

## Hume and Mutual Advantage

John Salter

University of Manchester

[john.salter@manchester.ac.uk](mailto:john.salter@manchester.ac.uk)

January 2009

Hume's theory of justice consists of three components: a theory of the origins of justice, which is based on the principles of self-interest and mutual advantage, a separate explanation of the particular rules of justice, which drew on Hume's theory of the imagination and the principle of utility, and finally a theory of the artificial virtues in which our approbation of justice and the duty to be obey the laws of justice are grounded in sympathy with the public interest.

It is the principle of mutual advantage, however, that has come to occupy centre stage in contemporary interpretations of Hume. His theory of justice as a whole is now routinely classified as a theory of 'justice as mutual advantage' and it has been given a prominent position in a history of such theories, which begins with the sceptical views of Glaucon and Thrasymachus in Plato's *Republic*, which includes the theories of Hobbes and Hume, and is represented in the modern era by David Gauthier and James Buchanan.

As a consequence of this modern reading, Hume's theory of justice is widely thought to manifest all the unattractive features of theories of justice as mutual advantage. In particular, the terms of justice, being the result of some kind of bargain or agreement, reflect the initial bargaining power of the parties involved. The weak inevitably receive relatively little and in some cases nothing. Justice, far from protecting the weak and poor, is little more, as

Thrasymachus claimed, than what is in the interest of the stronger party, a conclusion that Hume is commonly thought to be drawing explicitly in the *Enquiry* when he argues that justice is only possible amongst creatures of roughly equal strength.

Allen Buchanan has pointed out that in excluding the weak and defenceless from the protection of justice, theories of justice as mutual advantage clash both with common sense morality and with our most fundamental legal institutions.<sup>1</sup> This may not, in itself, be a serious problem for normative theories of distributive justice, which are supposed to yield principles by which existing institutions are to be judged and common sense morality to be challenged. But for Hume, whose primary purpose was to *explain* both existing legal institutions and common moral beliefs, it is a damning criticism.

My aim in this paper is to defend Hume's theory from this critical reading. I shall argue that it is not the case that Hume excludes the weak and disabled from the scope and protection of justice, nor more generally, is it the case that the terms of justice reflect bargaining power. The critical reading in question goes astray because it fails to see that mutual advantage is only part of Hume's theory, the part that explains the origins of justice in a very general sense, and that this is bracketed off from those parts of the theory that explain who is included within the scope of justice, how much each receives, and why and to whom we have a duty to be just.

The interpretation of Hume's theory of justice as a theory of justice as mutual advantage is thus an attenuated and misleading one, which has obscured Hume's own particular purposes as well as the coherence of his

moral and political theory as a whole. His attempt to show, against the proponents of the 'selfish system of morals', that although justice originates with self interest, our moral approbation of justice, and our duty to be just, involves adopting an impartial and disinterested point of view, is either lost sight of altogether or is judged to be incompatible with the main thrust of Hume's theory of justice as mutual advantage and part of a completely different kind of theory. At the same time, Hume's political purposes in challenging the different forms of contractarianism that were dominant in his day are entirely overlooked. In both these areas the interpretation of Hume's theory as a theory of justice as mutual advantage not only fails to convey Hume's complex purposes, but it portrays Hume's own theory as the kind of theory he was most concerned to refute.

When an interpretation of a past thinker becomes as well established as the one under consideration, it is important to understand why it has taken hold in the way it has. With this in mind, part I attempts to set out as clearly as I can the textual basis for the interpretation as well as the principle theoretical concerns that have driven it. In part II I challenge the interpretation in its two most prominent forms: what I shall call the standard version, which is based on Hume's discussion of the circumstances of justice in the *Enquiry* and the version that Brian Barry presents in the second volume of *A Treatise on Social Justice*. I conclude with some observations on the question of the nature of Hume's supposed conservatism.

I

### **The Circumstances and Scope of Justice**

The most direct textual evidence that Hume's theory of justice excludes the weak and defenceless from the scope of justice is the following passage from the *Enquiry*. Because of the centrality of this passage to what follows, it is necessary to quote it in full.

