of that action” (T.477)), it would seem the conclusion he should draw is that, once favorable social conditions have established the rules of justice, there is a moral obligation for agents to regulate their conduct by them. Still, Hume struggles mightily to squeeze ideas into his usual moral psychology with its orthodox theory of will even here.

II. Some Beginning Puzzles

As soon as one starts to sort out Hume’s views about obligation in the *Treatise*, especially in relation to what he there says about justice, various puzzles arise. Here are several:

(a) The very idea of obligation, Hume tells us, like those of justice, right, and property, depends on mutually advantageous conventions having already been “enter’d into.” (T.490–1) “Those...who make use of the word...*obligation*, before they have explain’d the origin of justice, or even make use of it in that explication, are guilty of a very gross fallacy, and can never reason upon any solid foundation.” (T.491) Several pages later, however, Hume announces that, having “fully explain’d” the “*natural* obligation to justice, viz., interest,” it remains to demonstrate the “*moral* obligation,” evidently meaning by the latter the sentiment of moral approbation towards justice and disapprobation towards injustice. (T.498) If the idea of obligation depends on convention, then since neither the moral sentiment nor the motive of self-interest do, how are we to understand Hume’s referring to both as *obligations* to the practice of justice?

(b) Sometimes Hume appears to treat ‘justice’ (and ‘honesty’) as terms of moral approbation. Thus following his general argument that “the first virtuous motive, which bestows a merit on any action, can never be a regard to the virtue

of that action," (T.478) Hume applies this argument to the special case of justice. The "honesty and justice" of restoring a loan, he says, "can never be a regard to the honesty of the action. For 'tis a plain fallacy to say, that a virtuous motive is requisite to render an action honest, and at the same time that a regard to the honesty is the motive of the action." (T.480) At other points, however, Hume treats 'justice' (and 'honesty') as morally neutral, defined by mutually advantageous rules concerning possessions, transfer, and promise. And he there supposes that whether justice is a virtue—or whether, as he also puts it, there is a moral obligation to be just—is something that must be established. We must ask, he says, "*Why we annex the idea of virtue to justice, and of vice to injustice.*" (T.498) So is the idea of virtue annexed to justice or contained within it?

(c) The moral obligation to justice, again, consists in the sentiment of approbation and disapprobation towards justice and injustice, and love and hatred towards the just and unjust, respectively. But none of these sentiments or emotions can be a direct motive to action according to Hume, since, for one thing, the object of moral approbation is always a motive and never an "external performance".¹⁵ Nonetheless, Hume also writes that, for example, "'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.518)¹⁶ This suggests we are morally obligated, and through this realization motivated, to perform just *acts* in the first instance, even though, if moral obligation consists in approbation, and if approbation can never have external conduct as object, this cannot be.

(d) As I mentioned above, Hume generally takes the view that ("by an *original* instinct") will aims at some good (i.e., pleasure) in prospect (T.438–439,399). But it is important to his view that justice is a rule-regarding rather than a good-regarding virtue. Unlike benevolence and self-love, the trait of justice *itself* aims at no good to self or others, but at regulation by mutually advantageous rules (even though moral sentiment and self-interest both endorse the trait for its usefulness). How, then, is voluntary just conduct possible?

Puzzles (b) through (d) combine in Hume's astonishing remark that we have "no real or universal motive for observing the laws of equity, 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.483)¹⁷ Taken literally, this says that justice is not simply puzzling, but paradoxical.

No interpretation can dissolve all these puzzles. They result, I believe, from Hume's trying to combine the following three positions: (a) his deeply insightful account of justice (as realized both in social practice and in the motivation of moral agents), (b) the fundamental principle of his virtue ethics (that the direct object of moral approbation and disapprobation is always some trait of character), and (c) his theory of the will as invariably aiming at some prospective good or the avoidance of some prospective evil. Hume could have held (b) and (c) as

Hutcheson had) had he not made the important discoveries about justice that today we regard as among his most important contributions. This account (a) is, however, incompatible with (c). In what follows, I consider how, if Hume's official theory of will is jettisoned, his insights about justice can be combined with his ethics of virtue to yield a highly distinctive contribution to early modern thinking about the obligation to be just.

III. The Hutchesonian Background

What Hume calls "natural obligation" and "moral obligation," respectively, come directly from Hutcheson.¹⁸ 'Obligation' is ambiguous, Hutcheson maintains, as between an action's being "*necessary to obtain Happiness to the Agent, or to avoid Misery*" and its being the case that "every Spectator...*must approve*" an action, "*and disapprove...omitting it, if he considers fully all its Circumstances.*" (II.i) Following Shaftesbury, Hutcheson holds that (what Hume will call natural) obligation consists in "a Motive from Self-Interest, sufficient to determine all those who duly consider it..." (In.vii.i)¹⁹ But he denies Shaftesbury's contention that self-love is the only reflectively rational motive. This, he objects, would make the only obligation to virtue a motive which is not itself virtuous. It would unacceptably disconnect natural from moral obligation.²⁰

Hutcheson's own view is that every virtuous motive is an instance of benevolence and that universal benevolence is both the morally best, and no less a rational, motive than self-love. His famous *Illustrations* arguments that reason alone cannot motivate (from which Hume borrowed liberally) notwithstanding, Hutcheson advances a theory of rational desire and will in the *Essay* according to which calm self-love and universal benevolence are equally rationally authoritative human motives.

Any rational motive, Hutcheson holds, aims at some good (pleasure) to self or others. And will just is "the disposition of the Soul to pursue what is presented as good, and to shun Evil." (II.i)²¹ What makes a motive rational for Hutcheson is that it arises through the use of reason, i.e., theoretical reason apprehending natural good and evil (pleasure and pain) in prospect. It is simply a fact of our psychology, Hutcheson believes, that calm reflection on facts about natural good and evil, from an agent-relative and an agent-neutral standpoint respectively, leads in the direction of "two grand determinations" or organizing desires, "one toward our own greatest happiness, the other toward the greatest general good."²² In this sense, then, universal benevolence is every bit as much an ultimate rational motive for human agents as is self-love. Fortunately, God has arranged things so that these never conflict.²³

When Hutcheson distinguishes a source of obligation other than self-interest, however, he identifies it with the sentiments or "perceptions" of moral sense, and not with any motive of which moral sense approves. This is curious, since according to his psychology, moral sentiment cannot be a motive; only desires for *natural* good (pleasure) can. To Shaftesbury's proposal that benevolence is

insufficient for genuine virtue or merit, the latter requiring in addition the agent's reflective approval through moral sense, Hutcheson objects that the only motive the agent gains thereby is a self-interested desire for the pleasurable contemplation of his own benevolence, and this "does not increase our approbation." (II.v) Approbation is no motive, and the only motive to which it gives rise is no virtue.

Given his objections to Shaftesbury, and his thesis that universal benevolence (the morally best motive) is no less rational than calm self-love, one might expect Hutcheson to maintain that these two motives are both obligations to virtue, one extrinsic, the other intrinsic. With no evident philosophical rationale, however, Hutcheson terms moral sense's approbation itself an obligation, and Hume simply follows his lead.²⁴

IV. Hume's Departure: Justice

It is when Hume comes to justice, the virtue whose natural and moral obligation he undertakes to establish, that his thought departs radically from Hutcheson's. There are important similarities even here, however, since, like Hume, Hutcheson advances a rule-consequentialist theory of rights and justice.²⁵ Where they disagree starkly is in Hume's belief that justice is a virtue quite independent of benevolence that aims, not directly at any good, but at complying with generally beneficial rules. Hutcheson simply assumes that violating such rules "is always exceedingly evil, either in the immediate, or remote Consequences of the Action." (In.vii.ix.10)²⁶ By his lights, right-defining rules can enter into agents' deliberations only as *summary* rules (in Rawls's sense), guiding benevolence, the master virtue. And "Justice, ...if it has no regard to the Good of Mankind...is a Quality proper for its ordinary Gestamen, a Beam and Scales, than for a rational Agent." (In.ii.i)²⁷

For Hume, however, a convention of justice arises when members of a society communicate their common interest in "regulat[ing] their conduct by certain rules" (most prominently, to abstain from the possessions of others), and come actually to do so—when, that is, they establish mutually advantageous *practices*, and come to regard the rules structuring them as *practice rules*, in Rawls's sense.

I observe, that it will be for my interest to leave another in the possession of his goods, *provided* he will act in the same manner with regard to me. He is sensible of a like interest in the regulation of his conduct. When this common sense of interest is mutually express'd, and is known to both, it produces a suitable resolution and behaviour. (T.490)

This, Hume concludes, "may properly enough be call'd a convention or agreement betwixt us." (T.490) Slowly, by gradual accretion, the common sense of interest and resolve spreads throughout society until enough members regulate their conduct by the rule of property, "upon the supposition, that" similar rule-required actions are "to be perform'd" by others. (T.490)

The latter construction brings out nicely the prescriptive/predictive aspects of the situation. From an external standpoint, each expects that, by and large, others will in fact regulate their conduct, and regulates himself on that expectation. But it is also crucial for Hume, I shall be arguing, that the rules come to have for each what Hart called an “internal aspect”—that each regard the rules prescriptively.²⁸ If the expectation were simply *that* others will in fact regulate their conduct, rather than an expectation *of* them (that applies also to oneself), then this could give no one a basis for regulating her own conduct. Only for a person who is resolved to regulate her conduct by mutually advantageous rules if others do (and thus who regards the rules prescriptively on that condition), would the knowledge that others will in fact regulate their conduct provide any ground for conforming herself. Such a person, we might say, regulates her conduct by a *meta-rule of cooperation*: regulate your conduct by mutually advantageous rules (of promise, transfer, and property) so long as others do.

Note how the conditional works in the passage indented above. Hume is not saying that each expresses to others the prediction that if they abstain from one’s possessions, then it will be in his interest to abstain from theirs. What each expresses to others is the sense that it will be in his interest to follow a certain rule: abstain if others do. Or perhaps: follow mutually advantageous rules provided others do.

