

# Sophie Botros

---

## Hume, Justice and Sympathy : A Reversal of the Natural Order?

---

Diametros nr 44, 110-139

---

2015

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

## HUME, JUSTICE AND SYMPATHY: A REVERSAL OF THE NATURAL ORDER?

– Sophie Botros –

**Abstract.** Hume’s view that the object of moral feeling is a natural passion, motivating action, causes problems for justice. There is apparently no appropriate natural motive, whilst, if there were, its “partiality” would unfit it to ground the requisite impartial approval. We offer a critique of such solutions as that the missing non-moral motive is enlightened self-interest (Baier), or that it is feigned (Haakonssen), or that it consists in a just disposition (Gauthier). We reject Cohon’s postulation of a moral motive for just acts, and also Harris’s attempt to dispense with motive as the source of their merit, by invoking extensive sympathy, and citing their beneficial societal consequences. These solutions assume that, if Hume remains a virtue ethicist, the natural virtues supply the paradigm. Taylor claims that a revolution in motivational psychology follows the inauguration of the artificial convention of justice, remoulding the natural virtues. This solution founders, we argue, upon unresolved contradictions besetting even these virtues.

**Keywords:** justice, sympathy, virtue, sentimentalism, passion, motive, character, self-interest, impartiality, consequences.

Hume is famous as a moral sentimentalist who uncompromisingly excludes reason as “utterly impotent”<sup>1</sup> to influence action from the process whereby we reach moral judgments. He writes:

To have the sense of virtue, is nothing but to *feel* a satisfaction of a particular kind from the contemplation of a character. The very feeling constitutes our praise or admiration. We go no further; nor do we inquire into the cause of the satisfaction. We do not infer a character to be virtuous because it pleases: But in feeling that it pleases ... we in effect feel that it is virtuous.<sup>2</sup>

He is also, equally famously, a virtue ethicist. This position, which is already apparent in the quotation, can be captured briefly thus, when it is sometimes called<sup>3</sup>

---

<sup>1</sup> Hume [2000] 3.1.1.6; for a detailed discussion see Botros [2006].

<sup>2</sup> Hume [2000] 3.1.2.3.

<sup>3</sup> Garrett [2007] p. 258.

his “Core Virtue Ethics Thesis”: the object of moral approbation, or disapprobation, is always some trait of character which provides the motive of moral action. As Hume comments: “The external performance has in itself no merit. We must look within to find the moral quality.”<sup>4</sup> He did not, however adopt this view, it is worth noting, as some medical ethicists do today, when they recommend morally controversial acts, such as mercy killing, and argue that even transgressions of such commandments as “Thou shall not kill,” can still be morally acceptable, even praiseworthy, when performed out of the desire to alleviate suffering by a beneficent doctor. On the contrary, Hume was morally conservative, largely rubber-stamping the contemporary moral mores. Thus he held that the “passions and affections” which provided the motives of those actions which we morally approve were typically those “partial” and “unequal” ones<sup>5</sup> that are commonly (and unexceptionably) found in human nature. This partiality and inequality was not merely reflected in our moral sentiments, but given “additional force” by them. We approve a father’s love for his children or the feelings of generosity between two friends. But where this “limited benevolence,” or “confined generosity,” are absent, or where they “depart too much from the common measures” – say a father loves his children excessively, or at the opposite extreme, “in any opposition of interest, gives the preference to a stranger, or mere chance acquaintance” – so that natural partiality is “transgressed,” we disapprove, and speak of “a defect” of nature.

The various elements of Hume’s moral position have typically been thought to cause most trouble for him when he passes from a consideration of natural virtues, such as benevolence and generosity, to that of justice. Some writers maintain that if his treatment of this virtue, with its appeal to the establishment of a system of just rules, is not to be riven with tension and inconsistency, he must abandon his virtue ethics altogether in its regard. Others claim that he has effectively done so. I shall devote most of this paper to a discussion, and critique, of the main attempts to resolve these tensions and inconsistencies confined, as they are mostly taken to be, to his account of justice. I will then however place these proposed solutions within a broader context, suggesting that they all take as their paradigm the natural virtues, which justice must then be brought to conform with, so tending to give the impression that the former virtues are *comparatively* unproblematic. This will lead me to look at a ground-breaking account which reverses this direc-

---

<sup>4</sup> Hume [2000] 3.2.1.2.

<sup>5</sup> Ibidem, 3.2.2.8, 3.2.1.17.

tion of influence. Jacqueline Taylor<sup>6</sup> argues that the inauguration of the convention of justice sets in motion a revolution in our motivational, and moral, psychology which not only resolves the supposed contradictions in the artificial virtues, but remoulds our natural passions and affections and makes them for the first time objects of proper moral approval. In the last few pages I will argue that even her account faces difficulties not least because, ironically, of the recalcitrance of these same contradictions in Hume's treatment of the *natural virtues*.

### 1. Hume's dilemma: justice and the missing motive

Now, Hume is not shy to rehearse the difficulties that beset his discussion of justice: he almost seems to revel in them. Thus, he makes clear early on in this discussion his commitment to his Core Virtue Ethics Thesis, deriving from it on the grounds that otherwise one would "reason in a circle," the Principle (which again following Garrett I shall call the "First Virtuous Motive Principle") that "the first virtuous motive, which bestows a merit on any action, can never be a regard to the virtue of that action, but must be some other natural motive or principle."<sup>7</sup> Nor is it easy to dispute this derivation as it stands. The unacceptable circularity that would follow if the First Virtuous Motive Principle was denied can be characterized in either of two ways. It is that of an attempt to explain one notion in terms of another when actually the explanation one offers has recourse to the very notion to be explained. Thus the inquirer is first referred *from* the virtuousness of the action *to* the motive, and then offered a description of the motive as "regard to the virtue of the action" which refers her back again *from* the motive *to* the virtuousness of the action. Alternatively, it may be thought of as a transgression of that fundamental tenet of Hume's experimental method that an item that precedes and causes another item cannot be identical with this second item. The motive of a virtuous action must be identifiable in some other way than in terms of the moral approval it gives rise to, otherwise, as here, regard to the virtue of an action appears to be causing itself. Having expounded his Principle in several ways, he summarizes his findings in the form of the "undoubted maxim" that "... no action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality."<sup>8</sup> Yet, almost immediately, he appears to stand this maxim on its head where justice is concerned. Using the ex-

---

<sup>6</sup> Taylor [1998].

<sup>7</sup> Hume [2000] 3.2.1.4.

<sup>8</sup> Ibidem, 3.2.1.6. For our purposes "undoubted maxim" and "First Virtuous Motive Principle" can be treated as interchangeable.

ample of a loan, and considering and rejecting a series of candidate natural motives, he argues that there is in fact “no motive to acts of justice and honesty, distinct from regard to the[ir] honesty” (my italics – S.B.).<sup>9</sup> But applying his “undoubted maxim” in advance to this example, he had already declared that “the [requisite] motive can never be a regard to the honesty of the action.”<sup>10</sup> Still more remarkably, and perplexingly, Hume neither, on the one hand, ever explicitly withdraws the “undoubted maxim” – he states it unreservedly seven times in the course of his discussion of the artificial virtues, and offers a defence of it<sup>11</sup> – nor, on the other, proposes a substitute *virtuous* motive for just acts.

Hume does not however leave us at this stage of his exposition completely bereft of resources. He provides some important additional materials concerning the nature of justice’s *artificiality*. He explains<sup>12</sup> how a *convention* of justice was at some time established by people coming together, and realizing that, given both the scarcity of goods and the instability of their possession, it was in everyone’s self-interest to submit to the rules enacted by such a convention. He does not suggest that this was in anyone’s *immediate* self-interest. It might mean for instance returning a loan when one would rather keep the money and spend it. But one could come to see that it was in his “redirected” self-interest to do so. If he paid back the loan, his creditor would trust him in the future, and perhaps be willing to advance another when he needed it. His reputation as trustworthy may rise with other people too who might be encouraged to lend him their services. Moreover, every time people were seen to keep their word, in a small community, trust would more generally be strengthened and practices could flourish that made it possible for business to prosper, and for society as a whole to be advantaged. Nevertheless, it is by no means immediately clear how this new material can be used in combatting the difficulties that we are faced with in approaching justice, and which Hume himself has presented to us. For first it is soon apparent that there is no quick way of turning redirected self-interest into that virtue-imparting motive, referred to in the “undoubted maxim”, which by producing the appropriate actions, could inspire moral approval. Indeed, Hume only deepens our perplexity by pointing out that redirection does not “correct... the selfishness and ingratitude of men” or “transform unkindness into kindness. All it does is to enable

---

<sup>9</sup> Ibidem, 3.2.2.10.