'Were there a species of creatures intermingled with men, which though rational, were possessed of such inferior strength, both of body and mind, that they were incapable of all resistance, and could never, upon the highest provocation, make us feel the effects of their resentment; the necessary consequence, I think, is that we should be bound by the laws of humanity to give gentle usage to these creatures, but should not, properly speaking, lie under any restraint of justice with regard to them, nor could they possess any right or property, exclusive of such arbitrary lords. . . .[A]s no inconvenience ever results from the exercise of a power, so firmly established in nature, the restraints of justice and property, being totally *useless*, would never have place in so unequal a confederacy.

This is plainly the situation of men, with regard to animals; and how far these may be said to possess reason, I leave it to others to determine. The great superiority of civilized Europeans above barbarous Indians, tempted us to imagine ourselves on the same footing with regard to

them, and make us throw off all restraints of justice, and even of humanity, in our treatment of them.’<sup>2</sup>

Despite the fact that Hume here refers to a ‘species of creature’, intermingled with human beings, the passage is usually interpreted as a veiled reference to inferior human beings. Martha Nussbaum, for example, thinks that Hume is applying his insights about the need for rough equality of strength ‘explicitly to the case of disability’<sup>3</sup>. After quoting the central passage, she says: ‘His exclusion of people with severe disabilities, . . . derives solely from his focus on rough equality of power as among the Circumstances of Justice.’ And she concludes: ‘In short, the much weaker, whether in body or in mind, are simply not part of political society, not subjects of justice.’<sup>4</sup> The essential point identified by Nussbaum is that ‘we do not need to cooperate with people who are much weaker than the normal case, because we can simply dominate them, as we now dominate nonhuman animals.’<sup>5</sup> Allen Buchanan interprets the passage in a similar way. It shows, Buchanan says, that in Hume’s theory our rights depend entirely on our capacity to harm others. Buchanan perceives a family resemblance between Hume’s theory and David Gauthier’s theory. In both, our rights depend on our strategic capacities – in Gauthier’s case, our capacity to contribute productively to the cooperative surplus and in Hume’s case our capacity to inflict harm. Those who cannot harm others, or cannot retaliate to the harm done against them, ‘are entitled to nothing because they have nothing to offer, nothing with which to bargain.’<sup>6</sup> Buchanan calls theories of this kind theories of reciprocity and contrasts them with subject centred theories, according to which, what we get is determined by such things as need or desert. Jonathan Wolff says that the

passage in question ‘rather chillingly’ illustrates the idea that justice is a type of bargain, in which ‘might makes right’. The logic of Hume’s theory, according to Wolff, is that ‘if others have nothing to offer us, or things we can take from them independently of what they decide or want, then we have no duties of justice, strictly speaking, towards them.’<sup>7</sup>

In *Theories of Justice*, Brian Barry, who provides by far the most detailed discussion of Hume’s theory as a theory of mutual advantage, reads the *Enquiry* passage in much the same way as the other theorists referred to above. Barry does acknowledge that Hume says that the Europeans only *imagined* themselves to be above justice in relation to the Indians. But he says that ‘Hume must be accused of drawing back from the full implications of his doctrine . . . . . on his theory, they were above justice in relation to the Indians’ because they could impose their will by force.<sup>8</sup> The implications, which Hume was not prepared to face up to, are that justice was entirely useless to the Europeans, who therefore had no reason to seek agreements with the Indians or to restrain their self interest in any other way. Barry is sufficiently confident of this reading of the passage to systematically replace Hume’s phrase ‘species of creatures’ with ‘race of creatures’.

In *Justice as Impartiality*, however, Barry revised his understanding of the passage, along with its applicability to the encounter between the Europeans and the American Indians. As Barry points out, the Europeans did seek agreements with the native inhabitants because the latter were not in fact ‘incapable of all resistance’ as were the imagined ‘race of creatures’ referred to by Hume.<sup>9</sup> However, the inequality of power, while not so extreme that the Europeans could have achieved all they wanted by force, enabled

them to renege on the terms of the treaties whenever it suited them without any repercussions. Moreover, according to mutual advantage theories of justice, *it was not unjust for them to do so*. This, Barry now claims, is the point Hume was 'prepared to accept' in the separate species passage of the *Enquiry*.<sup>10</sup>