Now as I shall discuss in Section VI below, Hume is somewhat ambivalent in the *Treatise* concerning the relation between regulating conduct by rules of justice and acting for the good, either the agent’s own or the public’s. But whatever the exact relation, Hume appears to consider justice a distinct virtue, and thus a distinct trait of character, from any focussed on the goods of self or others—for example, from prudence or benevolence. And he seems to be saying that what is distinctive about this virtue is the just person’s regulation of her conduct by rules that define the practices of property, transfer, and promise. Just persons, he tells us, “lay themselves under the restraint” of these rules (T.499,532); “strict[ly]” “regulate their conduct” by them (T.490,534); “impos[e]” on themselves these “general inflexible rules” and regard them as “sacred and inviolable” (T.499,533). Each person, he writes, must “fix an inviolable law to himself, never, by any temptation, to be induc’d to violate” them. (T.501)^{29,30}

What this *says* is that just persons regard the rules of justice as prescriptive or normative, conditional on others doing so. They conditionally treat these rules, or perhaps the meta-rule of cooperation, as authoritative, as *in themselves* giving reasons, indeed conclusive reasons, not to take the property of others, keep their promises, and so on. In one sense of the word, therefore, 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.³¹

The problem, of course, is that this motive can find no place in Hume’s theories of action and will as advanced in the *Treatise*.³² The “perception of pain

and pleasure," Hume writes, is "the chief spring and moving principle of all...actions." (T.118) 'Perception' is ambiguous here as between two different ways that pain and pleasure "mak[e] their appearance in the mind." (T.118) They appear in person, as it were, as impressions, and since "impressions always actuate the soul" they can then cause action directly. (T.118) Touching the hot stove I may pull back. Or testing the pleasantly warm tub water, I may immerse myself further. But *ideas* of pleasure and pain also arise in the mind, including of pleasures and pains that would result from possible actions. When these are sufficient in force and vivacity to constitute beliefs, they come to have a "like influence on the passions" as do pleasurable and painful impressions. (T.119, see also T.414) So believing pleasure will result if I immerse myself in the tub can have an effect similar to that of the pleasurable impressions of testing the water. And the belief that touching the stove will cause pain can result in an aversion to doing so. "'Tis from the prospect of pain or pleasure," Hume writes, "that the aversion or propensity arises towards any object." (T.414) Good and evil ("or in other words, pain and pleasure") "consider'd simply" give rise to desire and aversion respectively. (T.439) And "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.439)

Hume's theory of action thus not only employs the traditional idea that the will invariably aims at the good. It interprets that idea hedonistically and egoistically. Desires and aversions arise from the prospect of pleasure or pain, respectively, for the agent. Hume recognizes that this theory won't work for all cases. Indeed, he says that "direct passions *frequently* arise from a natural impulse or instinct, which is perfectly unaccountable" and unconnected to any thought of pain or pleasure. (T.439, emphasis added) As examples he gives "the desire of punishment to our enemies, and of happiness to our friends," and such bodily appetites as hunger and lust. (T.439) To these we should add calm passions or "instincts originally implanted in our natures, such as benevolence and resentment, the love of life, and kindness to children." (T.417) Pleasure and pain will still be connected to these passions, but through their satisfaction, not because they are caused by the idea of some pleasure or pain for the agent. (T.439)

Now Hume doesn't tell us what he takes the will to aim at in such cases. Resenting an injury, I may desire to punish my enemy, and see an opportunity to do so. If, however, will can exert itself only when a good (pleasure), or the absence of an evil (pain), to the agent is attainable, then my motive can only be something like the satisfaction of my desire to punish, rather than, say, that he deserves it, or that this will lay him low. On the other hand, if the will can have the same object as these "nonstandard" direct passions, then it can aim directly at an enemy's (deserved) unhappiness. Since Hume appears to be assuming that the will exerts itself to attain the object of desire, or the absence of the object of aversion, it is hard to see how he can avoid the latter possibility, once he has

admitted the existence of nonstandard direct passions. If this is right, Hume will be committed to revising his theory of action and will quite independently of anything he says about the just person's regulation of her conduct by conventionally established rules of property, transfer, and promise.

We shall return to the relations between the virtue of justice, regulation by the rules of justice, and the motives to just action in Sections VI and VII below. Before taking our leave, however, we should note a further textual problem in understanding Hume's talk about regulation by rules. When Hume discusses "what reason or motive" there is to just acts when the rules of justice have been established and people have been "train'd up according to a certain discipline and education" in the practices of property, transfer, and promise, what he says is not that this consists in any ground mentioned in the rules, or in (what just persons take to be) the rules' (conditional) authority. Rather, Hume cites as the "civiliz'd" motive to justice in those who "have the least grain of honesty, or sense of duty and obligation," the "regard to justice, and abhorrence of villainy and knavery." (T.479) And by this he means the approbation of justice and disapprobation of injustice, i.e., what he later calls the moral obligation to justice. (T.498, see also T.483, T.518) As I shall argue in Section VII below, however, since the direct object of approbation or disapprobation is always some mental quality ("all virtuous actions derive their merit only from virtuous motives" (T.478)), the "regard to justice and abhorrence of villainy and knavery," must be directed towards the virtue and vice of justice and injustice, respectively. It will follow that, unless justice involves "a sophistry" (T.483), the virtue of justice must itself involve some motivational state other than the regard to justice.³³ This, I shall ultimately argue, is nothing other than the just person's regulating her conduct by the rules of justice.³⁴

V. Three Species of Obligation

We are now in a position to appreciate the sense Hume must have in mind when he says that the concept of obligation arises only *after* the conventions establishing the rules of justice have been "enter'd into." (T.490–491)³⁵ As I mentioned at the outset, neither natural obligation (the motive of self-interest) nor moral obligation (the sentiment of approbation) requires background conventions to be intelligible. Granted, the only place Hume uses 'natural obligation' is in connection with justice, but that is not the only place he uses the concept. Twice he refers to the motive of self-interest as obligating in contexts that have nothing to do with conventions and justice. (T.312,314) And even if he had not called the motive of self-interest an obligation in any other context, it would not follow that the idea somehow depends on background conventions. Likewise with moral obligation. Hume uses 'moral obligation' in a context that has no special connection to the justice-conventions (T.479), but again, it wouldn't matter if he hadn't. The concept he has in mind evidently depends on convention in no way. "[W]hen any action, or quality of the mind, pleases us *after a certain manner*, we say it is

virtuous; and when the neglect, or non-performance of it, displeases us *after a like manner*, we say that we lie under an obligation to perform it." (T.517) The moral obligation is created by an approving or disapproving moral sentiment. And even if the only thing moral sense ever disapproved were violation of the rules of justice, it would not follow that it is necessary to explain the conventional origins of justice in order to explain the idea of moral obligation.³⁶ After all, Hume borrows the ideas of natural and moral obligation from Hutcheson, and Hutcheson's treatment gives convention no role whatever.

So what can Hume mean? He must have some third concept in mind. And we can see pretty clearly what it must be; he must be referring to a notion of obligation defined through the rules of justice themselves. Rules are in their nature directive. The rule of property, for example, directs agents to abstain from others' possessions. It requires that agents so abstain, and if they do not they violate the rule. We might say, meaning nothing more, that it obligates them to abstain.

Here we should recall Hart's distinction between the internal and external aspects of rules.³⁷ From an external point of view, we may say that the rule of property obligates abstinence from others' possessions, meaning only to convey the content of the rule without endorsement. But if we *accept* the rule or norm as a basis for regulating conduct, then from an internal standpoint we shall not simply be reporting that this is what the rule requires, we shall be saying, as we might put it, that this is what the person really ought (or is obligated) to do. The external claim requires a background rule in the obvious sense that it is a claim *about* a rule (from the outside). Were there no rule of property, for example, there would be no rule-obligation to report. But internal standpoint claims do not similarly refer to rules as part of their very content. They assert nothing about any rule. Someone who accepts the rule or norm of property, and says from the internal standpoint that people are obligated to abstain from the possessions of others, makes a normative claim that *expresses* the rule, or perhaps, his acceptance of it.³⁸ But a claim with the very same content could be made were there no *established* rule of property or conventions of justice at all.

An external standpoint rule-obligation claim, therefore, requires reference to a rule for its very intelligibility, but an internal standpoint claim does not. Still, we can understand why Hume would think there is a sense in which even the normative claims that agents make when they accept a rule are not really intelligible without an existing background convention. Hume clearly regards the attitude self-regulators take towards the rules of justice as *reasonable* only so long as enough others share it. And what makes this possible, he thinks, is the publicly expressed common sense that it serves mutual advantage, i.e., the existence of what he calls a convention.³⁹

For Hutcheson, recall, there are no act-regarding virtues. All virtuous motives are "kind affections," desires for the good of others. And any rational motive is a desire for some good to self or others. Humean justice, however, aims at no

condition or state to which it regards action as instrumental, but at action, or the governance of action, itself. This is, as I have noted, a significant departure from Hutcheson.⁴⁰ But it is also important that for Hume it is the *only* act-regarding virtue, because it is the only case in which the motivational state of rule-regulation is supported by mutually advantageous conventions. No such state is natural; it arises only as a result of a convention, through publicly expressed mutual advantage.

Hume, then, recognizes three species of obligation: those he calls “natural” and “moral”, and that we have called rule-obligation. While, on its face, the third concept has nothing essentially to do with rational motive, I shall be claiming that a central piece of Hume’s account of justice is that just persons are rationally motivated by their acceptance of the rules of justice, and that this motivation is not itself a species of any desire for the good, either for themselves or generally.⁴¹

VI. Justice, Motive, Rule, and Good

Again, Hume clearly says that rules of justice are established only when individuals come to *regulate* their conduct by these rules (T.490,499,532,533,534). They “lay themselves under the restraint of such rules,” (T.499) “impose” them on themselves (T.533), and “suppos[e them] to be sacred and inviolable.” (T.533) It is hard to see how to interpret these remarks in any way other than as entailing that rules of justice are established only when individuals come to treat rules of justice as having authority in their deliberations, when, that is, they give them independent weight (indeed conclusive independent weight), thereby deliberating on a basis different from their own good (or, for that matter, the public good). Nonetheless, Hume appears to take the position in the *Treatise* that what the rules of justice require is invariably extensionally equivalent with the dictates of enlightened self-interest. If so, then despite their intensional inequivalence, Hume seems to believe in the *Treatise* that enlightened self-interest invariably leads to compliance with rules of justice, so that agents can actually “regulate” themselves by these rules without giving them independent deliberative weight.