<sup>10</sup> Ibidem, 3.2.1.9.

<sup>11</sup> Ibidem, 3.2.1.14–18.

<sup>12</sup> Ibidem, 3.2.2.

us to “better satisfy our appetites in an oblique and artificial manner”<sup>13</sup> or “make much greater advances in the acquiring [of] possessions, than by running into the solitary and forlorn condition.”<sup>14</sup> Secondly, even if we are tempted, despite Hume’s unpromising observations, to try and close the gulf between this self-interested motive and an ordinarily virtuous one, we shall be pulled up short by other remarks of Hume’s. For these imply that after all, and despite his own affirmation of the “undoubted maxim,” this avenue of inquiry would be counter-productive. The remarks concern the *exceptionlessness* of justice. Taking an example, Hume writes: “a man’s property is supposed to be fenc’d against every moral, in every possible case.”<sup>15</sup> But this poses the question, as Hume goes on to observe, of how any passion or affection found in human nature, being inherently “partial and unequal,” and giving rise to moral sentiments which reflect this partiality and inequality, could have the required universality to ground an *impartial* moral approval.

## 2. A critique of key solutions

There is a highly complex diverse literature dealing with the tangled skein of issues and evident contradictions, with which Hume’s treatment of justice challenges us. It will be helpful to group the key solutions into three general categories depending on how they answer the following question concerning Hume’s Core Thesis and First Virtuous Motive Principle: can both Thesis and Principle, or just one of them, or neither, be applied to justice, and does Hume intend them so to apply? In the first category are those commentators who answer ‘both’, and that it was Hume’s intention that they should be so applied. This category can be further sub-divided depending on whether the virtue-imparting motive required by the Principle is claimed to be that redirected self-interest which was the original motive for the establishment of just rules, or another natural, non-moral, but, as it turns out, merely feigned motive

Annette Baier<sup>16</sup> whose suggested solution falls within the first subdivision, accepts Hume’s description of the “interested affection” as that aspect of self-interest in general concerned with “love of gain,”<sup>17</sup> or with a desire to accumulate transferable goods, and as potentially constituting the sort of “avidity,” or material

---

<sup>13</sup> Ibidem, 3.2.5.9.

<sup>14</sup> Ibidem, 3.2.2.13.

<sup>15</sup> Ibidem, 3.2.2.16.

<sup>16</sup> Baier [1991, 1992].

<sup>17</sup> Hume [2000] 3.2.2.13.

greed, which could prompt violence. She offsets her admission by pointing out that Hume's conception of human nature is very different from Hobbes. She maintains that when Hume warns of the "forlorn and solitary condition" of human beings outside the convention, he is not referring to some kind of brutalized state existing *before* the convention had been instituted, but rather to that state which would ensue, if after it had been instituted, and people have developed a sense of their rights, the convention was dissolved. Now, she is right that the original convenors are, on Hume's view, already quasi-social beings, in the sense that they are typically family members who care about their children and friends, and may even, as she suggests, have an almost civilized love of beauty in their surroundings, and who are not warlike. They even, as she notes,<sup>18</sup> feel sympathy for strangers, although it is not strong enough to prevent bitter strife breaking out when they seek to accumulate for the benefit of themselves, their family and friends, goods which are scarce, and of which they can easily be dispossessed. Perhaps the contrast she wants is implicit in Hume's remark that "self-love when it acts at liberty *instead of engaging us to honest actions* is the source of all *injustice and violence*" (though these are my italics)<sup>19</sup>. However there is no definitive evidence – other passages being even more ambiguous – that Hume is talking, as Baier contends,<sup>20</sup> about the ravages which would follow on a dissolution of the convention, as opposed to an essentially peaceable pre-civilized life which had carried them into the convention, but for which their experience of the convention has now unfitted them. Moreover it must seem that so long as Baier's interpretation cannot be uncontrovertibly established, the motive of redirected self-interest, if not that of essentially vicious or predatory agents, cannot exactly be supposed to be *naturally* virtue-imparting.

Baier may reply that she does not rest her case on this interpretation, nor can she, if she is also to agree with Hume that there is no *natural* virtue imparting motive for just acts. For what she argues is that, though redirected self-interest *is* the requisite motive, it is to be counted *artificial*, not natural, or not "fully"<sup>21</sup> natural – or, if "natural," not "spontaneously" so – in that it leads to moral approval only after human invention has been exercised in the institution of a convention. She states that the transformation of "the interested affection" does not happen

---

<sup>18</sup> Baier [1991] p. 221.

<sup>19</sup> Hume [2000] 3.2.1.10.

<sup>20</sup> Baier [1991] p. 235–236.

<sup>21</sup> Baier [1992] p. 436.

overnight.<sup>22</sup> Emphasizing the “dynamics and dialectical” character of Hume’s account, as problems with the First Convention (concerning the stability of property) lead to a Second (its transfer by consent), and this in turn is corrected and supplemented by a Third one (instituting promises and contracts), she equally emphasizes the dynamical character of the “interested affection” itself as people become increasingly civilized and broad-minded. It is perhaps better, she suggests, to describe this motive, by the time just practices are flourishing, as “enlightened,” rather than (narrowly) redirected. In any event this cannot, she claims, be considered a virtue-imparting motive until “we [come as spectators to] overlook our own special interest and attempt to sympathize with all right-holders in the scheme and with the ‘public interest’.”<sup>23</sup> It may be wondered how, even as a moral *spectator*, making a moral judgement about someone else’s motive or action, a person finds himself even at this stage able, or prepared, to overlook his own special interest. That Hume stipulates quite generally, that such impartiality is a condition of proper moral regard only exposes the problematic nature of this stipulation. For we know that it is supposed to apply even where a natural passion is the relevant motive, whose partiality is, as Hume observes, reflected in our moral approval. Even, however, accepting Baier’s claim, if the motive for just action remains in any sense self-interest, as it does for Baier, we come to see that she is allowing it to become dangerously unstuck from the moral approval whose object it is supposed to be, since now the reason for that approval is public interest, or rather the publically beneficial consequences of the action it leads to. Indeed she says as much: “What is inadequate as a motive may be perfectly adequate as a reason for approbation of other less sublime motives.”<sup>24</sup> She continues:

It takes contrivance or artifice to create a public interest, a means of increase of public goods, and so an interest we can all share. Each person’s motive in observing the rules of justice can be enlightened self interest, awareness of her own share in the public interest, but when it comes to approving such motivated acts, her own or others, it is the public interest that becomes the relevant concern not just anyone’s share in it. The moral judge must sympathize with the public interest [which is benefitted by just practice].

---

<sup>22</sup> Baier [1991] p. 248.

<sup>23</sup> *Ibidem*, p. 241.

<sup>24</sup> *Ibidem*, p. 242.



Baier seems to be suggesting that the approving moral spectator, since *his* concern has become the public interest, overlooks, or looks beyond, the actual *self-interestedness* of the motive which prompts the agent to just observance, to the tendency of such “motivated acts” to be publically beneficial. She evidently wants to hold on in some sense to the idea that it is the motive that is the source of merit, but only that motive, whatever it is, which *leads, or has the tendency to lead, to actions with publically beneficial consequences*. Here “motive” is little more than a placeholder without any full-blooded virtue-imparting content of its own, and deriving whatever it has from the results of the action which, given the imposed contrivance, it produces. But this is essentially to get it the wrong way round, putting enormous pressure on the Core Virtue Ethics Thesis. After all Hume states repeatedly that the order of moral priority is from motive to action, not *vice versa*. He writes: “the immorality [of an omission to relieve a person’s suffering] arises from its being a proof, that *we want the natural sentiment of humanity,*”<sup>25</sup> and again:

... we regard [benevolent] actions as proofs of the greatest humanity. This humanity bestows a merit on the actions. A regard to this [action’s] merit is, therefore, a secondary consideration, and deriv’d from the antecedent principle of humanity, which is meritorious and laudable... it is only the motive... whose moral beauty renders the action meritorious [my italics].<sup>26</sup>

Baier would point out perhaps that Hume’s example is of a natural – or at least of a “spontaneously natural” – virtue, but justice is not *spontaneously* natural, and in that sense it is artificial, and hence a just action cannot be held, via its motive, to the standard of the former as proof of the natural sentiment of humanity. But if for her, as we noted, redirected or enlightened self-interest is artificial, does not she also chip away at the First Virtuous Motive Principle which requires not only that the relevant motive be independent of morality – a requirement that Baier satisfies – but that it be fully natural? She would reply perhaps that “natural” in this context means “non-moral”, and it is not meant in any case to be contrasted with “artificial”. It is moreover, she might add, its artificiality that enables her chosen motive to escape the partiality that, for Hume, attends those natural passions and affections which give rise to merely partial and unequal moral approval. Thus, it is only *as an artificial passion* that the interested motive seems capable of producing the strict, and exceptionless, conformity that, for Hume, charac-

---

<sup>25</sup> Hume [2000] 3.2.5.6.