Barry reaches this remarkable conclusion in the following way. Theories of justice, Barry maintains, are characterised by the answers they give to two questions: what is the motive for behaving justly, and what is the criterion for a just set of rules. Mutual advantage theories, Hume's included, answer these questions in the following way: the motive for behaving justly is self-interest, and the criterion for a just set of rules is mutual advantage, that is, the rules are just if everyone is better off with them than with no rules. It follows, Barry claims, that:

'A theorist of justice as mutual advantage would have to say that any treatise was just so long as it was better for both parties than the alternative of fighting. He would also have to say that, since the only motive either side could have for adhering to the terms of the treaty was a sense of the gains to both sides if both sides observed it, it would also be just for the whites to tear up the treaty . . . as soon as they reached a position in which it would be more advantageous to drive the Indians by force from the territory they had been granted.'<sup>11</sup>

Now for this question of compliance to arise at all, both sides, having made an agreement, evidently perceived the advantages of cooperation and so neither regarded the other at the time of the agreement as an inferior species in the relevant sense. Furthermore, the signing of agreements over

such things as land ownership and trading rights, implies not merely that they regarded cooperation as mutually advantageous, but that both sides recognised a common set of property rights and conventions about promising. It is evident, therefore, that we have two quite distinct interpretations of the *Enquiry* passage and of the significance of inequality of strength in Hume's theory as a whole. The first, which I will refer to as the standard reading, interprets the passage in the context of what, following Rawls, has come to be known as the circumstances of justice - those facts about human circumstances and dispositions that make justice both useful and necessary.<sup>12</sup> According to this reading a capacity for mutual harm is relevant in the context of Hume's explanation of the *origins* of justice: institutions of property and promising only develop between creatures who have the capacity for mutual harm. To paraphrase Hume's remark about the state of nature: it is not that it is allowable for humans to violate the property of inferior creatures or break their promises to them, rather, there is no such thing as justice and injustice in these relationships and so the question of whether humans have a duty to respect the property of inferior creatures or keep promises to them never arises. When the Europeans encountered the indigenous people of America they, rightly according to this reading, regarded them as an inferior species of being, on the same footing as farm animals, and so they quickly reached the conclusion that they had no duties of justice towards them. This is the interpretation that is to the fore in *Theories of Justice* and which seems to be the reading favoured by the other authors referred to above.

In *Justice as Impartiality* however, Barry introduces the idea that the capacity for mutual harm is relevant more directly in that it is the criterion that



determines whether our treatment of other creatures is just. On this reading the reason why the Europeans could disregard the claims that the indigenous Americans made to their land is because these claims were unjust, and the reason they were unjust was because there were no reciprocal benefits to the Europeans in the mutual respect for each others property. Notice that on this reading, but not on the first, the factors that shape our duties of justice are, so to speak, historically reversible. When the Europeans first arrived in the new world they found that they needed the cooperation of the inhabitants and since the inhabitants also needed their cooperation, both sides had a duty to treat each other justly. But this all changed when the Europeans, largely through growing numbers and greater independence, found that they could do without the restraint and cooperation of the Indians. From that point on, the Europeans no longer had a *reason* to comply, and because the terms of the treaties were no longer just, according to the stipulated criterion, they no longer had a *duty* to comply.

### **The Terms of Justice**

The case of extreme inequality of power is, from the point of view of the standard reading, an extreme case of what is commonly seen as a more general feature of mutual advantage theories, namely that they are all, in one way or another, theories of bargaining. According to Barry, for example, all mutual advantage theories explain the terms of justice as the outcome of a self-interested bargain, or settlement of some kind, in which 'the amount received by the parties under the settlement must reflect their relative bargaining power.' The whole point of such theories, Barry says, is to 'translate strength into advantage as smoothly as possible.'<sup>13</sup> In *Theories of*

*Justice* Barry presents a systematic and detailed analysis of mutual advantage theories using two person bargaining theory, and he compares the solutions of John Nash, R. B. Braithwaite and David Gauthier to what he takes to be the central problem of justice – dividing up a scarce resource between self interested parties. The case of extreme inequality of power discussed by Hume in the *Enquiry* is interpreted in terms of Braithwaite's example of the two musicians, Matthew and Luke, who live in adjacent apartments and have the same hour of the day in which to practise. It is, Barry says, a generalisation of the case where one of the musicians is unaffected by the practising of the other and does not, therefore, need an agreement.<sup>14</sup>