Hume certainly there says that it is only because each person can expect to gain by the existence of rule-structured practices of property, transfer, and promise that each has reason to establish them.

To the imposition then, and observance of these rules, both in general, *and in every particular instance*, they are at first mov’d only by a regard to interest; and this motive, on the first formation of society, is sufficiently strong and forcible. (T.499, emphasis added)

Once the “common sense of interest” is mutually express’d, moreover, no more is needed “to induce any one of them to perform an act of justice, who has the first opportunity,” (T.498) Each sees that doing so will become an example of

good will to those equally prepared to follow the rules if others do, and this makes it advantageous for her to do so. "On the first formation of society," when each appreciates the "disorder and confusion" that result from "every breach of these rules," self-interest dictates observing the rules "in every particular instance." (T.499) When the circle of interagents is sufficiently small, and the rules insufficiently well established, the personal costs of flouting them are likely to be high; the situation will resemble, not a "one shot" Prisoner's Dilemma, but an iterated form of that game, changing the payoffs to make justice and self-interest coincide.⁴²

Were this situation to continue, agents could successfully comply with the rules of justice by following their own interests. Hume does not expect it to, but in the *Treatise*, at any rate, the reason he doesn't is apparently not that he thinks that when "society has become numerous, and has increas'd to a tribe or nation," individuals can sometimes actually gain by violations, but because he thinks that the interests individuals have in maintaining order are there "more remote" and ones they "may frequently lose sight of," so that they come to follow "a lesser and more present interest." (T.499)⁴³ In a larger, more anonymous setting,

[t]he consequences of every breach of equity seem to lie very remote, and are not able to counterballance any immediate advantage, that may be reap'd from it. They are, however, never the less real for being remote... (T.535)

Before the mutual acknowledgment of a shared interest in establishing rule-structured practices of property, transfer, and promise, self-interest "is the source of all injustice and violence." (T.480) But once there is a convention, "a general sense of common interest," this is sufficient to make self-interest the servant of justice. When society becomes sufficiently numerous, individuals may tend to lose sight of their real interests, but these continue to dictate abiding by the rules of justice in every case.

Because Hume stresses that the primary benefits of justice for individuals are benefits they reap from the *practice*, from the "whole plan or scheme," (T.497) and because he says that "single acts of justice may be contrary, either to public or private interest," (T.497) it may seem that he is willing to allow in the *Treatise* that justice may occasionally conflict with the agent's interest, that it is possible for individuals to act unjustly and "free ride" on the just actions of others. In this context, however, Hume is excluding from the consequences of individual just acts any compliance they may cause by others.⁴⁴ "A single act of justice is frequently contrary to *public interest*; and were it to stand alone, without being follow'd by other acts, may, in itself, be very prejudicial to society." (T.497) So the "somewhat singular" connection Hume asserts in the *Treatise* between the rules of justice and self-interest is not, as we might expect from recent discussions of the free rider problem, that though the "whole plan or scheme" is essential to promoting mutual advantage, individual acts of justice may be con-

trary either to the agent's or to the public interest. Hume's point at T.497 is that, considering all consequences other than those that run through the rule-structured practices of property, transfer, and promise, individual acts of justice may do more harm than good. Once we include these practice-mediated effects, however—e.g., that trust may be threatened and therefore others may be less likely to comply—then, Hume appears to believe in the *Treatise*, it is invariably in the agent's interest to comply. The “singular connection” is that it is only because others are likely to comply only if one does—that their compliance with rules of justice is among the consequences of one's compliance—that each individual agent's complying with the rules of justice invariably has the best consequences.

As far as the *Treatise* goes, therefore, it would appear that we need not take Hume's talk of agents regulating themselves by the rules of justice too seriously. Compliance with these, or with what I have called the “meta-rule of cooperation,” is invariably dictated by self-interest once the common sense of interest in establishing just practices has been mutually expressed. Individuals comply with these rules only because others do, but this latter consideration functions for them not as a condition for giving intrinsic weight to these rules, nor as the antecedent of an intrinsically weighty meta-rule of cooperation, but as the fact that one had better comply since others are likely to comply only if one does, the good consequences of their compliance being consequences of one's compliance as well.

VII. Moral Obligation, Motive, and Act

Still, even in the *Treatise*, Hume says that just practices will be stable among interagents sufficiently numerous to constitute a “tribe or nation” only if they have some other motive to comply. Self-interest will continue to dictate compliance, but when the bad consequences of injustice are sufficiently remote, persons are far likelier to be attracted by “lesser and more present interest[s].” (T.499) So some other motive is necessary. The problem is that Hume says that the requisite motive is the sense of duty or moral obligation, and, as I have said, this requires some further motive as its object.

Indeed, Hume holds that the sense of duty is not simply a fail safe, but the motive uniquely appropriate to justice. “[W]e have naturally no real or universal motive for observing the laws of equity,” he writes, “but the very equity and merit of that observance.” (T.483)⁴⁵ And:

'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.518)

Justice's having merit, being a duty, and being morally obligatory all come to the same thing for Hume, viz., that justice is the object of moral approbation and injustice the object of moral disapprobation. (T.477,498–500,517) But approba-

tion and disapprobation always take some "mental quality" as direct object. An "external performance"—for example, an action that complies (extensionally) with a rule of justice—"has no merit" in itself. So a just action can only acquire merit derivatively, by evidencing a meritorious motive. (T.483–484) It follows that "no action can be equitable or meritorious, where it cannot arise from some separate motive," (T.483) i.e., some motive other than the sense of merit, duty, or moral obligation.

These ideas seem to be in irreconcilable conflict. How can Hume both hold the fundamental thesis of his virtue ethics, that the sense of duty cannot be the "first virtuous motive" (T.478) of justice or any other virtue, and also think that it is the only "real or universal motive for observing the laws of equity"? (T.483) This conflict has led commentators to the conclusion that Hume cannot really mean to maintain his distinctive virtue ethic when it comes to justice. Although with the natural virtues, it is clearly the case that what we approve is some motive or mental quality that might be manifested in various actions, when it comes to the artificial virtues, what is approved in the first instance, and thus morally obligatory, are *actions* themselves, viz., those that comply with the rules of justice.⁴⁶ In this section, I shall argue that this interpretation faces numerous obstacles. To avoid these, justice must be treated, like every Humean virtue, as having moral qualities only as "a mental quality," specifically, as a motivational state. Seeing things this way, however, will leave the substantial question what this motivational state can be, especially since what is distinctive about justice in Hume's view is the way a sense of obligation is supposed to enter into it.

On any reasonable interpretation of Hume's ideas, the moral obligation to justice is supposed to derive from the fact that sympathy with the usual effects of justice and injustice leads to moral approbation of justice and moral disapprobation of injustice, respectively. (T.479, see also T.483,518) And it is uncontroversial that the relevant consequences are (at least primarily) consequences of acts of justice and injustice. What is at issue is whether Hume thinks that, uniquely with justice (and the artificial virtues more generally), the proper object of moral approbation and disapprobation are kinds of acts. Or whether he holds that with both artificial and natural virtues the object of moral sentiment is a mental quality of which an "external performance" is only a sign.

Since Hume puts forward the thesis that "all virtuous actions derive their merit only from virtuous motives, and are consider'd merely as signs of those motives," right at the outset of Part II ("Of Justice and Injustice"), and since it is the centerpiece of his argument that justice is an artificial virtue, it is hard to see how he can mean it not to apply to the case of justice. (T.478) But perhaps we should not conclude from this directly that Hume denies that acts of justice and injustice can be morally approved or disapproved in themselves. After all, common sense distinguishes analytically between evaluations of persons and their characters (and of actions as signs of these), on the one hand, and evaluations of acts as things to do, on the other. Hume could be saying that "merit" concerns the former while believing that another category of moral evaluation concerns the

latter. This thought might be encouraged by the account he gives of moral obligation in his discussion of promising:

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

Perhaps his view is that while merit, virtue, and vice all concern mental qualities of the agent in the first instance, *moral obligation* primarily concerns acts—“external performances”. And if just *acts* can be morally obligatory, then perhaps the just person is moved actually to follow the rules of justice by this conviction or sentiment.

This passage notwithstanding, however, Hume makes no principled distinction between the categories of virtue and vice, merit and demerit, on the one hand, and duty and obligation, on the other. Hume does not say that the (moral) motive to justice is the sense of duty or obligation as opposed to a sense of merit and virtue. Rather: “the very equity and merit” of just acts is the only “real or universal motive for observing the laws of equity.” (T.483) And he includes the category of moral obligation within the scope of the thesis that links the moral quality of an action to that of its motive:

But tho’, on some occasions, a person may perform an action merely out of regard to its moral obligation, yet still this supposes in human nature some distinct principles, which are capable of producing the action, and whose moral beauty renders the action meritorious. (T.479)

Whenever we require an action, he says, “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.” (T.478)

How thoroughgoing Hume’s commitment is to this thesis can be illustrated by his discussion of the rare instance in which “the sense of morality or duty [can] produce an action, without any other motive.” (T.479) Whenever we approve of a motive or principle in human nature,

a person, who feels his heart devoid of that principle, may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty, in order to acquire by practice, that virtuous principle, or at least, to disguise to himself, as much as possible, his want of it. (T.479)

By “the action” here, Hume evidently means an action characteristic, and thus ordinarily a sign, of a virtuous motive or “mental quality.” If what the sense of duty (later in the paragraph: “regard to its moral obligation”) motivates is acting *in order to* acquire a virtuous principle by practice or to sustain ignorance of its lack, then the object of moral duty or obligation must be having the virtuous principle itself.

Were Hume to maintain that just acts are literally morally obligatory (by his definition) independently of the agent's state, and that this approbation is what directly motivates just acts after the "first formulation of society," he would face numerous problems. Most obviously, he would have to abandon the fundamental principle of his virtue ethics, that all merit derives from "mental qualities." But second, it is not obvious that Hume could count a moral approbation of just acts that was independent of their relation to a state of the agent as any kind of motive to justice. One reason is that the way Humean approbation motivates is indirectly, through the agent's hatred or esteem of himself on account of his qualities.⁴⁷ Approbation or disapprobation are "nothing but a fainter and more imperceptible love or hatred." (T.614) And only insofar as the objects of approbation and disapprobation are qualities in the agent will they be related to love and hatred in the right way. Knowing that an agent performed a kind of act we regard unfavorably does not yet connect the act to her in a way that reflects on her as an appropriate object of love or hatred. Thus, while Hume is willing to allow that approbation and disapprobation can as a psychological matter be transferred by association from thing signified to sign (T.479), in order to motivate through self-esteem or self-hatred, the object of moral sentiment must be some quality in the person.⁴⁸

Additionally, in a manuscript amendment to the *Treatise's* original edition, Hume makes it clear he thinks that moral approval of justice is an insufficient motive to just acts in any case.