<sup>26</sup> *Ibidem*, 3.2.1.6–8.

terizes just and honest practices, and which can properly be the object of a similarly strict, and exceptionless, approval. But we may just not find it believable, if this exceptionlessness is supposed to be a definitive mark of artificiality, that such an *interested* motive, however transformed, could be artificial *in that sense*. Rachel Cohon,<sup>27</sup> identifying redirected self-interest at best with prudence, or the “calm and informed pursuit of one’s long term interest,” points out that even the latter could not be said credibly to motivate just, or honest, acts since it could not guarantee the requisite strict, and exceptionless, observance. On the contrary, where a prudent person fails to perceive, whether correctly or incorrectly, a link between a just act and his interests, however long-term, he will – since this is what prudence amounts to and demands – refrain from performing the act.

I turn now to the other type of commentators within the first category: those who, like Knud Haakonssen,<sup>28</sup> are prepared, in order to save both Hume’s Core Virtue Ethics Thesis, and his First Virtuous Motive Principle, to postulate a natural, non-moral, yet virtue imparting, motive for just acts which is nevertheless merely *imagined*. Haakonssen proposes a sort of *bifurcation* of motives, telling a two-stage story. According to the first stage, when we see that just behaviour, motivated only by redirected self-interest, has beneficial consequences for society, we “through sympathy with its beneficial tendency, come to approve [the] behaviour.” In doing so, since “we see behaviour as an expression of motive,” we imagine that there is some natural, non-moral, but virtue-imparting motive behind the behaviour, though we do not know what it is<sup>29</sup>. The second stage, as Haakonssen narrates it, involves our reacting to this false supposition by feeling self-hatred since clearly “[we] do not have this [virtue-imparting] motive.” The desire to escape these feelings of moral inferiority gives us a new motive for acting justly, namely what he calls “*a sense of duty*,” or “*moral obligation*.” Conceived, then, under the description of “self-hatred,” our motive, being a mere psychological response, has the requisite naturalness, and independence of morality, but conceived under the description of “a sense of duty or obligation” it is apparently virtue-imparting, and so it can properly evoke moral approval. But how, it will be inquired, can a gross psychological reaction be elevated, within an account such as Hume’s, to “a sense of duty,” or of “moral obligation,” or at least do so without forfeiting the natural, non-moral, status which is also required by the Principle. Furthermore, what basis is there for any of this in Hume’s text? Haakonssen, citing Hume’s fa-

---

<sup>27</sup> Cohon [2008] p. 184–185.

<sup>28</sup> Haakonssen [1978].

<sup>29</sup> *Ibidem*, p. 13.

mous “self-hatred” example, replies that his explanation of how a moral obligation can arise to behave in accordance with the artificial virtues is exactly parallel to that which Hume gives of how such an obligation can arise to behave in accordance *with the natural virtues*. Thus, in the relevant passage Hume describes how a person who “feels no gratitude in his temper,” and so hates himself for lacking a “virtuous motive... common in human nature,” performs the “grateful” action, without the virtuous motive, from “a sense of duty or morality,” in order either “to acquire by practice that virtuous principle, or at least to disguise from himself... his want of it.”<sup>30</sup>

I have recounted the story that Haakonssen tells in a way that is ambiguous between two readings, both of which are at different times suggested by what he says. Sometimes he gives the impression that it takes a *crisis* in the practice of justice, involving a breakdown in individual cases, for people to engage in the sort of reflection that he describes, with the psychological and moral repercussions envisaged. At other times, it seems that no such crisis is required, these reflections simply occur in the ordinary course of things. Told in the first way, the story goes like this: there is a point in the development of justice when society has expanded to such an extent that we can no longer easily perceive the connection between acting justly and our own individual self-interest. As a result, the self-interested motive falters, or grows “fainter.”<sup>31</sup> We are then tempted, and perhaps give way to the temptation, to act unjustly. Looking around, however, we see that there are people who still observe the rules of justice, and we think that they must have a motive to act so, though we have none. Since, however, we cannot but register the socially beneficial effects of this behaviour, we find ourselves approving it, and think that the motive, if it is to justify this approval, must be a virtue-imparting one. This leads us to hate ourselves when we compare ourselves with these other people, and the desire to escape such unpleasant feelings (as in the analogous case of the natural virtues) provides us with a new motive for performing just acts, namely the sense of duty. This reading is particularly suggested by the fact that Haakonssen presents his story as supplementing the one that Hume tells,<sup>32</sup> which may seem to feature just such a moral ‘watershed’ in the development of justice, when people lose sight of the link between acting justly and their self-interest. However, though Hume announces at the start of his discussion here his intention to explain how we come to morally approve just acts, he never, Haakonssen points

---

<sup>30</sup> Hume [2000] 3.2.1.8.

<sup>31</sup> Haakonssen [1978] p. 10.

<sup>32</sup> Hume [2000] 3.2.2.24.

out, proposes an appropriate virtue-imparting motive which could be the object of this approval. Thus, a supplementary story, Haakonssen claims, is needed “*at a vital turn* in Hume’s argument, namely the development of moral obligation out of interested motivation” (my italics – S.B.).<sup>33</sup>

No individual or societal crisis or breakdown in the practice of justice figures in the second version of the story. It is merely that finding ourselves approving this practice, which we see to be as a whole in the public interest, we are puzzled: how can a practice that is, as we know from our own case, motivated by self-interest be praiseworthy? We can only justify our moral regard by imagining that other people have a different motive than ours for acting justly: a properly virtue-imparting one. This leads us to hate ourselves for our moral inferiority to them, and gives us, as in the first version, a new motive to perform just acts, namely a sense of duty which becomes “the real moral motive” of justice.<sup>34</sup> Haakonssen cites as textual evidence for his claim that we merely *imagine* the requisite motive, a passage on promises where Hume states of “the difficulties in supposing a moral obligation to attend promises” that we “either surmount or elude [them] by “feign[ing] a new act of the mind.”<sup>35</sup> But there is no hint in this passage that this “feign[ing] of a new act of mind” comes about as a result of a crisis in the practice of keeping promises as individual self-interest falters as a motive. Rather, people, approving of keeping promises because of its generally beneficial consequences, look for a suitable motive on which this approval can be primarily focussed, and reject self-interest, though they know that this is the motive in their own case. Haakonssen writes: “it is hardly likely that Hume thought self-interest as a general character trait, morally approved by men.”<sup>36</sup> He continues:

Certain actions done out of... self-interest have on the whole so good consequences, and seem so clearly aimed at these consequences, that men naturally come to imagine that there is a specific motive for the actions which directs them toward these consequences.<sup>37</sup>

These actions do not cease to be performed out of self-interest but come to have what is effectively an additional motive in a sense of duty. It is also suggested by the question he poses and aims to answer: how “coming to hate themselves

---

<sup>33</sup> Haakonssen [1978] p. 14.

<sup>34</sup> *Ibidem*, p. 11.

<sup>35</sup> Hume [2000] 3.2.5.12.

<sup>36</sup> Haakonssen [1978] p. 12.