Barry's interest in conveying Hume's ideas through the theories of other, more contemporary, writers is also evident in his claim that Hume's theory is much the same as the theory advanced by James Buchanan, according to which, fighting in a state of nature establishes a natural equilibrium, in which there is no further advantage to any one in continued predation. This establishes *de facto* possession as the non-agreement point, and the basis for subsequent contractual arrangements, whereby 'sheer physical possession gets transformed into assured property'. Thus, Barry says: 'the position people have at the non-agreement point (possession) form the basis for the move to the Pareto frontier (property). Those with more at the non-agreement point retain their advantage'. This reading enables Barry to characterize Hume's theory as a two-stage bargaining theory. The most important point he takes from Hume's discussion of the rules of property is that 'the determination of justice in Hume's theory requires a reference to what people had at the non-agreement point.'<sup>15</sup>

Jeremy Waldron also takes James Buchanan's theory of the emergence of property rights as the appropriate modern exemplar of Hume's model. According to Waldron: 'On the Humean model, we start from an assumption of conflict driven by limited altruism and moderate scarcity. People grab things and use them; they argue and fight over them. Over time, the holdings determined in this way are going to be largely arbitrary. Nevertheless, if any sort of stable pattern emerges, then something like a peace dividend may be available.'<sup>16</sup> Waldron discusses a number of technical issues involved with this so called Humean model, such as the bargaining or transaction costs of reaching precise terms of the agreement and the problem of instability – people who do less well from the bargain 'will try constantly to force renegotiation to redistribute the cooperative surplus.'<sup>17</sup> But the most serious deficiencies of Hume's theory, Waldron thinks are not technical but moral. It is all, he says, 'a matter of grabbing and fighting and defending what one can' and some people will inevitably end up with nothing – 'Hungry, cold, terrified, and despairing, they will doubtless face a proposal to ratify and legitimize their dispossession with considerable alarm.'<sup>18</sup>

## II

This critical interpretation of Hume has not gone unchallenged. Thomas Pogge, in his (1990) review of *Theories of Justice* argued that Barry's interpretation of Hume's reference to the encounter between Europeans and American Indians conflates explanation and justification. Hume was not making a *normative* point about inequality but a *factual* point about the origins of justice. It is certainly true that Barry entirely overlooks this distinction.

However, the factual point in question is about who is included within the scope of justice. So if Barry is correct in his interpretation of what Hume says about this factual point, the implication is that duties of justice are not owed to human beings who are incapable of causing harm to others.<sup>19</sup>

Arthur Kuflik has also defended Hume from his critics, arguing that in view of what Hume has to say about justice elsewhere in the *Treatise* and the *Enquiry*, the suggestion that Hume thought weak Human beings were outside of the protection of justice is an ‘obviously false or highly implausible claim.’<sup>20</sup> It is particularly at odds, Kuflik thinks, with Hume’s opposition to the idea that self-love is the foundation of morality, and his attempt to ground our duties of justice on an impartial regard for the public interest. However, Kuflik accepts that a literal reading of the ‘separate species’ passage of the *Enquiry* suggests that the Europeans *were* unencumbered by duties of justice, and that this cannot be squared with Hume’s remark that they only imagined themselves to be so. Kuflik concludes that it is necessary to ‘discount some elements of Hume’s text.’<sup>21</sup> In particular, he thinks we need to disregard those aspects of the text that suggest Hume actually meant that rough equality of power was one of the circumstances of justice.<sup>22</sup>

Can we interpret this passage in a way that does not require that we disregard any of its components, that acknowledges that it is clearly about equality of strength, but that it refers to an imaginary species of creatures rather than to inferior human beings, and that the Europeans only *imagined* they were above justice and that their treatment of the indigenous Americans *was* unjust? Is there a reading that is also consistent with what Hume

elsewhere says about justice, in particular with his account of the virtue and duty to be just?