Thus *Self-interest* is the original Motive to the *Establishment* of Justice: but a Sympathy with *public* Interest is the Source of the *moral* Approbation, which attends that Virtue. This latter Principle of Sympathy is too weak to controul our Passions; but has sufficient Force to influence our Taste, and give us the Sentiments of Approbation or Blame. (T.670, see also T.586)⁴⁹

A third and final point. The idea that moral approbation of just acts can directly motivate agents conflicts no less with Hume's official theory of the will than does the suggestion that agents are motivated by their acceptance of mutually advantageous rules. If will exerts itself only when "the good or the absence of the evil may be attain'd" by an action, then it will not be moved by moral obligation *per se*. At most, the agent will be moved to follow the rules of justice by the thought that doing so will enable him to avoid painful contemplation of (his own) injustice. If somehow he could do injustice without having to contemplate it, this would give him no reason not to.

VIII. Justice as an Agent-state

If the moral obligation to justice derives from our approbation and disapprobation of justice and injustice as states of the agent, respectively, the question obviously arises, which states are they? Hume never says explicitly, so the

interpretive problem is to work out what it might be indirectly.⁵⁰ David Gauthier suggests that in the *Treatise* Hume holds that justice as a motivational state of the agent is nothing other than self-love (“the interested affection”) “redirected towards its fuller satisfaction through its own restraint by the conventionally instituted laws of society.”⁵¹ Hume rejects the possibility that “a concern for our private interest or reputation is the legitimate motive to all honest actions” (T.480) in arguing that justice is an artificial virtue, but that is on the grounds that *before* the convention arises “self-love...is the source of all injustice and violence.” (T.480) And, as we’ve seen, Hume holds in the *Treatise* that once the sense of common interest has been mutually expressed, a person appropriately mindful of justice’s long-term benefits, and able to weigh them properly with “lesser and more present interests” (T.499), will always have an adequate motive of interest to comply with its rules. Hume has a name for the trait that enables a person to act in her greater long term interest when that requires foregoing lesser, more immediate interests: *strength of mind* (T.418, see also, E.205,239)⁵²

Suppose we take Hume to hold that suitably enlightened self-interest, or strength of mind, is the motivational state distinctive of the just person. If self-interest invariably dictates compliance with the rules of justice, as Hume appears to believe in the *Treatise*, then a fully-informed, strong-minded person will invariably comply. Moreover, we can understand on this hypothesis how someone lacking this trait might be moved by the sense of duty or moral obligation to act justly in situations where he otherwise would not. If lesser, more immediate interests incline him to act unjustly he may still reflect, contemplate his lack of a strong mind, and “hate himself upon that account.” (T.478) And this will give him an additional, and more immediate interest in acting justly than he would have had if he didn’t disapprove weak-mindedness, since by so acting he will be encouraging a trait of which he would be proud, “or, at least, disguis[ing] to himself, as much as possible, his want of it.” (T.479)

Now while this shows how an obligation to justice other than the natural obligation (self-interest) can figure in motivating justice as a backup if the virtue of justice is strength of mind, it does raise various puzzles, even in the *Treatise*. First, although Hume’s general view is that virtues include various kinds of “mental qualities,” he says at the beginning of his discussion of justice that “all virtuous actions derive their merit from virtuous *motives*.” (T.478, emphasis added) This suggests that the agent-state of justice involves a distinctive motive to just acts. But on the current proposal, the motive of just acts is simply enlightened self-love.

Second, it is hard to explain on this proposal the prominence Hume gives to a distinctive motivational source (and thus obligation) in the case of justice and the other artificial virtues. I have already mentioned Hume’s remarks in the *Treatise* that the sense of moral obligation is the only “real or universal motive for observing the laws of equity,” (T.483)⁵³ and that “if we thought promises had no moral obligation, we never shou’d feel any inclination to observe them.” (T.518)

In "Of the Original Contract," Hume makes clear that what he has in mind here is nothing less than a fundamental difference between natural and artificial virtues:

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.⁵⁴

If Hume thinks strength of mind is the trait we admire in the just person, and that enlightened self-interest invariably dictates complying with rules of justice, then it is odd that in his description of the just person's practical thinking he would give a prominent role to a sense of any obligation to justice other than the "natural obligation" of self-interest.⁵⁵

Third, it is a mystery why Hume should say that there is *any* idea of obligation that is "unintelligible" before the "convention, concerning abstinence from the possessions of others, is enter'd into." (T.490) As I've argued, neither of the ideas of natural obligation or moral obligation require conventions to be understood. On the present suggestion, the idea of natural obligation would be adequate to capture the practical thinking of the just person, and that of moral obligation would suitably explain the evaluative thoughts of observers of justice and injustice, and of agents as observers of themselves. So why does Hume think there is any idea of an obligation to justice that requires the existence of convention in order to be understood?

In Section V, I claimed that Hume's remarks about the unintelligibility of obligation before the existence of a convention concerning property make sense if we take him to be speaking of *norm-* or *rule-obligation*. The convention in question is a sense of common interest in regulating conduct by rules of justice (or perhaps, by what I called the meta-rule of cooperation). Each communicates to the other his sense that it is in his interest to accept and abide by the rule of abstinence "*provided* he will act in the same manner with regard to me." (T.490)⁵⁶ Again, this is not a prediction that if others abstain, then it will be in one's interest to do so. Of course, Hume believes that it is in each person's interest to encourage others to abstain by abstaining as an example, but that is a different matter. Rather, it is an expression that it is one's interest to accept and follow a conditional rule or norm: abstain if others do. On this hypothesis, there would indeed be a sense of 'obligation' for which Hume believes people would have no use before the convention, before the expression of common interest in regulating themselves by the rule, viz., the idea of rule-obligation.

Still, although Hume speaks as though participants in a rule-structured practice of justice genuinely regulate themselves by rules they accept, regarding them as authoritative (and overriding), what he really seems to think in the *Treatise* is that self-interest always provides an adequate deliberative basis. Only meta-

phorically do just persons “regulate” themselves by the rules; their motive is always given by enlightened self-interest.

IX. Justice and the Sensible Knave

In the *Enquiry*, however, Hume is no longer prepared to say that justice is invariably advantageous:

a man, taking things in a certain light, may often seem to be a loser by his integrity. And although 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. (E.282)

Hume now appreciates what we have come to call the “free rider problem.” Or, at least, its appearance. People “may think” that, occasionally, they benefit more by injustice, even in the long run, since, owing to the “imperfect way in which human affairs are conducted,” a single injustice may not cause “any considerable breach in the social union and confederacy.”

Nor, Hume must believe, is this *mere* appearance. If it were, he would answer the knave differently than he does. The knave’s position is to regard “honesty is the best policy” as a “good general rule” that is nonetheless “liable to many exceptions,” and that “he...conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions.” (E.282–283) If Hume really believed at this point that, despite appearances, justice is invariably advantageous, he would try to correct the knave’s beliefs about the consequences of injustice. Indeed, even if Hume believed the knave is right that particular unjust acts may actually be advantageous, but nonetheless thought it disadvantageous, because too risky, to execute a policy of making exceptions to the rules when doing so appears advantageous, he could still recommend inflexible conformance to the rules of justice solely on the grounds that it is the most advantageous general policy. But Hume does neither of these things. While he does point to ways in which knaves may be betrayed by their own frailties, the overall burden of his response is quite different. Hume despairs of adequately answering a knave who can regard the rules of justice as no more than strategically valuable.

If his heart rebel not against such pernicious maxims, if he feel no reluctance to the thoughts of villainy or baseness, he has indeed lost a considerable motive to virtue. (E.283)

If, that is, there were no moral obligation to justice—if normal human beings did not disapprove of knavery—then we might lack an adequate motive to justice. It is only the moral sentiment and our tendency to love and hate ourselves on the

strength of our virtue and vice that give us moral interests that make the just life advantageous: "Inward peace of mind, consciousness of integrity, a satisfactory review of our own conduct; these are circumstances, very requisite to happiness, and will be cherished and cultivated by every honest man, who feels the importance of them." (E.283, see also T.501).

Hume does not say why the uncertainties and dangers inherent in the knave's strategy would be insufficient in themselves to warrant treating the rules of justice as inflexible constraints, but we can see some problems that would develop. Even if an agent lacking moral interests can be convinced she would be better off were she so to treat the rules of justice as a matter of general policy, there will inevitably be cases where the most advantageous policy will dictate what she reasonably believes to be less advantageous acts. How, in such cases, is she to deliberate? She knows that she is better off in general if she deliberates by justice rule-regulation. But she may also have very good reason for thinking that in this case she will be better off by violating. The very reasons that recommend that she generally deliberate by rule-regulation, recommend that she violate. In such circumstances, her practical convictions may well prove unstable.

Moral interests counteract this instability. Prone to self-esteem or self-hatred through inescapable reflection on our own character, and deeply concerned with the approval of others, we have more at stake in violating the rules of justice. "After the opinion, *that a merit or demerit attends justice or injustice*, is once firmly establish'd among mankind," the interest in our reputation (than which nothing "touches us more nearly") comes to depend more on "our conduct, with relation to the property of others" than on anything else. (T.501) "For this reason," Hume concludes, "every one, who has any regard to his character, or who intends to live on good terms with mankind," must fix the rules of justice as an "inviolable law." (T.501) Likewise, if we did not think that infidelity to promises was vicious (contrary to moral obligation), "we never shou'd feel any inclination to observe them." (T.518)⁵⁷ But because we do, anyone who makes use of the conventional signs of promising "is immediately bound by his interest to execute his engagements." (T.522) The moral interests in integrity, reputation, and so on, that support regulating conduct by the rules of justice, therefore, do not also support acts that violate these rules.