<sup>37</sup> *Ibidem*, p. 13.

enables [people] to develop a sense of duty *on top of* the interest that they have in such [just] behaviour" (my italics – S.B.).<sup>38</sup>

There is, it will be suggested, apart from the puzzle here as to how on this *second* version self-interest and the sense of duty are both supposed to contribute to motivation, an obvious difficulty with both versions of the story. Moral approval and the inference to a natural, non-moral, virtue-imparting motive are again the wrong way round if they are to satisfy the Core Virtue Ethics Thesis. For, according to both stories, moral approval comes, even if only by a split second, first, being aroused in the moral spectator by the socially beneficial tendency of just actions, who *then*, in order to justify his moral approval, casts around for "a specific motive for these actions which directs them toward these consequences,"<sup>39</sup> only to invent one, when his search turns up nothing else. Admittedly, here, unlike with Baier and the self-interested motive, there is no question but that the motive is independently virtue-imparting since it has precisely been invented to meet this specification. Moreover once this motive is imagined to be in place, it can apparently lead, by way of a transition which Hume (in the "self-hatred" passage) has already described in relation to the natural virtues, to the development of a "sense of duty or obligation." Two crucial differences remain which may lead to doubts concerning the claimed parallel. Hume's example only makes sense against a flourishing background of practice of the relevant natural virtue. But it is precisely the falling off, or threatened falling off, of the practice of justice that Haakonssen, by his appeal to the example is trying to rectify. Secondly, and relatedly, whereas Hume invokes self-hatred only in dealing with peripheral cases of the natural virtues, in the context of justice it is appealed to to provide for the motivation of just acts in general.

Now while the above criticisms (and mitigations) are, I believe, justified, it is nevertheless worth observing that the first version of the story has certain resources for dealing with the issues at hand which, even if not wholly adequate, sets it apart from the second version, and tells us something interesting and important both about justice and about Hume's notion of it. The key contrast between the two versions is, as we saw, that the first centres around some kind of breakdown in just practice, some crisis of confidence in its ability to deliver what is wanted, which results in certain individuals being tempted to act *unjustly*. Thus, crucially, this solution focuses on *infringement*, rather than observance, of just

---

<sup>38</sup> Ibidem, p. 10.

<sup>39</sup> Ibidem, p. 13.

rules. Moreover this is Hume's focus in the famous passage,<sup>40</sup> to which Haakonssen refers. For Hume, having posed the question: "why we annex the idea of virtue to justice, vice to injustice?" describes just such a crisis

... when society has become numerous... and interest... [which on the first formation of society is sufficiently strong and forcible]... is more remote... [and when] men [do not] so readily perceive that disorder and confusion follows upon every breach of these rules.<sup>41</sup>

This crisis may then be thought to be intrinsic to Hume's answer in that he suggests that it is precisely in being assailed in such circumstances by the temptation to break the rules and act unjustly that people come to experience moral sentiments with regard to justice and injustice. His explanation therefore starts, significantly, as we shall see, with our reaction to injustice. Moreover, since, as he remarks, we are likely to deceive ourselves where our own unjust acts are concerned, this explanation must also start with our displeasure at *other peoples'* unjust acts (though eventually we extend this displeasure through sympathy with other spectators' moral opinions to our own unjust actions). Nor is this reaction, he observes, limited to those occasions when we personally are the victims of these others' unjust acts: "Nay," he exclaims, "when the injustice is so distant from us, as no way to affect our interest, it still displeases us." He offers two reasons for our displeasure. First, "we consider it as prejudicial to human society." Secondly, we consider it as "pernicious to everyone that approaches the person guilty of it." Reminding us of the link between such "uneasiness" and moral disapproval, he continues: "We partake of their uneasiness by *sympathy*; and... everything which gives uneasiness in human actions, upon the general survey is call'ed Vice..."<sup>42</sup> A crucial difference may be discerned between the two reasons he offers for our disapproval. Only the second mentions an individual victim of an act of injustice, one who, because of his propinquity to the perpetrator, would be in a position to feel moral disapproval at the perpetrator's character, as is required by Hume's Core Virtue Ethics Thesis, if he is properly to disapprove of the act. This does not yet explain how we as moral spectators who may be removed from the scene either in time or place – a scene which may indeed be purely hypothetical – can also disapprove of it. But Hume has already allowed for the possibility of such disapproval by stipu-

---

<sup>40</sup> Hume [2000] 3.2.2.24.

<sup>41</sup> Ibidem.

<sup>42</sup> Ibidem.

lating that no sentiment can be deemed “moral” unless the spectator adopts a “general point of view.” For to adopt such a point of view is here precisely to put ourselves in the victim’s position, so that the latter’s feelings of displeasure, when he contemplates the character of the perpetrator of the unjust act, will be communicated to us through sympathy.

Hume has exploited the fact (unnoticed as I think by most commentators, including Haakonssen) that injustices may be, and are often typically thought of as, committed by one individual against another, so that there is a clear perpetrator and victim. But he seems to assume that he can also explain under the same rubric, not only the disapproval which we feel when we observe the deleterious effects upon society at large of infringements of the rules of justice, but also our approval of their ongoing observance. As Hume has earlier pointed out, however, in obeying on a particular occasion the rules of justice, whatever the ultimate benefits for society as a whole, we are as likely to disadvantage, as advantage, the individuals immediately affected by our action. As far as the fruits that accrue to an individual through the system of justice as a whole are concerned, for instance through society becoming more prosperous, it will usually be impossible for this individual to identify any single person who, by their observance of its rules, is responsible for this individual’s share in the wider prosperity, or whose character the individual, thus benefitted, could morally approve. Moreover, the only pleasure that could be communicated to us as moral spectators, as a result of our reflection, aided by extensive sympathy, upon this individual’s situation, would be his satisfaction at his improved lot. But *that* is not, so long as Hume insists on the moral priority of character over the effects of action, strictly *moral* pleasure, because its object is not any virtuous character, and so it could not constitute our moral approval. Not even then by taking the first version of Haakonssen’s story, and appealing to the relevant *Treatise* passage, can we – except where one person perpetrates an injustice against another – ensure that moral sentiment is directly, and in the first place, aroused by the contemplation of character, not consequences, and so dispense with Haakonssen’s resort to a *subsequently imagined* motive. Even in the amenable case, it would seem, the trait deplored would have to be both identifiable independently of, yet linked in the right way with the infringement, to count as injustice, rather than as some natural vice which happened in the circumstances to cause the infringement. But it is dubious whether the conceptual resources for this are thus far forthcoming.

Still falling into the first category, but combining elements of both solutions so far considered is that of David Gauthier.<sup>43</sup> This commentator seeks to meet Hume's First Virtuous Motive Principle, as Baier does, by suggesting that the motive of just action – one which is “natural,” in the sense of “independent of morality,” though not in a sense which excludes its artificiality, and which is also virtue-imparting – is redirected self-interest. He therefore does not resort, as Haakonssen does, to an imagined motive to fulfill this role. He does not think however that redirected self-interest is virtue-imparting from the moment the convention is established. His argument for its being thus transformed, unlike Baier's, invokes that threatened breakdown of just practices (already familiar from our first version of Haakonssen's story) when “society becomes more numerous,” and mistaken perceptions lead to the weakening of the motive of interest. Again, like Haakonssen, he imports<sup>44</sup> Hume's “self-hatred” example from the context of the natural to the artificial virtues in order to establish a moral obligation to act justly, over and above redirected self-interest. He maintains, unlike Haakonssen, that the latter, “redirected towards the conventions on which society depends,” is sufficient on its own to give rise to this moral obligation in the self-hatred situation.

Gauthier does hold that redirected self-interest, right from the start, constitutes a *natural* obligation to justice, and in so doing already resembles moral obligation in a crucial way. This is because it is redirected, or rather because self-interest “redirects itself,”<sup>45</sup> and this redirection of the love of gain's “heedless and impetuous movement” is experienced by the agent as a form of *restraint*. Thus actions motivated by redirected self-interest – now essentially an artificial motive in that “it arises only as part of a conventional practice” – share this phenomenological feature with moral actions. There are however difficulties with the claim that Hume uses the word “obligation” to signify a feeling of restraint. He does not say that to be naturally obliged is to *take oneself* to be so obliged, nor that this is a consequence of recognizing that, as Gauthier puts it, “what is required is restraint.” For Hume, the claim that each person has a natural obligation to justice is not, as Gauthier states, “a further claim” *on top of* the claim that “each person has an interest [in acting justly].”<sup>46</sup> The natural obligation to justice is simply one's interest in being just. Thus Hume writes of “the natural obligations to justice, viz. inter-

---

<sup>43</sup> Gauthier [1992].

<sup>44</sup> Gauthier [1992] p. 412.

<sup>45</sup> *Ibidem*, p. 406.

<sup>46</sup> *Ibidem*, p. 409.



est"<sup>47</sup> and also of "the natural obligations of interest."<sup>48</sup> Though this interest exists whether or not one realizes it, and even if one mistakenly thinks one's interest lies elsewhere, it can of course only become one's motive when one does recognize it.