### **The Duty to be Just**

It is interesting to note that Barry is struck just as much as Kuflik by an apparent contradiction between Hume's emphasis on a capacity for mutual harm on the one hand, and his attempt to ground the virtue of justice on an impartial regard for the public interest on the other. Barry, however, argues that this reflects the presence of two quite distinct and incompatible theories – the theory of justice as mutual advantage, and the theory of justice as impartiality. Now the theory of justice as mutual advantage is a theory of the origins of justice. But according to Barry, as we have seen, it also answers the two questions that he says any theory of justice must answer - the question of motive and the question of the criterion of a just set of rules. Justice as impartiality is the term Barry uses for Hume's explanation of why justice is a virtue – why it is morally approved of and why injustice is disapproved of. Thus, Barry says, 'What makes justice a virtue is not that it is mutually advantageous, but something else about it'. That something else is the fact that justice has beneficial consequences for society in general, and the fact that we sympathise with the beneficiaries of justice and with the victims of injustice.

The consequence of this 'two theory' reading of Hume is, according to Barry, that the rules or institutions of justice can be assessed from two entirely different viewpoints: the viewpoint of mutual advantage and the impartial viewpoint. This incompatibility is nowhere more apparent than in the issue at hand – relationships involving extreme differences in power. Institutions that

simply reflect relations of power and exclude the weak entirely from their scope, but are nevertheless just from the point of view of the stipulated criteria of mutual advantage, can clearly be criticized from the impartial viewpoint Hume adopts when discussing why justice is a virtue. The fact that Hume did not pursue this line of thought, Barry declares, shows that he was 'a better conservative than he was a philosopher.'<sup>23</sup>

It is instructive at this point to note that David Gauthier, in making the case for regarding Hume as a contractarian, provides much the same reading of Hume as Barry does, although being more favourably disposed to Hume, he does not draw attention to the problematic nature of the reading. Gauthier explains that when Hume says in the *Enquiry* that 'public utility is the *sole* origin of justice' he means *mutual expected utility* 'so that a rule or practice has public utility if and only if each person reasonably expects that rule or practice to be useful to himself.'<sup>24</sup> But when Hume says that 'reflections on the beneficial consequences of this virtue are the *sole* foundations of its merit' he is not referring to mutual expected utility but to the usefulness of the rule to human society in general.<sup>25</sup> Thus, Gauthier says 'we distinguish public utility as the origin of justice, from general utility, or overall advantage, as the basis of our moral approbation of justice.'<sup>26</sup> But having made this important distinction, Gauthier then confuses the issue by saying: 'Arrangements may be expected to be useful to each person; therefore they are just. These arrangements may also be expected to have beneficial consequences; therefore they receive moral approval, and justice is a virtue.'<sup>27</sup> But if this is the correct reading of Hume, then it is easy to see how our approbation of justice, our impartial reflections on the virtue of justice, are separate from, and

may well conflict with our judgements about whether the arrangements in question are in fact just. And since our duty to be just is derived from our moral approbation of justice, we could easily have a duty to obey rules that are not just. This contradiction, to which Barry draws attention, is an inescapable result of the attempt by both Barry and Gauthier, to present Hume as a contractarian.

It is now possible to see that Barry's two theory interpretation of Hume, and the interpretation of Hume's 'separate species' passage that he puts forward in *Justice as Impartiality*, are closely related and follow from the same interpretive step, namely: *to read the mutual advantage component of Hume's theory not only as an explanation of the origins of justice, but also as a theory that is intended to justify the rules of justice and to explain the motive and duty to be just*. In other words, the principle of mutual advantage is taken to underpin the whole of Hume's theory of justice.

However, this contains two serious interpretative mistakes. First, according to Hume, the ordinary motive for behaving justly is not self interest but a sense of duty, which is entirely separate from the self interested considerations that explain the origins of justice.<sup>28</sup> Hume explains the foundation of this duty in the second part of Book III, Part II, Section II of the *Treatise* where he turns to the question 'Why we annexe the idea of virtue to justice, and of vice to injustice.' He has already explained that self interest is the original motive for the establishment of justice, and he says that in the early stages of social development, when societies are small, self interest also accounts for compliance with the rules. As societies grow, however, this interest becomes more remote. We no longer see that our own compliance

contributes to the maintenance of the whole system. But we never fail to see the harm done to ourselves by the injustices of others. We are also aware of and moved by the harm that is done by injustices to society as a whole and to other individuals: 'when the injustice is so distant from us, as no way to affect our interest, it still displeases us; because we consider it prejudicial to human society, and pernicious to every one that approaches the person guilty of it.' It is this displeasure, or 'uneasiness' with acts of injustice and our pleasure or satisfaction with just acts that explains our sense of moral right or wrong; it explains why we regard injustice as a vice and justice a virtue and also why we regard justice as a duty.