Thus while in the *Treatise* Hume appears to believe that self-interest invariably dictates justice, by the time of the *Enquiry* he has apparently abandoned this view, having come to think that, with large numbers, occasional injustice can be advantageous owing to the free rider problem. Any doubt about Hume's having recognized the free rider problem (finally, at least) is removed by the following passage from the "The Origin of Government":

All men are sensible of the necessity of justice to maintain peace and order; and all men are sensible of the necessity of peace and order for the maintenance of society. Yet, notwithstanding this strong and obvious necessity, such is the frailty and pervers-

seness of our nature! ... Some extraordinary circumstances may happen, in which a man finds his interests to be more promoted by fraud or rapine, than hurt by the breach which his injustice makes in the social union. But much more frequently, he is seduced from his great and important, but distant interests, by the allurements of present, though often very frivolous temptations.⁵⁸

But if this is so, the virtue of justice cannot be strength of mind, since that will sometimes dictate injustice. Adding in the moral interest in avoiding self-hatred does not help, since that interest exists only if the approved agent-state (virtuous motive) would issue in the (virtuously just) action independently. And Hume's response to the knave presupposes that he no longer believes that enlightened self-interest leads to "regulation" by the rules of justice.

So the problem re-emerges: what agent-state is the virtue of justice? Even apart from the "sensible knave" passage there are several problems with taking it to be strength of mind, as we noted at the end of Section VIII. And that passage is flatly inconsistent with the proposal, in any case.⁵⁹

Hume continues in the *Enquiry* to emphasize his theme that justice involves *regulation by rules*. Rules "found requisite" for society's subsistence are "immediately embraced." (E.192, see also E.193) And men's "understanding and experience tell them that [their] combination is impossible where each governs himself by no rule, and pays no regard to the possessions of others. ..." (E.307, see also E.305,306) But if self-interest does not invariably dictate compliance, Hume's claims that just practices are established when "the whole scheme or system [is] concurred in by the whole, or the greater part of society," (E.304) and that this involves just persons regulating their conduct by the rules of justice, "lay[ing] themselves under the restraint of such rules" (T.499) and regarding them as "sacred and inviolable," now commit him to the virtue of justice being a distinct motivational state of the agent than enlightened self-interest or strength of mind. Since he cannot now think that rule-regulation is reducible to the pursuit of interests in maintaining rule-structured practices viewed *externally*, Hume's continuing talk of acceptance of ("embrac[ing]" (E.192) and regulation by rules must now be taken seriously. And this requires interpreting Hume as holding that just persons regard the rules *internally* as agents. They take them to have a normative relevance to their conduct distinct from a consideration of any good or evil that "may be attain'd by any action of the mind or body." (T.439)

Just persons acquire the "habit of justice." (E.203) They are determined to follow the rules of justice, and although their original motive for being so determined is self-interest, they continue "without recalling, on every occasion, the reflections, which determined [them]." (E.203) But rules of justice enter into the determination of their conduct in a way that is different from the way general rules and habit usually operate in Humean psychology, where "we extend our motives beyond those very circumstances, which gave rise to them, and form something like *general rules* for our conduct." (T.531) Usually "these rules are

not perfectly inflexible, but allow of many exceptions.” (T.531) The rules of justice, however, “are unchangeable... by particular views of private or public interest.” (T.532) They “are artificially invented for a certain purpose, and are contrary to the common principles of human nature, which accommodate themselves to circumstances, and have no stated invariable method of operation.” (T.532–533) The convention that establishes the rules of justice depends, indeed, on a mutual recognition of the “disorders that result from following their natural and variable principles,” including “natural” general rules. (T.533) And the convention, again, is the shared recognition of a mutual interest in *regulation* by the “inflexible” rules of justice. “I see evidently, that when any man imposes on himself general inflexible rules in his conduct with others, he *considers certain objects as their property, which he supposes to be sacred and inviolable.*” (T.533, emphasis added)

These thoughts could be developed in different ways. Perhaps justice (the agent-state) involves regulation by specific rules of property, transfer, and promise, and this is a virtue only if enough others have it. Or perhaps justice involves conditional regulation by these specific rules—conditional, that is, on enough others doing so also. Or perhaps just persons regulate themselves by a meta-rule of cooperation that dictates following mutually advantageous rules so long as others do. For our purposes, the differences between these do not matter. All have in common the notion that just persons regulate themselves by rules they regard as authoritative. And the crucial point is that by the time of the *Enquiry* there seems to be no way to understand such claims other than as positing an agent-state distinct from the desire to promote any good.⁶⁰

If we take the virtue of justice to consist in the agent’s regulating herself by the relevant rules (or her disposition to do so), we can solve various puzzles of Hume’s text. For example, Hume can thereby avoid the circle involved in claiming that, as with any virtue, there must be a “first virtuous motive” of justice other than the sense of duty, and that the sense of duty is the “only real or universal motive for observing the laws of equity.” If the virtue of justice is the motivational state of rule-regulation, then the object of the moral sentiment can be specified noncircularly. Justice, the virtue, consists in the trait of justice-rule-regulation as realized in the agent’s practical reasoning. And its being morally obligatory or virtuous consists in an observer’s approbation of this trait generated by Humean association and sympathy. In thinking about the beneficial effects of the trait (including of the “whole plan or scheme” of which it is an ineliminable part), sympathy turns ideas of contemplated pleasure into the pleasurable sentiment.

We can also explain on this hypothesis why Hume lapses into speaking as though just *acts* are morally obligatory *and* why he gives a substantial motivational role to the moral sentiment in explaining just action in our “civiliz’d state.” Hume believes that in the case of justice, and only here, is what is morally approved a motivational state whose direct object is a kind of act. Unlike such

natural virtues as benevolence or parental concern, the virtue of justice is no aspect of the heart, no emotion or desire that takes some state or condition (such as another's welfare) as object, and that arises only by involuntary response. A parent lacking in the naturally virtuous parental responses can drag himself out of bed in the middle of the night to tend a sick child out of the sense of moral obligation, according to Hume, only in the hope thereby to create circumstances in which virtuous parental feeling will in time arise. The virtue of justice, on the other hand, is a disposition, not to an emotional response that might be reflected in action, but to a form of practical reasoning that is realized in the decision to abstain from unjust acts on the grounds that it is required by the rules of justice. Someone lacking but approving the virtue, can be motivated by his moral interest in being just to follow the rule, to perform just acts, hoping to acquire "by practice" the stable trait of governing himself by it without appeal to this interest. Strictly speaking, what is morally obligatory is the agent's mental quality of rule-regulation, not just acts themselves. But the direct object of this state, unlike that of other virtuous principles, is a kind of act. Uniquely with justice, therefore, it is almost true that the just act is itself morally obligatory, regardless of its relation to a virtuous principle.

If we regard the moral obligation to justice as consisting in moral approbation of the "mental quality" of justice, and identify the latter with rule-regulation, we can explain why Hume would still *say* that the motivation to be just depends on the moral obligation to justice. The latter, on the present suggestion, is generated (through sympathy) by contemplating the beneficial consequences of general justice-rule-regulation. Holding psychology fixed, justice (rule-regulation) would fail to be approved only if its normal consequences were not beneficial. But if that were so, it could not be mutually advantageous. So a person disposed to act on mutually advantageous rules, so long as others do, would be disposed to act on the rules of justice only if justice is morally obligatory.

But this does not explain why Hume says that, uniquely with the artificial virtues and justice in particular, virtuous actions are "performed entirely from a sense of obligation." ("Original Contract," p. 480, see also T.483,518) It is an important part of my case, of course, that it *would* explain it, if he there meant "rule-obligation." But what he seems to have in mind in these passages is moral obligation: the sense of the "merit" of just action. (T.483)

The interpretation I am suggesting must attempt to explain this away as confusion. It is indeed crucial to Hume's distinction between the artificial and natural virtues that only the former are performed from a sense of obligation. But the relevant obligation can't be the moral obligation, because that depends on artifice and convention in no way. The notion of obligation that is "unintelligible" without convention is rule-obligation. The convention, recall, is the jointly acknowledged mutual interest in regulating conduct by the rules of justice, that is, in regarding the rules of justice as authoritative and obligating. So the relevant

virtue—justice as an agent-state—is realized when agents act because they so regard the rules, that is, from a sense of (rule-)obligation. The only sense of obligation, therefore, that can enter uniquely into artificial virtues is rule-obligation. Because it does not fit with his usual categories, however, Hume confusedly asserts that it is the moral sentiment that uniquely motivates artificially virtuous acts. And that is what lands him in the circle.

So far as I can see, all alternatives to this interpretation require Hume to hold that the virtue of justice involves error or a sense of “sophistry” that he thinks the insight that justice is conventional should enable us to see beyond. If, for example, we take Hume’s remarks that the sense of merit is the motive distinctive of justice, then we are either involved in the circle or we have to suppose that just actions can have merit, and thus be morally obligatory, as “external performances.”⁶¹ The former alternative is the sophistry of which we are to be disabused by understanding the conventional roots of justice. (T.483) And while the latter is given psychological support by an association of ideas between “thing signify’d” (virtuous motive) and “sign” (virtuous action), this also involves error since “the external performance has no merit.” (T.483,477) “To find the moral quality” of any action, “we must look within.” (T.477) We cannot do this “directly,” so we “fix our attention on actions, as on external signs.” (T.477)

Another possibility would be to hold that the virtue of justice consists, not in an accurately informed strength of mind, but either in a tendency to overestimate the disadvantages of injustice sufficiently so that, conjoined with a strong mind, it leads to invariable compliance with justice’s “inflexible” rules.⁶² Or perhaps, a tendency vividly to imagine and dwell on disadvantageous consequences of injustice out of proportion to their place in what (the agent accurately believes) to be her long-run interest.⁶³ Both alternatives implicate the just person in error. She either mistakenly believes that injustice is invariably disadvantageous, or she does not believe this, but dwells disproportionately in her practical thinking on distant disadvantages. The latter involves no cognitive mistake, only the same kind of vice (the mirror image) that Hume laments when people lacking strength of mind give greater weight in deliberation to lesser, more immediate interests.