Furthermore, to make good his claim that redirection is *felt* as a restraint, Gauthier has to "suppose that the natural tendency of the passion to its unrestrained fulfilment is still present" even *after* redirection.<sup>49</sup> But this means that, even after I have arrived at a true understanding of where my self-interest lies, I still have either to conquer, or to succumb to its contrary "natural tendencies." But Hume sees neither a need, nor an opportunity for such an exercise of will. According to him, once individuals are sensible of their proper interest, they express this to their fellows and resolve to act accordingly. "No more," he writes, "is it requisite to induce any one of them to perform an act of justice who has the first opportunity."<sup>50</sup> If reflection is in the first place impeded as when "a great present advantage"<sup>51</sup> causes me "to over-look [a] remote [but much greater] interest," this is essentially a failure of knowledge – it has to do, he says, with "the degrees of men's sagacity or folly"<sup>52</sup> – not of will.

Finally, morality itself, at least where natural virtues are the model, is not, for Hume, constraining. Rather, it is natural and pleasurable. Even where an individual is moved by self-hatred to act as if he had a virtue which he does not possess, he is "pleas'd" to do so, says Hume,<sup>53</sup> banishing any suggestion of unwillingness. On Hume's account,<sup>54</sup> we say someone is under a moral obligation when as *spectators* we are disappointed that he has failed to perform an action which would, we believe, have aroused in us the pleasure of moral approval if he had performed it. There is no direct or necessary link between an agent's believing this ought-statement (which is just after all an expression of the onlookers' disappointment) and his performing the requisite act. This is why, when the agent has no appropriate natural motive, there has to be some other passion – self-hatred, or the desire to hide from himself his deficiency, or to court favour – if he is still to

---

<sup>47</sup> Hume [2000] 3.2.2.23.

<sup>48</sup> *Ibidem*, 3.2.8.7.

<sup>49</sup> Gauthier [1992] p. 410.

<sup>50</sup> Hume [2000] 3.2.2.22.

<sup>51</sup> *Ibidem*, 3.2.8.7.

<sup>52</sup> *Ibidem*, 3.2.2.13.

<sup>53</sup> *Ibidem*, 3.2.1.8.

<sup>54</sup> See particularly: *ibidem*, 3.2.5.4.

act. Regard alone for the morality of the action will not suffice since, according to Hume, we cannot change our sentiments by an act of will.<sup>55</sup>

Perhaps the last objection might be met by suggesting that the natural virtues ought not to be assumed to be the model, and that the artificial ones may set a precedent. This raises an interesting question to be considered later concerning the direction of influence in Hume's account of morality as a whole. Suppose we allow here for the sake of argument that a natural obligation does resemble a moral one, in involving "felt restraint."<sup>56</sup> We still lack an adequate explanation of why redirected self-interest should become, during the course of just practice, that virtue-imparting motive which could give rise to moral approval. As we have seen, it must do this if it is to figure plausibly in the self-hatred scenario, and in the origination of the moral obligation to justice. In fact, Gauthier interposes a "sense of justice" between redirected interest, as merely a natural obligation, and as that virtue-imparting motive which could be thus approved, and precede the formation of the moral obligation.<sup>57</sup> What does "the sense of justice" signify for Gauthier then? He points out that it "is not a natural trait whose absence is a defect in one's temper, rather it arises" (he now quotes Hume's *Treatise*<sup>58</sup>) "artificially, tho' necessarily from education and human convention." He nevertheless identifies "the sense of justice" with "the interested affection" which, he claims, has the requisite independence of moral approbation. In accounting for the implied transformation of the "interested affection" – or "sense of justice" – he stresses its artificiality, *as the sense of justice*. This artificiality consists not only in the "redirection [of the interested affection] toward its better satisfaction" being a matter of its restraining itself, but also of its being brought about "by the conventionally instituted laws of society" which permit a distinction between "particular acts as just or unjust." He maintains that this natural, in the sense of non-moral, yet artificial, motive finally gives rise in us to moral approval when, as societies expand, our short-sightedness causes us to doubt whether the practice of justice really does pay off for us. For only then are we caused to "reflect" on "[the practice's socially] beneficial tendencies" and "[to] observe the [beneficial] effects of the universal practice of justice." He suggests however that something else happens: "[We start to] take adherence to the practice to indicate a just disposition." Now this last remark looks like an attempt to head-off an anticipated criticism. This is that his so-

---

<sup>55</sup> Hume [2000] 3.2.5.4–5.

<sup>56</sup> Gauthier [1992] p. 410.

<sup>57</sup> *Ibidem*, p. 413–414.

<sup>58</sup> Hume [2000] 3.2.1.17.

lution, no less than the others, represents moral approval as only derivatively dependent on motive, or character, that is to say, only in so far as they produce actions with socially beneficial consequences. It thus reverses the order of moral priority as stipulated in the Core Virtue Ethics Thesis again. Gauthier's response to such an anticipated criticism seems to be to tie motive, or character trait, in the form of "just disposition," to adherence to just practice, in such a way that the claim, that it is merely derivatively dependent for its moral merit on the practice, can no longer get a purchase. But if (as is often held, and as Gauthier would need to hold for this response to work) to say that someone has a just disposition is simply to say that he adheres to just practice, the former being merely a logical construction out of the latter, what has happened, we may want to know, to Hume's original virtue-imparting motive? For, being a passion in human nature that causes action, it is independently specifiable in its own right. It may seem nevertheless that Gauthier wishes to cast his logical wraith in this substantial role. For he writes, paraphrasing Hume's remark, and replacing Hume's "quality of the mind" with his "disposition": "the idea of such a disposition pleases me *after a certain manner*; this pleasure in turn leads me to judge that your disposition is a virtue, and that you are virtuous."<sup>59</sup> But in appealing to the "self-hatred" example in order to explain the *moral* obligation to justice, he seems to revert to the properly substantial motive. For he writes that it is the feeling of being "devoid of the redirected interest that constitutes justice," that leads a person to "perform a just action without the motive from a certain sense of duty." He does not make clear exactly what relation the just disposition is supposed to bear to the redirected self-interest – the sense of justice – which is also therefore claimed to be, as in the last quotation, the virtue imparting motive of just action. Nor, if there are two items here – and they are certainly characterized very differently – does he give a systematic account of their respective contributions.

It is finally worth noting that though Gauthier, like Haakonssen on our first version of his story, appeals to a threatened breakdown in just practice, he does not, like Haakonssen and Hume, exploit this fact by concentrating in the first place on *injustice*. He puts the genesis of approval and disapproval immediately on an absolutely equal footing. This makes his subsequent appeal to the "self-hatred" example in order to furnish a supposedly missing motive, namely that of moral obligation, somewhat superfluous. (It remains true that only Hume sees the potential of those particular cases, where one individual commits an injustice against another, for the restoration of motive as the immediate object of moral sentiment.)

---

<sup>59</sup> Hume [2000] 3.2.5.4.

All the solutions, considered so far, each attempting to hold on to both the Core Virtue Ethics Thesis and the First Virtuous Motive Principle, and so belonging to our first category, have serious weaknesses. I turn now to the second category and to the suggestion that if the First Virtuous Motive Principle, or “undoubted maxim,” is deliberately relinquished with regard to just acts, Hume’s denial that these acts have a natural, in the sense of a non-moral, motive will no longer seem contradictory. Moreover, the way will be open for the postulation of a *moral* motive for just acts. The outstanding example of this approach, which professes at the same time to adhere to the Core Virtue Ethics Thesis, and the only one I will consider here, is represented by Rachel Cohon.<sup>60</sup> This author, dismissing the resort either to redirected self-interest (Baier), or to the just disposition (Gauthier), or yet again to a fictitious motive (Haakonssen), asserts, with uncompromising audacity, that “our approval of honest action, instead of depending upon the motive of honest action, ultimately provides it.”<sup>61</sup> Her approach shares with Haakonssen’s, on our first version of his story, and Gauthier’s, a reliance on that critical juncture in the development of justice when, societies having expanded, the motive of redirected self-interest wavers, and she seeks to exploit it to introduce her specifically moral motive. The story that she tells however is interestingly, and contentiously, different from those of the other commentators. In the first place, as we have noted, she deliberately abandons, as regards justice, Hume’s “undoubted maxim” (or First Virtuous Motive Principle). She argues that he could not reasonably be thought to adhere to it in its generality whilst at the same time denying the existence of a natural, non-moral motive for just acts, and alleging that there was a vicious circle in trying thus to explain them. Rather, she suggests, he perceived that justice, or honesty, requires impartiality and willingness to cooperate with strangers, which often conflict with our natural proclivities, and concluded that such virtues were “prosthetic.” They were socially fabricated in order to remedy our shortcomings. However, the success of this enterprise required concealment. Therefore, people were kept under the illusion that just or honest actions had natural non-moral motives. Indeed, they were unaware that moral approval itself, socially enhanced, provides a new moral motive for these actions.