The duty to be just is not then, for Hume something externally given, either in the sense of deriving from a higher authority, or from natural law, or from the justness or unjustness of the rules in question. It is simply the motivating force of our sense of moral right and wrong that becomes attached to acting justly as a result of the accumulated experience of living in societies bound together by justice and the knowledge of the harmful consequences of injustice.

Barry's second mistake, which he shares with Gauthier, concerns what he says about the criterion for a just set of rules. In Hume's theory, as Barry himself acknowledges in other contexts, there is no criterion for a just set of rules. Justice and injustice for Hume are not qualities of rules or institutions or arrangements, rather they are qualities of people in complying with or failing to comply with the prevailing rules of justice. It is not the case, therefore, that Hume's theory provides us with two conflicting viewpoints for evaluating the rules of justice. In fact, there is no viewpoint for evaluating the rules of justice



– there is a single, impartial, viewpoint for evaluating the merit of justice and this provides the grounds for our duty to be just as explained above.

What are the implications of all this for the encounter between the Europeans and the Indians and for relationships between the strong and the weak generally? Much depends on what we make of Hume's use of the phrase *human society* when he says that injustice displeases us 'because we consider it prejudicial to human society, and pernicious to every one that approaches the person guilty of it.' Certainly there is no reason to suppose that at this point in the *Treatise* he was talking about *particular* societies, separated by geographical, national or political boundaries. It would also be a forced and implausible reading to think that in this context Hume meant only able bodied human beings, those capable of mutual harm. The same impression is created by Hume's discussion in the *Enquiry*, where he says that the duty to be just is founded on its usefulness: 'Usefulness is agreeable, and engages our approbation.' He then asks: 'But *useful*? For what? For somebody's interest surely. Whose interest then? Not our own only: For our approbation frequently extends farther. It must, therefore, be the interest of those; who are served by the character or action approved of; and these we may conclude, however remote, are not totally indifferent to us.'<sup>29</sup> So it seems that our duty to be just extends to whoever benefits from the character or actions of the just person, a definition sufficiently broad to include the indigenous inhabitants of America and the weaker members of our own society.

However, we cannot reach this conclusion quite so easily. It cannot simply be that we owe duties of justice to all creatures that can benefit from

them since animals can certainly benefit from the kind of property rights Hume has in mind. Now in view of what Hume says about the principle of sympathy, it is not hard to see how behaviour that originates with self-interested co-operation can be extended beyond the original group once the relationship between the behaviour and the public interest comes into view. And the fact that, as human beings, we can only sympathise with the pleasures and pains of other human beings provides the limits to how far this behaviour can extend. However, if what I have called the standard reading of Hume is correct, then I doubt whether this would be wholly convincing for Hume's sternest critics. So it is to the standard reading that I now turn.

### **The Origins of Justice**

Hume's explanation of the origins of justice appears in Book III, Part II, Section II of the *Treatise* where he presents his famous theory of the convention. The convention, he says, can be seen as a kind of agreement in as much as '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.'<sup>30</sup> However, it differs from the traditional idea of a contract in at least two ways: it does not involve a promise, which Hume says is itself part of justice, and it is a dynamic evolutionary process of some kind, which leads to the gradual acceptance of the conventional behaviour, and at the same time, *to the formation of the group within which the convention is accepted.* Relations of justice are thus established only between creatures that have the capacity to harm each other, to prey on each other's possessions and to retaliate when preyed upon, and this seems to be the point that Hume graphically underlines with the separate species passage in the *Enquiry*.

Perhaps the most natural way of exploring the implications of Hume's theory of the convention for the encounter between the Europeans and the indigenous inhabitants of America is to think of it as some kind of return to a state of nature. Neither side of the encounter was encumbered by the restraints of justice so neither could be accused of acting unjustly. Moreover, contact between them taught the Europeans that they had no need to restrain themselves by entering into any kind of convention with the Indians because they could get all they wanted by force. So relations of justice never developed between them.