Finally, there are Haakonssen’s and Gauthier’s error-theoretic proposals. Taking his cue from Hume’s curious description of promising as including a feigned willing of an obligation, Haakonssen constructs Hume’s position as follows.⁶⁴ In order for us to regard justice as a virtue we must suppose that there is a “first virtuous motive” to justice—one which we would approve and hate ourselves for lacking. But there is no such motive. So we feign its existence and hate ourselves for lacking it, with this then being the motive to act justly—either because we also imagine we might acquire this motive by practice or because we want to maintain the fiction that we have it. Gauthier rejects this account for the *Treatise*, since enlightened self-interest can apparently function there as the “first virtuous motive,” but suggests that the sensible knave passage makes something like it necessary. In order for the moral obligation to justice to help motivate just

actions, we must at least believe there to be a first virtuous motive of justice. The sensible knave passage commits Hume to denying that it can be enlightened self-interest, and this, Gauthier thinks, is the only real contender. The sole remaining possibility is that we pretend otherwise, hate ourselves for lacking the pretended motive, and act to convince ourselves either that we have it or that we are on the way to acquiring it.

The best way for Hume to avoid the possibility that justice involves some kind of error is to hold that the moral obligation to justice consists in moral approbation for the trait of justice, and that this consists in the just person's disposition to engage in a form of practical reasoning substantially different from any countenanced by his official theory of the will, viz., by regulating her conduct by rules she regards as authoritative. It is, as he puts it in "Of the Original Contract," "a regard to the property of others," that is morally obligatory, not acts considered independently of this regard.⁶⁵

X. Conclusion

On this interpretation there is a resultant congruence in the case of justice between the three things Hume identifies as forms of obligation. First, Hume's approach is distinguished most sharply from Hutcheson's by the idea that justice is realized by social practices that require a different concept of obligation from either Hutchesonian natural or moral obligation, viz., rule-obligation. The just person acts, not simply from desires for the good; she regulates her conduct by rules of justice. In accepting these, she regards action falling under them as what she ought or must do. She can, of course, step back from her acceptance of these norms and consider that critically. When she does so from the general point of view, the evident benefits of the "whole plan or scheme" lead to moral approval (the moral obligation). Critical endorsement results also from the perspective of her own good, not least because moral sense gives her interests in her own character. Rawls's distinction between justifying a practice and justifying a particular action falling under it can thus be drawn within the just agent's practical reasoning, as well as within the social practice it helps to realize.⁶⁶

I remarked above that early modern debates concerning obligation were connected to issues about the nature of the will. On the one hand, there was a powerful philosophical tendency to understand obligation as a constraint that operates on the will *internally*, from the agent's point of view in practical reasoning. But so long as the will was thought of as a kind of desire for, or "tend[ency] to unite with," the good (T.438), obligation had then to be reduced to such terms. On the other hand, writers like Pufendorf and Crusius resisted attempts to reduce obligation to the pursuit of good and avoidance of evil. But they understood obligation *externally*, as deriving from dictates of a will the agent recognizes as authoritative because superior to his own—a will that rightly subjugates him.

Kant later synthesized these two elements with the reciprocal doctrines that autonomous will involves self-direction under a conception of law and that

obligation derives from law the agent imposes on himself in autonomous practical reasoning.⁶⁷ This departs from the traditional, rational appetite theory of will, since it holds that agency requires a conception of authoritative rule of conduct or law, and that this cannot be reduced to consequential goods or evils. And it also diverges from Pufendorf and Crusius by locating the authority requisite for obligation inside the agent, within his own autonomous practical reasoning.

While Hume's official theory of the will is thoroughly traditional, his account of the agency of the just person, as I have interpreted him, is not. The just person regulates himself by rules (norms) he accepts. He abstains from the property of others not (just) because doing so advances goods he desires and avoids evils to which he is averse, but because he "embraces" the rule of property, and abstinence is what the rule of property requires. In accepting the rule, he recognizes it as authoritative and obligating.

In conclusion, it is worth remarking briefly some tantalizing similarities between this idea and Kant's doctrines. Hume's just person acts on a principle of conduct he regards as authoritative (a maxim, if you like) and not simply out of the desire for any consequential state. And his doing so is conditional on taking action on this principle itself to be something he can critically endorse. Of course, Hume's picture of the relevant reflective endorsement is no model of Kantian autonomy. The perspectives of self-interest and the general point of view are both "material" by Kant's lights. While Kantian autonomy is norm-based "all the way down," Humean reflective endorsement proceeds, respectively, from the perspectives of the agent's and the general good. And it is only contingent for Hume that these agree in any case. Still, in giving prominence to the idea of self-regulation, Hume's account of the obligation to be just marks a substantial departure from earlier, good-based accounts of obligation in the direction of the Kantian idea that obligation derives from a conception of authority located within the autonomous rational agent's will.

Notes

¹Earlier versions of this paper were presented to the Hume Society at the 1992 APA Central Division meetings in Louisville, Kentucky, to the 1992 summer meeting of the Society in Nantes, France, to a Hume Workshop at the University of North Carolina, to a faculty colloquium at the University of Michigan, and to the Philosophy Department at Harvard University. I am indebted to commentators on those occasions, Rachel Cohon, Jan Narveson, and J. B. Schneewind, and to members of the audience. I am also indebted to Charlotte Brown, William Frankena, Christine Korsgaard, Louis Loeb, David Fate Norton, Gerald Postema, T. M. Scanlon, and Robert Shaver, and to Amy Wesa, for help in checking references.

References to the following will be placed parenthetically in the text:

David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd ed., with text revised by P. H. Nidditch (Oxford: Clarendon Press, 1978). (T.)

David Hume, *An Enquiry Concerning the Principles of Morals*, in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 3rd ed., with text revised by P. H. Nidditch (Oxford: Clarendon Press, 1975). (E.)

Francis Hutcheson, *An Inquiry Into the Original of Our Ideas of Beauty and Virtue*, 4th ed. (London, 1738). References will be to section and article number, and will be to this edition, unless otherwise noted. (In.)

Francis Hutcheson, *An Essay on the Nature and Conduct of the Passions and Affections with Illustrations on the Moral Sense*, 3rd ed. (London, 1742). References will be to section and article number, and will be to this edition unless otherwise noted. (P.) (II.)

²On Kant's relation to modern natural law, see J. B. Schneewind, "Kant and Natural Law Ethics," *Ethics*, forthcoming. For his relation to Shaftesbury and Hutcheson see, e.g., Christine Korsgaard, "Kant's Analysis of Obligation: The Argument of Foundations I," *The Monist* (1989): 312–313. These were not, of course, the only traditions to which Kant was responding; he was also reacting against the perfectionism of Leibniz and Wolff.

³Actually, Hobbes distinguishes between obligation and its "tie" or bond. While he officially defines the former in terms of the transfer and abdication of rights, he accounts for obligation's bond through the practical necessity of necessary means to human agents' inescapable end of self-preservation. It can be shown, moreover, that in *De Cive* and in *Leviathan* Hobbes attempts to provide (differing) naturalistic reductions of obligation proper to contingently inescapable features of an agent's practical situation. Locke's official view is that moral obligation consists in the commands of a (divine) superior authority. Nonetheless, he also believes that such commands can exist only if agents have rational motives to comply (which God creates through the eternal joys and torment of Heaven and Hell), and a capacity for self-determining practical reasoning that engages these. In his (mostly unpublished) manuscripts on freedom of the will, Cudworth agrees that obligation requires a capacity of self-determination which, if properly exercised, engages deliberatively-conclusive motives for compliance. In addition, he argues, moral obligation requires that these motives be intrinsically moral.

I discuss the views of these writers, along with Cumberland, Shaftesbury, Hutcheson, Butler, and Hume, in *The British Moralists and the Internal 'Ought': 1640–1740*, (Cambridge University Press, forthcoming). For a brief discussion, see S. Darwall, "Obligation and Motive in the British Moralists," *Social Philosophy & Policy* 7 (1989): 133–150.

⁴*An Inquiry concerning Virtue*, II.i.i. First published in an unauthorized edition in 1699, Shaftesbury included this as a part of *Characteristics of Men, Manners, Opinions, and Times* (London, 1711).

⁵On this aspect of Grotius's ideas see Richard Tuck, "The 'modern' theory of natural law," in Anthony Pagden, *The Languages of Political Theory in Early Modern Europe* (Cambridge: Cambridge University Press, 1987), pp. 99–119; for Pufendorf, see J. B. Schneewind, "Pufendorf's Place in the History of Ethics," *Synthese* 72 (1987): 123–155, and "Modern Moral Philosophy: from Beginning to End?" in *Philosophical Imagination and Cultural Memory*, ed. Patricia Cross (Durham, N. C.: Duke University Press, 1992) and "Kant and Natural Law Ethics." See also Knud Haakonssen, "Divine/Natural Law Theories in Ethics," in *The Cambridge History of Seventeenth-Century Philosophy*, ed. Michael Ayers and Daniel Garber (Cambridge: Cambridge University Press, 1992).

⁶Although the conception of morality as involving action-guides that are distinct from (and potentially conflicting with) prudence occasions this question, this does not mean that the question "how does morality obligate?" was itself a question about the relation of morality to self-interest. Far from it.

⁷For some important qualifications, see note 3.

⁸Unlike Hutcheson's and Hume's, Shaftesbury's moral sentimentalism was more rationalist than empiricist. I discuss this in *The British Moralists and the Internal 'Ought'*.

⁹Bishop Richard Cumberland, *A Treatise of the Laws of Nature*, trans. J. Maxwell, (London, 1727), p. 233.

¹⁰*de Jure Naturae* (1672) I.i.21 (*On the Law of Nature and of Nations*, trans. Oldfather and Oldfather (Oxford: Oxford University Press, 1934), p. 20). For the following discussion of Pufendorf I am indebted to Schneewind, "Pufendorf's Place in the History of Ethics," and "Kant and Natural Law Ethics," and to Haakonssen, "Divine/Natural Law Theories in Ethics."

¹¹There is, of course, a difference between the practical necessity of an agent's either taking a necessary means to an end (or giving up the end) and the natural necessity that the end will be achieved only if the means are taken.

¹²*de Jure Naturae*, I.vi.5, I.vi.9 (pp. 90–91,95). Locke takes a similar line in his *Essays on the Law of Nature*, ed. W. von Leyden (Oxford: Oxford University Press, 1954). While he retains the same official theory in the Essay, other more naturalistic elements assume a much more prominent role in his thinking there.

¹³*Ibid.*, I.vi.5, p. 91, III.iv.6, p.386.