Now on the one hand, as I indicated earlier, there is no direct textual evidence that Hume does reject the “undoubted maxim” in its generality. He states it unreservedly seven times and offers a defense of it<sup>62</sup> while never explicitly repudi-

---

<sup>60</sup> Cohon [2008].

<sup>61</sup> Ibidem, p. 181.

<sup>62</sup> Hume [2000] 3.2.1.14–18.

ating it. On the other, it is difficult not to sympathize with Cohon when she reasons that he would hardly have propounded a maxim without qualification at the same time as pointing out its failure to apply in certain cases. David Wiggins,<sup>63</sup> whom Cohon does not mention, offers a possible solution: Hume did originally propound the maxim as applying to all virtues, but then “silently revokes it,” his “infamous” circle being “an expository device.” Wiggins, who interprets the *Treatise* in light of the *Enquiry*, acknowledges that this revocation would also require the abandonment, as regards justice, of the claim which Cohon professes to hold on to, namely that there has to be a virtue-imparting motive to approve, and that “moral beauty” cannot lie “in observance as such” (Hume’s Core Virtue Ethics Thesis). We may yet wonder how Cohon herself *can* hold on to that Thesis, given her remark, quoted above, that moral approval need not depend on a *motive* at all. It turns out, however, that for Cohon,<sup>64</sup> unlike for Wiggins, evaluation can be of actions alone, or “bare dispositions to act,” only during that “vulnerable stage in human social development” which precedes justice becoming a fully fledged virtue. Once moral approval has taken hold, due to our having come, through sympathy with the public interest, to appreciate, in our newly precarious state, the beneficial social consequences of the observance of just practice, there is generated from these sentiments a moral motive which “fills the breach” left by waning interest. But this does not occur without more human intervention. As she puts it, “... parents and politicians engage in a further ‘artifice’ which... converts their charges’ approval of rule-following acts... into a motivating sentiment.” This is later strengthened by a concern for reputation, which is also fostered in these charges. Consequently, the Core Virtue Ethics Thesis, after this interval, can be taken to be properly established, or if one prefers, “regularized.”

Even leaving aside her stance toward Hume’s “undoubted maxim,” and the hiatus during which transitional period approval has no motive, there are difficulties with Cohon’s claim that, for him, the motive that supposedly transforms justice into a virtue is generated by moral approval itself. For the passage which she cites as textual support<sup>65</sup> resists such an interpretation. Cohon (like Gauthier and others) is impressed by the fact that Hume explains how moral sentiments come to attach to justice at the same time as he describes how, as societies expand, the self-interested motive falters. But I am not sure how much we should read into this order of exposition. Cohon may reply that our doubts are due to our intransigence

---

<sup>63</sup> Wiggins [2006] p. 80–82.

<sup>64</sup> Cohon [2008] p. 174–175.

<sup>65</sup> Hume [2000] 3.2.2.24.

in refusing to recognize that Hume countenanced a moral motive for justice. But Hume does not state, in the cited passage, that moral approval, or the underlying sympathy, generates a motive. He concludes only that “self-interest is the original motive to the establishment of justice, but sympathy with public interest is the source of our approval of it.”<sup>66</sup> Does not his explanation of how an individual extends her moral disapproval at other peoples’ injustices, with their socially damaging consequences, to her own infringements, “suggest” that she will be moved thereby to refrain from future infringements? The last remark in the paragraph appears to rule it out: “Sympathy is too weak to controul our passions but has sufficient force to influence our taste.”<sup>67</sup> Moreover, this reflects Hume’s moral sentimentalism with spectators, not agents, at its core, disinterestedly contemplating moral, much as they might aesthetic, beauty. Moral sentiments are a kind of love and hate, but being disinterested are less vivacious. As Cohon concedes, such “indirect” passions, unlike the “direct” passions of desire and aversion, cannot directly motivate us. Of course, Hume had argued earlier<sup>68</sup> that morality, in contrast to reason, was precisely an active force. Almost everywhere else, however, he accords the disinterested sentiments, that he goes on to identify it with, scarcely even a tangentially active role.

So we come to the third category, and to those commentators who dispense with and take Hume to dispense with both the Core Virtue Ethics Thesis and the First Virtuous Motive Principle, alleging that the consequences of both just and unjust acts are sufficient alone, without allusion to a state of character, to explain their approval and disapproval. It may seem that there is not much left to say about this appeal to consequences since we have already had occasion to discuss and criticize it in relation to the commentators considered above. They too, we saw, are obliged to make such an appeal at a certain stage in their arguments, even if only as an interim measure before hastily, and not entirely convincingly, re-embracing the Thesis. However, it is worth evaluating the third stance, when it is adopted consistently and in its own right, especially as it forces a clarification of the role that, for Hume, sympathy plays in moral judgement. As with the second category I will content myself with one particularly well worked out and cogently defended example of the proposed solution: that of James Harris.<sup>69</sup>

---

<sup>66</sup> *Ibidem*.

<sup>67</sup> Cohon [2008] p. 175-176.

<sup>68</sup> Hume [2000] 3.1.1.

<sup>69</sup> Harris [2010].

Harris thinks that anyone who takes the salient question posed by justice to be why we are motivated to act justly is already asking the wrong question. They have been misled by failing to keep separate Hume's explanation of the origins of justice, which does indeed appeal to motives, but is purely historical, from his account of how we become morally obligated to act justly. "Showing how the conventions developed is not," Harris observes, "the same thing as showing why we believe that we are morally obliged to respect the conventions."<sup>70</sup> The answer to this further question – though Hume deals with it<sup>71</sup> after having drawn attention to the weakening, in "more numerous societies," of the interested and morally neutral motive which got us into the convention – is not directly motivational. For it concerns in the first place how we, as moral *spectators*, come to *esteem* just practice. Harris suggests that what is crucial to this process, as Hume describes it, is the spectators' realization of the socially beneficial consequences of compliance with the rules of justice and, conversely, the socially deleterious ones of their infringement. This realization, together with extensive sympathy for the affected parties, explains the ensuing moral pleasure, or displeasure felt by these spectators, even when their own interests are not involved. Harris suggests that at no stage during which their instinctual sympathetic processes are thus refined and corrected, do motives matter, "sympathy [being] excited primarily [for Hume] by the consequences of action (and inactions)." Nor does this give Harris qualms. For, in rejecting what he calls a "virtue theoretical construal," he goes further than Cohon, dismissing as relevant to justice not only (as she does) Hume's "undoubted maxim" in its generality, but also his Core Virtue Ethics Thesis. He has done this despite Hume's prominently placed, repeated assertions and reformulations of the Thesis, and despite its standing, as attested by several distinguished writers, as the fundamental principle of Hume's kind of philosophy. It is hardly surprising then that he neglects to accord any special significance to the fact that Hume's explanation starts from peoples' response to *injustice*, or discern in this Hume's struggle to hold on to some vestige of his Core Thesis. Nor does he mention Hume's allusion to individuals who, being around the perpetrator of an injustice, suffer directly at his hands, and who would be in a position to feel that "uneasiness" at his *character* which, communicated through sympathy, could still evoke in them moral disapproval, in the full virtue ethical sense.

There is a suggestion in Harris's account that, because sympathy is focused upon the consequences of actions, on effects, not causes, when Hume appeals to

---

<sup>70</sup> Ibidem, p. 39–40.