This way of thinking about the European ventures into the new world has an important antecedent in the theory of John Locke, but I do not think that it fits Hume's theory. On Hume's view it is not possible to go back to a state of nature. It is a hypothetical time before justice, not merely in the sense that there were no property rights in existence, but in the sense that *the very ideas* of justice, property and promising had never occurred to people. By contrast, both the Europeans and the Indians had long established practices of justice. The fact that there were no past agreements or bargains between the two sides, the fact that they had not even come into contact before, is not enough by itself to establish the conclusion that they did not have mutual duties of justice. So if the Europeans were justified in their treatment of the indigenous people it could only be because they correctly identified them as being the kind of creatures to whom they do not owe duties of justice. But what possible grounds could they have for reaching such conclusions?

Now on any plausible reading of Hume's convention, the answer to this question must depend on the consequences of the evolutionary dynamics of

group formation not on whether or not the Europeans could actually dominate the Indians at the time of the encounter. The reason human beings do not typically show the kind of self restraint associated with justice in their dealings with farm animals is not because they reason that they have the better option of getting all they want by force. Refraining from the possessions of farm animals is an idea that never occurs to human beings. And the reason it never occurs to us must be sought in the evolutionary dynamics of group selection. Our behaviour towards other human beings and other species, whether we treat them justly or not, is not, according to Hume, determined by a continual process of calculation of advantage. It is determined by conventions, rules and habits. Calculation of advantage may explain the origin of those conventions, rules, and habits, but in our everyday dealings with others we are not conscious of these origins. So the *explanation* for our differential treatment of farm animals and other human beings may, as Hume thinks, include the fact that in our distant history our capacity to dominate farm animals shaped our institutions and ideas, but our *reasons* for our differential treatment of farm animals and other human beings, and our moral beliefs governing our differential treatment of them, have completely different explanations and this is reflected in the two part structure of Hume's theory.

Now I think we can safely conclude from this that Hume cannot be accused of drawing back from the implications of his theory in the 'separate species' passage of the *Enquiry*. The superiority of the Europeans consisted in the superiority of their weapons and this was irrelevant from the point of view of the dynamics of group selection. Military superiority no more justified the European treatment of the inhabitants of the new world than it would have

justified the English in invading Scotland and seizing the lands and possessions of the Scots. It is for this reason that Hume says that the superiority of the Europeans *tempted them to imagine* that they were on the same footing with the indigenous inhabitants as they were with animals. His point is that their military superiority was misconstrued in the imagination of the Europeans as the kind of natural superiority they had over animals. It is the critical note in this phrase that is so striking: Hume is criticising, not justifying their actions.

What then, are the implications of Hume's theory for the situation of people with natural disabilities of a kind that could influence the structure of interaction in a state of nature and the subsequent process of group selection? Hume's critics may be wrong in thinking that the 'separate species' passage is a veiled reference to weak human beings, but are they not right in thinking that the implications of Hume's theory are that relations of justice are established only between people who have the capacity to harm each other?

The answer depends on whether the criteria for inclusion within the group operate at the level of the individual or at the level of the species. Is Hume's convention of the form: I abstain from the possessions of x because I need x's co-operation, or is it of the form: I abstain from the possessions of x because x is the sort of creature with whom it is beneficial to cooperate?

The tendency to see Hume's convention as a kind of bargain encourages the view that the criteria operate at the level of the individual. The convention is typically seen as a series of two person bargains each with a determinate outcome shaped by the relative strengths of each pair – Matthew is in the circumstances of justice with Luke, but not with Peter so there are

relations of justice between Matthew and Luke but not between Matthew and Peter.<sup>31</sup> Human society thus becomes fragmented into the strong and the weak and justice emerges as a relationship only between human beings who are capable of mutual harm, leaving the rest to form a group that is regarded in much the same way as animals, except that their superior intelligence makes them more suited to exploitation by the strong.

The alternative reading of Hume's convention, that the criteria for inclusion within the group operate at the level of species characteristics, is best illustrated theoretically, not by a two person bargaining model but by a repeated anonymous game in which we learn which general strategy works the best against the population as a whole, but we do not associate the individual outcomes with the characteristics of each opponent.<sup>32</sup> For a number of reasons I want to suggest that this is the superior reading.