¹⁴Christian August Crusius, *Anweisung, vernünftig zu leben (Guide to Rational Living)* (1744). The passages come from an excerpt translated by J. B. Schneewind in his *Moral Philosophy from Montaigne to Kant* (Cambridge: Cambridge University Press, 1990), v. ii, pp. 577–581.

¹⁵At least this is so when what is approved is a “virtuous action.” (T.478) Other “mental qualities” than motives can be approved also.

On the intrinsic motivating power of moral sentiment, it is worth noting that Hume remarks that moral sentiment is “a fainter or more imperceptible love or hatred,” (T.614) and that love and hatred are “always follow’d by” motivating desires, but do not themselves include them. (T.367) On the impotence of the moral sentiment to motivate directly, see Charlotte Brown, “Is Hume An Internalist?,” *Journal of the History of Philosophy*, 26 (1988): 69–87. I discuss this further in Section VII.

¹⁶Cf.: “we have naturally no real or universal motive for observing the laws of equity, but the very equity and merit of that observance.” (T.483, see also T.479)

¹⁷Hume’s manuscript amendment to the original edition of the *Treatise* adds ‘naturally’, which Nidditch includes in his edition thus: “we have naturally no real or universal motive... .” This is lacking in the Selby-Bigge edition.

Since Hume concludes that “unless...nature has establish’d a sophistry,” the sense of justice must be artificial rather than natural, we can see why he was drawn to add ‘naturally’ where he does. But it really doesn’t help. So long as the motive that renders an act equitable or meritorious is regard to equity or merit, this is “evident sophistry and reasoning in a circle” whether the source of the motive be natural or artificial. (T.483)

I have been helped here by reading some papers of Rachel Cohon’s, including “Hume’s Difficulty with the Virtue of Honesty,” unpublished manuscript.

¹⁸Hume substitutes ‘interested obligation’ for ‘natural obligation’ in the *Enquiry*. (E.278)

¹⁹From Shaftesbury, Hutcheson took (and passed on to Hume) the doctrines that there is a natural moral good that attaches most fundamentally to motives; that the morality of acts derives from that of motives; that moral goodness or virtue depends upon a moral sense, activated when contemplating motives; that this moral sense arises naturally and not from any specific acculturation; and that the virtuous affections approved of by moral sense prominently include benevolent concerns, which also arise naturally, and which provide a source of motivation independent of self-directed concerns.

²⁰See, especially, the Introduction to the *Inquiry*.

²¹Compare Hume: “*Desire* arises from good consider’d simply, and *aversion* is deriv’d from evil. 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.439) For Hume’s value hedonism, see T.276,399,438,439.

²²These latter passages are from *A System of Moral Philosophy*, (Glasgow: R. and A. Foulis, 1755), p. 50. Because of his arguments about “exciting reasons” in *Illustrations*, it might be thought that Hutcheson believes that such reflection can cause motivation only if the agent already has desires for his own good, or the good of all. But he actually denies this: “We need not imagine any innate ideas of good in general, of infinite good, or of the greatest aggregate: much less need we suppose any actual inclination toward any of these, as the cause or spring of all particular desires.” (P.ii) We form desires for good (either for oneself or others) generally conceived by abstraction from particular desires which themselves arise from awareness of the prospects of pleasure for ourselves or others.

²³In great part, by giving us moral sense through which we can enjoy contemplating our own benevolence.

²⁴One might reasonably ask, what is the genus of which natural and moral obligation are species? Obligation, perhaps, but what is that?

²⁵“Whenever it appears to us that a Faculty of doing, demanding, or possessing any thing, universally allow’d in certain Circumstances, would in the Whole tend to the general Good, we say, that one in such Circumstances has a Right to do, possess, or demand that Thing.” (In.vii.vi) For Hume, of course, the conventions of justice must actually be established, and their consequences must be good for everyone, not just overall. Nonetheless, Hutcheson and Hume are in agreement that justice and rights can be recommended only on grounds of the consequences that generally attend their establishment.

²⁶This passage actually asserts this to be true of violating “perfect rights.”

²⁷This passage occurs at the end of a remarkable paragraph in which Hutcheson critically reviews in similar fashion the cardinal Aristotelian virtues. This may begin a tradition of utilitarian criticism of intuitive or common sense morality that reaches maturity in Sidgwick’s *Methods of Ethics*. For the distinction between the “summary” versus the “practice” conception of rules, see John Rawls, “Two

Concepts of Rules," *The Philosophical Review* 64 (1955): 3–32. Note, by the way, that what I am calling Hutcheson's "rules of right" are only appropriately viewed as summary rules insofar as they enter into determining right action. The way rules actually enter into Hutcheson's definitions of rights, is more like a practice rule. See note 26.

Even had Hutcheson come to believe that benevolence can conflict with a desire to abide by (mutually beneficial) rules that structure just practices, he would, quite unlike Hume, have had powerful systematic reasons for holding that benevolence should then prevail. For one thing, benevolence is morally good, Hutcheson believed, not because of its effects, but because moral sense approves of it (and only it) intrinsically. This is probably a secular version of the Christian idea that love begets love. On the Christian roots of Hutcheson's thought, see Wolfgang Leidhold, *Ethik und Politik bei Francis Hutcheson* (Freiburg: Alber, 1985). Moral sense also approves (derivatively) of beneficial acts and establishing beneficial conventions, but that is because these are what benevolence, the immediate object of its approval, motivates. I discuss this aspect of Hutcheson's thought, in comparison with Hume's theory of the moral sentiment in "Hume and the Invention of Utilitarianism," in M. A. Stewart and J. Wright, eds., *Hume and His Connexions*, (Edinburgh: Edinburgh University Press, forthcoming). The moral goodness of benevolence is thus fundamental for Hutcheson; good consequences have moral relevance only because they are that at which benevolence aims. And even if, as Hume argues, agents' regulating themselves by (mutually beneficial) rules of justice as a general practice has better effects, that has no tendency to establish the moral goodness of motivational states of the agents who are so regulated. For another, since Hutcheson believes that any rational desire or motive must aim at some good for someone, if just persons are concerned to comply with mutually advantageous rules, then, since the object of this concern is no good as such, Hutcheson must regard it as more like an appetite or a passion than a rational motive. It can enter into rational deliberation only insofar as its object is related to the good.

²⁸On what he famously called the "internal aspect" of rules, see H. L. A. Hart, *The Concept of Law* (Oxford: Oxford University Press, 1961), pp. 55–57.

²⁹What Hume actually says here is that no one should ever violate "those principles, which are essential to a man of probity and honour." But these are violated precisely by (knowingly) violating the rules of justice. (T.501)

³⁰Sometimes, as in the current passage, Hume seems to be assuming that agents can "impose" or "fix" rules by which they then regulate themselves, something, apparently, like the Kantian idea that agents can freely adopt maxims. Nothing I will say about the way rule-regulation figures in Hume's account of justice depends on this assumption. My interest will be in the state of rule-regulation itself, and its difference from motivation by desires for goods (including those that will be realized by having the disposition of rule-regulation).

³¹Hume sometimes uses 'motive' to refer to a reason for acting (as in: "What reason or motive have I to restore the money?" (T.479)) and sometimes to refer to a motivational state of the agent (as in: "all virtuous actions derive their merit only from virtuous motives, and are consider'd merely as signs of those motives." (T.478)).

³²There are three places where Hume discusses will and the causes of human action: (a) "Of the influence of belief," in Book I (T.118–120); (b) "Of the influencing motives of the will," in Book II (T.413–418); and (c) "Of the direct passions," in Book III (T.438–439). I am much indebted to David Aman and Rachel Cohon for discussion of points in these sections.

³³Note that even if approbation of justice were to involve something like a desire to perform just acts (and an aversion to injustice), these direct passions would also be nonstandard by Hume's theory of action and will. Note also that nothing here affects Hume's thesis that every actions proceeds from a passion. What is at issue is whether all action-motivating passions are desires for some good.

³⁴See note 31.

³⁵Hume generally uses 'convention' to refer, not to the rules by which persons agree to regulate themselves, but to the state of agreement itself. The rules, he says, "deriv[e]" from the convention. (T.490)

³⁶I take this to refute David Gauthier's speculation that "moral obligation, for Hume, arises from a coincidence between an object of our moral sentiments and an object of our reflective interests." ("David Hume, Contractarian," *The Philosophical Review* 88 (1979): 28) Annette Baier makes a similar suggestion: "Hume seems to require that, for something to be a moral obligation, it must first satisfy the test of self-interest which convention imposes." (*A Progress of Sentiments* (Cambridge:

Harvard University Press, 1991), p. 243) See also Annette Baier, "Hume and Social Artifice," *Ethics* 98 (1988): 757–778.

³⁷See note 28 above.

³⁸On the idea that accepting a norm is a distinctive motivational state that normative judgments express, see Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge: Harvard University Press, 1990).

³⁹"[A convention] is only a general sense of common interest; which sense all the members of the society express to one another, and which induces them to regulate their conduct by certain rules." (T.490)

⁴⁰Hume was not the first to argue against Hutcheson that justice is a virtue independently of benevolence, and potentially conflicting with it. Butler had made these points in *Dissertation of the Nature of Virtue*.

⁴¹Although Hume also holds rule-regulation to be supported by both self-interest and moral approbation, and thinks these depend on that motivational state's furthering private and public interest, respectively.

⁴²See Francis Snare, *Morals, Motivation, and Convention: Hume's Influential Doctrines* (Cambridge: Cambridge University Press, 1991), pp. 297–302; Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); and Philip Pettit, "Free Riding and Foul Dealing," *The Journal of Philosophy* 83 (1986): 361–379.

⁴³I am indebted to Rosalind Hursthouse for stressing this to me. On this point, and the importance of T.535, see David Gauthier, "Artificial Virtues and the Sensible Knave," *Hume Studies* 18 (1992): 407, 410; also Barry Stroud, *Hume* (London: Routledge & Kegan Paul, 1977), pp. 204–218. It is not clear, by the way, that Hume must be understood (as Gauthier does) as implying that injustice is invariably disadvantageous when, at T.499, he writes, "nor do men so readily perceive, that disorder and confusion follow upon every breach of these rules, as in a more narrow and contracted society." He might be implying that, or he might be allowing the possibility that these consequences are sometimes not there to be perceived. The former is surely more likely, however, given the passages just discussed in the text.