<sup>71</sup> Hume [2000] 3.2.2.24.

that mechanism, in explaining how justice becomes a virtue, he is already signaling his preparedness to suspend his Core Virtue Ethics Thesis, with its concentration on motive as the object of moral approval. Obviously, if sympathy only becomes connected with consequences in the context of justice, and elsewhere focuses on motive, the appeal itself to this mechanism will not have this larger theoretical significance. It is worth therefore clarifying the sphere of operation of sympathy, as Hume understands it. Admittedly, in his famous “surgery” example<sup>72</sup> sympathy does not operate through any observation of or response to a motive or character trait expressed in action. On the contrary, the patient “seeing the preparation of instruments” is bitterly anxious at the thought of his coming suffering. This anxiety communicates itself, given that the principles of resemblance and of causation are met, through sympathy to the spectator, exciting in him “the strongest sensations of pity and terror.” Moreover this has led some writers to speak of such “sympathy” as an “emotional contagion,” or an affective “mirror[ing].”<sup>73</sup>

We must remember however that Hume also defends the view that “our *esteem* proceeds from sympathy” (my italics – S.B.)<sup>74</sup> and is “the chief source of moral distinctions.”<sup>75</sup> This is not only because, as he puts it, sympathy interests us directly “in the good of mankind,”<sup>76</sup> but because it has an important role in enabling us to adopt the general viewpoint necessary to any proper moral judgement, whether concerning the natural or artificial virtues. It does this, he explains, “by taking us so far out of ourselves as to give us the same pleasure or uneasiness *in the character of others* as if they had a tendency to our own advantage or loss” (my italics – S. B.).<sup>77</sup> There is here already a reference to character as the object of the pleasure or uneasiness that sympathy, though it may need correcting, communicates to us. But such a remark is still compatible with the view that what our sympathy is in the first place communicating to us are feelings which are merely effects of an action by an agent upon its recipients. Hume however goes much further in bringing character within the remit of sympathy. For the agent’s character can only be the object of our moral approval, for Hume, because of our sympathy with that peculiar pleasure or uneasiness felt by the recipient *herself* toward this

---

<sup>72</sup> Hume [2000] 3.3.1.7.

<sup>73</sup> For instance, see Collier [2010] p. 258, 269.

<sup>74</sup> Hume [2000] 3.3.1.14.

<sup>75</sup> *Ibidem*, 3.3.6.1.

<sup>76</sup> *Ibidem*, 3.3.1.19.

<sup>77</sup> *Ibidem*, 3.3.1.11.



character, which is *her* moral approval or disapproval. To take just one apposite quotation of many, Hume writes:

We blame equally a bad action which we read of in history, with one perform'd in our neighbourhood t'other day: The meaning of which is that we know from reflection, that the former action would excite as strong sentiments of disapprobation as the latter, were it plac'd in the same position.<sup>78</sup>

I conclude that it will be difficult to interpret Hume's appeal to sympathy in the context of justice, on grounds that this mechanism is concerned with effects not motives, as necessarily a sign of his willingness here to abandon virtue ethics.

### 3. A revolution in moral psychology?

This concludes the substantial part of my paper consisting in a conspectus and critique of key proposed solutions to the difficulties posed by Hume's treatment of justice. So far, our classification has been in terms of the stance adopted regarding his Core Virtue Ethics Thesis and his First Virtuous Motive Principle. The proposed solutions have something more general in common: they are largely pre-occupied with justice and the artificial virtues, and do not consider the implications for, or repercussions of, what they say on the natural virtues. The impression is thus inevitably created that this latter area of his thought is less plagued by tension and contradiction. In a moment I will briefly discuss a commentator, Taylor<sup>79</sup> who, as I indicated at the outset, takes no such view, setting a bold new agenda for work on Hume's moral philosophy. In approaching her position, let us first note two other salient modes of classification that we might have availed ourselves of above. One of these is in terms of method, more particularly *direction of inquiry*. Commentators belonging to our original first or second category (Baier is an example from the first category) seek a suitable motive for just acts, only aiming subsequently to explain how this can give rise to the requisite impartial and exceptionless moral approval of them. Less often, a writer may proceed in the opposite direction. Thus Taylor, while admitting that "there are complexities surrounding the motivation to be just" holds nevertheless that "in one sense Hume's explanation is quite straightforward: self-interest motivates us to establish and follow the convention of justice."<sup>80</sup> Accordingly, she starts her investigation else-

---

<sup>78</sup> Ibidem, 3.3.1.18.

<sup>79</sup> Taylor [1998].

<sup>80</sup> Ibidem, p. 5.

where, by focussing on that impartial and exceptionless moral approval that characterizes our response to an established just practice. She notes Hume's rejection, as an "improper foundation" for the "rigid, inflexible rules of justice," of natural motives since they "must accommodate themselves to circumstances and admit of all the variations which human affairs... are susceptible of."<sup>81</sup> She therefore asks what kind of revolution in our moral psychology must have taken place, along with the establishment of the artificial convention, to foster in us the requisite civilized sense of morality, together with appropriate motives. Now the *direction of inquiry* – that is, whether it is from motive to moral sentiment or *vice versa* – often tracks the *direction of influence*, namely whether this is from the natural to the artificial virtues or *vice versa*. Proposed solutions can now be seen to differ in terms of whether they take the natural virtues, in supplying a motive for action, independently of the moral sentiments they evoke in a spectator, to be the paradigm with which justice and the artificial virtues must conform, or reverse this order. Haakonssen and also Gauthier, following the first approach, and taking their model from the natural virtues, concentrate on the (apparently missing) motive for just acts, and attempt by an exercise of ingenuity to find one. Taylor, boldly turning on its head much orthodox commentary, argues that, for Hume, moral approval, even of the *natural virtues*, requiring that "a character trait be considered in general" or "impartially,"<sup>82</sup> can only be properly experienced *after* there has occurred a transformation of our natures following the establishment and observance of the artificial virtues, in particular the rules of justice.

Pivotal to Taylor's account are two intimately related distinctions: those between our "pre-just" and "post-convention" motivational psychology, and between our "uncultivated" and "cultivated" sense of morality.<sup>83</sup> Our passions, before the establishment of the convention, though the product of social interaction, are, as Hume puts it, "irregular" and "incommodious,"<sup>84</sup> having "to accommodate [themselves] to circumstances."<sup>85</sup> Uncultivated morality does little more than "rubber stamp" this partiality of natural motives. But Hume stipulates that moral sentiment involves the adoption by the spectator of a general and impartial point of view in surveying character or action. There is, it seems, nothing in uncultivated morality which could explain such a viewpoint. There is no natural love of

---

<sup>81</sup> Hume [2000] 3.2.6.10.

<sup>82</sup> Hume [2000] 3.1.2.4.

<sup>83</sup> Taylor [1998] p. 7.

<sup>84</sup> Hume [2000] 3.2.2.9.

<sup>85</sup> *Ibidem*, 3.2.6.9.

mankind at large, nor any natural or common conception of the public good which could take people beyond their merely “confined generosity” or “limited benevolence.” Nor can sympathy be relied upon since it grows fainter when the pleasures and pains of strangers are involved, especially if they are distant from us in time and place. Taylor therefore takes the adoption of a general viewpoint to be a distinguishing mark of cultivated morality. It requires that the force, and direction, of such natural passions be changed, and what is “irregular and incommodious” in them be corrected. The following question arises: how does entering into the convention effect such a change? Reflection is obviously involved in bringing us to see, once the idea of a convention is even muted, that the interested passion is far more likely to achieve its end when it has been redirected. But that alone will not get us very far. We will still prefer our own interests, or those of our friends, to that of strangers, and we will still have regard to them in evaluating character and action. Taylor now distinguishes between “the general sense that interest is common,” which still involves “[this] narrow prudential perspective,” and “a shared sense of the public interest,” which is effectively “a moral point of view” – “a general moral perspective founded on extensive sympathy.” She argues that “in comprehending the interest of each individual, the convention has an unintended result: our observance of the specific rules of justice leads us to form a shared conception of the public interest.”<sup>86</sup> In brief we are liberated from a narrow concern with “consequences, the relation of the agent to us, or our own private view of her character” to concentrate, in light of that public interest in which we have a stake, and with which we are now partially identified, on the relation of actions to collectively determined rules which serve that broader interest. Nor does it stop here. Our new shared evaluative perspective shines a light back upon even natural motives and traits of character. Thus, according to Taylor, convention and character have a profound reciprocal influence, reaching as far as the natural virtues, which can no longer be so sharply contrasted with artificial ones, as mere “givens” of human nature. Her examples show how radical her interpretation is. “Our acceptance of the operation and purposes of artificial contrivances,” she writes,<sup>87</sup> affects our understanding even of such fundamental notions as those of harm and injury, kindness and what it is to be humane. Our private conception of the virtue, for instance, may lead us to approve of someone for being ‘kind’ who, seeing the penury of another refuses to seize from him certain goods in order to return them to their rightful but rich owner. This, however, from our common evaluative per-

---

<sup>86</sup> Taylor [1998] p. 19.

<sup>87</sup> *Ibidem*, p. 22.

spective reveals itself as a misunderstanding: If a rule of justice applies, which is thereby infringed, what we take as a matter of kindness cannot really be so.