⁴⁴On this point, see Gauthier, "Artificial Virtues and the Sensible Knave," p. 407.

⁴⁵On "naturally", see note 17.

⁴⁶J. L. Mackie makes this suggestion in *Hume's Moral Theory* (London: Routledge & Kegan Paul, 1980), p. 80, as does Francis Snare, *Morals, Motivation, and Convention*, pp. 192–201, and Rachel Cohon, "Hume's Difficulty with the Virtue of Honesty," unpublished.

⁴⁷See, e.g., T.479. On this point, see Charlotte Brown, "Is Hume an Internalist?"

⁴⁸Hume's official psychological view is that love and hatred always have some other person as object, their self-correlates being pride and humility. (T.329)

⁴⁹Moral judgment's influence on the passions is, of course, an important theme of Hume's. (T.457) But this is fully explainable by the natural, if indirect, psychological mechanisms linking moral judgment to motivation via the pleasurable and painful moral sentiments and love and hatred. Hume's famous objection against the rationalists that they are in no position to "prove *a priori*, that [their favored] relations, if they really existed and were perceiv'd, wou'd be universally forcible and obligatory" is sometimes offered as proof that he thinks that moral judgment is intrinsically action-guiding (T.466) But Hume regards this complaint as apt here precisely because the rationalists, maintaining that morality derives from relations between ideas and not from impressions, hold it to consist in "immutable measures of right and wrong [that] impose an obligation, not only on human creatures, but also on the Deity himself." (T.456) They, therefore, are committed to explaining a relation between morality and motivation that is universal and *a priori*. A contingent natural connection will not do for their purposes, though it will for Hume's.

⁵⁰No doubt, this should count as some evidence that Hume is (at least) ambivalent as between carrying through his ethics of virtue in the case of justice and treating justice as a property of undervative property of acts, if not, indeed, that he holds the latter view.

⁵¹Gauthier, "Artificial Virtues and the Sensible Knave," p. 413.

⁵²In the course of arguing that there is no agent-state of justice, and thus that Hume holds that the moral obligation to justice motivates through self-hatred for lacking what we mistakenly believe to be such a natural state, Knud Haakonssen writes that "it is hardly likely that Hume thought self-interest, as a general character trait, morally approved by men..." (Haakonssen, *The Science of a Legislator*

(Cambridge: Cambridge University Press, 1981), p. 35. But there is no doubt Hume thinks strength of mind is a virtue.

⁵³Again, the Nidditch edition adds that this is “naturally” the only such motive. See note 17.

⁵⁴David Hume, *Essays, Moral, Political, and Literary*, ed. Eugene F. Miller (Indianapolis: LibertyClassics, 1987), pp. 479–480.

⁵⁵It should be noted here that Hume’s contrast between natural and artificial virtues in “Of the Original Contract,” brings in natural as well moral obligation. Unlike artificial virtues, natural virtues operate “independent of all ideas of obligation, and of all views, either to public or private utility.” *Essays*, p. 479.

⁵⁶Again, this is different from predicting that if others abstain from one’s possessions, then it will be in one’s interest to abstain.

⁵⁷Cf. “Of the Original Contract,” in *Essays, Moral, Political, and Literary*, p. 480.

⁵⁸“Of the Origin of Government,” in *Essays, Moral, Political, and Literary*, p. 38. This first appeared in the edition of 1777. It did not appear in any edition of the *Essays* published during Hume’s lifetime.

⁵⁹Note, however: “Had every man sufficient *sagacity* to perceive, at all times, the strong interest which binds him to the observance of justice and equity, and *strength of mind* sufficient to persevere in a steady adherence to a general and a distant interest, in opposition to the allurements of present pleasure and advantage; there had never, in that case, been any such thing as government or political society, but each man, following his natural liberty, had lived in entire peace and harmony with others.” (E.205)

⁶⁰The sensible knave’s point that noncompliance is occasionally advantageous for the agent can be extended to the case of the public interest in ways that Hume’s own examples suggest. (T.497,E.306) Again, for a contemporary (to us) example of a view that norm-acceptance is a distinctive psychological state, see Allan Gibbard, *Wise Choices, Apt Feelings*.

⁶¹See note 46.

⁶²Something like this is suggested by Marcia Baron’s, “Hume’s Noble Lie: An Account of His Artificial Virtues,” *Canadian Journal of Philosophy* 12 (1982): 539–555.

⁶³I am indebted to David Aman for this suggestion.

⁶⁴Knud Haakonssen, *The Science of a Legislator*, pp. 30–35. Haakonssen emphasizes that this is his construction and nothing to which Hume is explicitly committed.

⁶⁵*Essays, Moral, Political, and Literary*, p. 480.

⁶⁶“Two Concepts of Rules,” p. 3.

⁶⁷There are various interesting movements in the direction of these doctrines in the thought of Cudworth, Shaftesbury, Butler, and, to some extent Locke, which I discuss in *The British Moralists and the Internal ‘Ought’*.

References

- Axelrod, Robert. [1984] *The Evolution of Cooperation* (New York: Basic Books, 1984)
- Baier, Annette. [1988] “Hume and Social Artifice,” *Ethics* 98 (1988): 757–778.
- Baier, Annette. [1991] *A Progress of Sentiments* (Cambridge: Harvard University Press, 1991).
- Baron, Marcia. [1982] “Hume’s Noble Lie: An Account of His Artificial Virtues,” *Canadian Journal of Philosophy* 12 (1982): 539–555.
- Brown, Charlotte. [1988] “Is Hume An Internalist?,” *Journal of the History of Philosophy*, 26 (1988): 69–87.
- Cohon, Rachel. “Hume’s Difficulty with the Virtue of Honesty,” unpublished manuscript.
- Cumberland, Richard. [1727] *A Treatise of the Laws of Nature*, trans. J. Maxwell, (London, 1727).
- Crusius, Christian August. [1744] *Anweisung, vernünftig zu leben (Guide to Rational Living)*

- Darwall, Stephen. [1989] "Obligation and Motive in the British Moralists," *Social Philosophy & Policy* 7 (1989): 133–150.
- Darwall, Stephen. "Hume and the Invention of Utilitarianism," in M. A. Stewart and J. Wright, eds., *Hume and His Connexions*, (Edinburgh: Edinburgh University Press, forthcoming)
- Darwall, Stephen. *The British Moralist and the Internal 'Ought': 1640–1740*, (Cambridge University Press, forthcoming)
- Gauthier, David. [1979] "David Hume, Contractarian," *The Philosophical Review* 88 (1979):
- Gauthier, David. [1992] "Artificial Virtues and the Sensible Knave," *Hume Studies* 18 (1992): 401–427.
- Gibbard, Allan. [1990] *Wise Choices, Apt Feelings* (Cambridge: Harvard University Press, 1990).
- Haakonssen, Knud. [1981] *The Science of a Legislator* (Cambridge: Cambridge University Press, 1981)
- Haakonssen, Knud. [1992] "Divine/Natural Law Theories in Ethics," in *The Cambridge History of Seventeenth-Century Philosophy*, ed. Michael Ayers and Daniel Garber (Cambridge: Cambridge University Press, 1992)
- Hart, H. L. A. [1961] *The Concept of Law* (Oxford: Oxford University Press, 1961)
- Hume, David. [1739, 1740] *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd ed., with text revised by P. H. Nidditch (Oxford: Clarendon Press, 1978).
- Hume, David. [1751] *An Enquiry Concerning the Principles of Morals*, in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 3rd ed., with text revised by P. H. Nidditch (Oxford: Clarendon Press, 1975).
- Hume, David. [1777] *Essays, Moral, Political, and Literary*, ed. Eugene F. Miller (Indianapolis: LibertyClassics, 1987)
- Hutcheson, Francis. [1738] *An Inquiry Into the Original of Our Ideas of Beauty and Virtue*, 4th ed. (London).
- Hutcheson, Francis. [1742] *An Essay on the Nature and Conduct of the Passions and Affections with Illustrations on the Moral Sense*, 3rd ed. (London, 1742).
- Hutcheson, Francis. [1755] *A System of Moral Philosophy*, (Glasgow: R. and A. Foulis, 1755)
- Korsgaard, Christine. [1989] "Kant's Analysis of Obligation: The Argument of Foundations I," *The Monist* (1989): 312–313.
- Leidhold, Wolfgang [1985] *Ethik und Politik bei Francis Hutcheson* (Freiburg: Alber, 1985).
- Locke, John. [1954] *Essays on the Law of Nature*, ed. W. von Leyden (Oxford: Oxford University Press, 1954).
- Mackie, John. [1980] *Hume's Moral Theory* (London: Routledge & Kegan Paul, 1980).
- Petitt, Philip. [1986] "Free Riding and Foul Dealing," *The Journal of Philosophy* 83 (1986): 361–379.
- Pufendorf, Samuel. [1672] *de Jure Naturae (On the Law of Nature and of Nations*, trans. Oldfather and Oldfather (Oxford: Oxford University Press, 1934).
- Rawls, John. [1955] "Two Concepts of Rules," *The Philosophical Review* 64 (1955): 3–32.

- Schneewind, J. B. [1987] "Pufendorf's Place in the History of Ethics," *Synthese* 72 (1987): 123–155.
- Schneewind, J. B. [1990] *Moral Philosophy from Montaigne to Kant* (Cambridge: Cambridge University Press, 1990), v. ii, pp. 577–581.
- Schneewind, J. B. [1992] "Modern Moral Philosophy: from Beginning to End?" in *Philosophical Imagination and Cultural Memory*, ed. Patricia Cross (Durham, N. C.: Duke University Press, 1992) .
- Schneewind, J. B. "Kant and Natural Law Ethics," *Ethics*, forthcoming.
- Shaftesbury, 3rd Earl of (Anthony Ashley Cooper). [1711] *Characteristics of Men, Manners, Opinions, and Times* (London, 1711)
- Snare, Francis. [1991] *Morals, Motivation, and Convention: Hume's Influential Doctrines* (Cambridge: Cambridge University Press, 1991).
- Stroud, Barry [1977] *Hume* (London: Routledge & Kegan Paul, 1977).
- Tuck, Richard. [1987] "The 'modern' theory of natural law," in Anthony Pagden, *The Languages of Political Theory in Early Modern Europe* (Cambridge: Cambridge University Press, 1987), pp. 99–119.