#### **4. Contradictions in the natural virtues**

As I intimated earlier, I am pessimistic that any solution along the lines explored by Baier, Haakonssen, Gauthier, Cohon or Harris, will be found for the contradictions inherent in Hume's treatment of justice. Taylor, as we have seen, offers a novel approach to them. She seeks to integrate the artificial virtues within a much broader picture, alongside the natural virtues, arguing that the establishment of the convention sets in motion a revolution in our motivational and moral psychology, so profound that it not only remoulds the natural virtues, but changes the conceptual terrain in such a way that many of the old questions and difficulties disappear. I shall concern myself here, in contrast with my procedure earlier, with Taylor's stance only in so far as it bears on the natural virtues. However, it seems to me that, given the high degree of integration, and inter-dependence of the different parts of her picture, it will be enough to raise doubts in one area. At the same time, what I have to say will serve to reveal, with reference to just one contradiction (there are others<sup>88</sup>) that justice, in Hume's hands, does not have a monopoly where contradictions are concerned, but that they are already present, if slightly submerged, in his treatment of the natural virtues.

Consider Taylor's attribution to Hume of the distinction between "uncultivated," and "cultivated," morality. Certainly, Hume occasionally alludes, although vaguely, to an apparently earlier period of uncultivated morality. On the face of it, moreover, one might plausibly suppose that he has the first in mind when he describes moral sentiment as reflecting, indeed accentuating, the natural "partiality" and "inequality" of the ordinary run of "passions and affections" which are its object, and the second, when he stipulates that no sentiment can be "denominated moral" unless the spectator adopts a "general" viewpoint, namely an impartial one, without regard to "[his] particular interest."<sup>89</sup> Such a distinction

---

<sup>88</sup> The tension for instance between whether character (which is immediately emotionally responded to) or the effect of an action (which cannot be determined without "reflexion") is morally prior, is also present, if slightly submerged, in his treatment of the natural virtues. The passage on "virtue in rags" reveals him to have been acutely aware of it there. Sometimes, it looks as if he was trying to resolve it by taking the question "From what principles this pain or pleasure that distinguishes moral good and evil is derived?", and its answer: "from a character's "tendency to the good of mankind" as purely of concern to philosophers, to be settled after they have accurately described the phenomenon 'on the ground'. But later he so construes the "general viewpoint", which any ordinary moral spectator must adopt if his sentiment is to be properly moral as involving reflexion about consequences.

<sup>89</sup> Hume [2000] 3.1.2.4.

would also avert the contradiction which would otherwise occur on grounds that morality cannot be both partial and impartial. I will show that no such contradiction can in this way be averted. It is not just that the matter is complicated by the fact that Hume means different things both by “partiality” and by “the general viewpoint” at different times. It is also that in certain examples we find partiality as an apparently integrated element of the moral response of someone who is held to have adopted the requisite general, and *so impartial* viewpoint. It becomes impossible to extricate an uncultivated from a cultivated morality, supposing them to be associated in a one-one relation with partiality and impartiality.

The complication in his notion of partiality is this: sometimes Hume means, in describing a moral response as “partial,” that the spectator only approves or disapproves of a character in terms of the tendency of that character to benefit the spectator himself, or his family or friends. Thus he writes of “our first and most natural sentiment of morals [being] founded on the nature of our passions, and giv[ing] preference to ourselves and our friends above strangers.”<sup>90</sup> This sense in which moral sentiments are partial – let us call it the “narrow” sense – obviously clashes with the requirement that we adopt a general viewpoint, in the exposition of it just given. Clearly, if we were able to take partiality in this “narrow” sense as indicating an uncultivated morality and the general viewpoint, where that, by definition, takes no heed of the spectator’s own interests, as indicating a cultivated morality, we might seem to be able to avert contradiction. But in other remarks he seems to mean by “partiality” that we, as spectators, irrespective of our own personal interests, will morally approve or disapprove of a character trait in terms of whether *the possessor* of the trait is moved by those “passions and affections” which are “common in human nature.” Hume’s claim that our moral sentiments are partial in this “broader” sense, sits uneasily with several remarks that he makes when he returns, after he has tackled the artificial virtues, to examine in detail the natural ones,<sup>91</sup> and which also stress the importance of adopting a general point of view. For here he suggests that moral sentiments, even with regard to (most) natural virtues, are properly a response, now aided by “extensive” sympathy, to a given character trait in so far as it has a “tendency to benefit society” or “mankind” or to increase “the good of mankind,” or “the public advantage.” As he expresses it,

---

<sup>90</sup> Ibidem, 3.2.2.11 & 3.3.1.18.

<sup>91</sup> Ibidem, 3.3.1.

moral distinctions arise in great measure from the tendency of qualities and characters to the interest of society, and that 'tis our concern for that interest, which makes us approve or disapprove of them.<sup>92</sup>

It will be countered that the larger interests of society could be inversely proportional to those of the family and friends of the possessor of the trait. Thus “adopting the general viewpoint” may come into conflict with the claim that moral sentiments are partial, as much when partiality is understood in the “broader” sense, as when understood “narrowly,” so that they reflect exclusively the interests of the moral *spectator*.

It may be suggested that this is surely precisely where the uncultivated/cultivated morality distinction could usefully be invoked in order to avoid the impending contradiction. But it cannot easily be supposed that “broader” (along with “narrower”) partiality is a feature of our “uncultivated,” or “uncorrected,” moral ideas alone. For even in this later section<sup>93</sup> Hume continues, in many remarks, to understand the “correction” to be made by us as spectators as a matter of our trying disinterestedly to contemplate a character trait from the point of view of those who have, or might have, the relevant close ties with the possessor: friends, family, and associates. We are thus exhorted to imagine what *their* moral feelings would be toward him as the particular recipients of his favours and disfavours, rather than to work out the effects, happy or unhappy, on the world at large. In other remarks, Hume even suggests that it is natural – and proper – for us as spectators to “make allowance for [sheer] selfishness” on the part of an opponent whose character we are judging... because [we] know it to be inseparable from human nature, and inherent in our frame and constitution,” and so to refrain from “the blame which so naturally arises on any opposition.”<sup>94</sup> Thus we are presented with a curious hybrid in which the general point of view, adopted by a disinterested spectator, appears *morally* to accommodate – at least in so far as blame is resisted – an exhibition of the narrowest natural partiality by the possessor of the trait being judged, who blatantly considers only himself.

## References

Baier [ 1991] – A. Baier, *Progress of Sentiments*, Harvard University Press, Cambridge, MA 1991.

---

<sup>92</sup> Ibidem, 3.3.1.11.

<sup>93</sup> Ibidem, 3.3.1.16.

<sup>94</sup> Hume [2000] 3.3.1.17.

- Baier [1992] – A. Baier, “Artificial Virtues and the Equally Sensible Non-Knaves: A Response to Gauthier,” *Hume Studies* 18 (2) 1992, p. 429–439.
- Botros [2006] – S. Botros, *Hume, Reason and Morality: A Legacy of Contradiction*, Routledge, Oxford 2006.
- Cohon [2008] – R. Cohon, *Hume’s Morality: Feeling and Fabrication*, Oxford University Press, Oxford 2008.
- Collier [2010] – M. Collier, “Hume’s Theory of Moral Imagination,” *History of Philosophy Quarterly* 27 (3) 2010, p. 255–273.
- Darwall [1995] – S. Darwall, *The British Moralists and the Internal ‘Ought’ 1640–1740*, Cambridge University Press, Cambridge 1995.
- Garrett [2007] – D. Garrett, “The First Motive to Justice: Hume’s ‘Circle Argument’ Squared,” *Hume Studies* 33 (2) 2007, p. 257–288.
- Gauthier [1992] – D. Gauthier, “Artificial Virtues and the Sensible Knave,” *Hume Studies* 18 (2) 1992, p. 401–427.
- Haakonssen [1978] – K. Haakonssen, “Hume’s Obligations,” *Hume Studies* 4 (1) 1978, p. 7–17.
- Harris [2010] – J. Harris, “Hume on the Moral Obligation to Justice,” *Hume Studies* 36 (1) 2010, p. 25–50.
- Hume [2000] – D. Hume, *A Treatise of Human Nature*, D.F. Norton and M.J. Norton (eds.), Oxford University Press, Oxford 2000.
- Taylor [1998] – J. Taylor, “Justice and the Foundation of Social Morality in Hume’s *Treatise*,” *Hume Studies* 24 (1) 1998, p. 5–30.
- Wiggins [2006] – D. Wiggins, *Ethics: Twelve Lectures on the Philosophy of Morality*, Penguin, London 2006.