If we are to discriminate between different human beings and moderate our behaviour towards the strong but not the weak, it must be possible for us to identify the characteristics of our opponents independently of their behaviour, and to keep track of the payoffs from each encounter. And there must be some way of identifying the characteristics of "weakness" in people we meet for the first time. However, Hume's account suggests that the information that shapes our strategies in the state of nature is derived from our general experience of interacting with others. We do not know, therefore, what the chances are of coming out on top in any particular encounter, nor what the costs of doing so are likely to be. We only have our past experience to go on and what we learn from this is that preying on the possessions of other human beings is dangerous, not in the sense that we encounter

retaliation on every occasion, but on a sufficient number of occasions to teach us that forbearance is the best general strategy. We thus develop a general *disposition* to forbear from taking the possessions of other human beings, including those who have no capacity or desire to harm us, and including those we haven't even met before.

We will find, of course, that on some occasions we do better through aggression, when for example, our predation meets neither resistance nor retaliation. This may be because our 'opponent' is much weaker than us and has no capacity to resist. But it may also be because she is adopting a different strategy – she may be ahead of us in the game, realizing before we do that restraint is the best policy. Even if we wanted to, however, we cannot infer which kind of opponent we are dealing with from their behaviour alone. Our problem is that there is no distinct recognizable class of humans who stand out from the rest as being weak or disabled. Rather, there is a range of human strength and ability, and the 'weak' are weak in different ways and to different degrees. But if we imagine, along with Hume, that there are intermingled with us, creatures who are easily and instantly recognisable as belonging to a different species, and that we learn from our interaction with them that creatures of this kind *never* harm us, it is easy to see how we can develop different dispositions towards humans and non-humans.

Whether or not Hume provides us with enough textual evidence to distinguish these two readings, the second is certainly a plausible reading, and it is one that is consistent with what Hume goes on to say about our duty to be just. It thus enables us to avoid Barry's two theory reading, and it enables us to interpret the separate species passage in a straightforward way

without discounting any of its components and without accusing Hume of deliberately avoiding the consequences of his own theory. It is thus the reading that is recommended by the principles of interpretative charity if by nothing else.

### **The Rules of Justice**

I now turn to the question of the terms of justice and to those interpretations of Hume's theory of the convention that portray it as some kind of bargain in which the terms or rules of justice reflect the bargaining power of people in the state of nature. Much of what has been said above should suggest that I regard this as a complete misinterpretation of what Hume meant by convention, but the point I now want to stress is that Hume's convention explains the origins of justice in a rather general sense; it does not explain the origins of the rules or terms of justice. The convention establishes 'the general rule, *that possession must be stable*' but, Hume says, this rule must be 'modified' and 'applied' by other general rules, which 'must extend to the whole society'.<sup>33</sup> Hume's explanation of these other general rules comes in Section III, where he makes use of his theory of the imagination and of the principle of utility, principles that are quite distinct from and incompatible with the idea of a bargain between people in a state of nature.<sup>34</sup>

Consider the rule of present possession. Waldron paraphrases Hume's convention in the following way: 'I agree to respect what you have managed to hang on to, and you agree to respect what I have managed to hang on to: "By this means, everyone knows what he may safely possess".'<sup>35</sup> This reading suggests a direct link between the convention (understood as a bargain) and



the rule ratifying present possession. Similarly, Barry's two-stage bargaining theory interpretation of the present possession rule construes it as the result of a bargain, in which the advantages at the non-agreement point are translated to property rights. However, there is no support for such a reading in the text. In the *Treatise*, Hume gives two alternative explanations of the rule of present possession. The first, which appears in the main body of the text has to do with custom and our preference for things we already have and are used to: 'Such is the effect of custom, that it not only reconciles us to any thing we have long enjoy'd, but even gives us an affection for it, and makes us prefer it to other objects, which may be more valuable, but are less known to us. . . . 'Tis evident, therefore, that men wou'd easily acquiesce in this expedient, *that every one continue to enjoy what he is at present possess'd of.*'<sup>36</sup> Hume's second explanation of the rule of present possession is presented in a footnote where he draws on his theory of the imagination. He says: 'As property forms a relation between a person and an object, 'tis natural to found it on some preceding relation; and as property is nothing but a constant possession, secur'd by the laws of society, 'tis natural to add it to the present possession, which is a relation that resembles it.'<sup>37</sup>

Not only is there no textual support for the Barry/Waldron reading of the convention but it is incompatible with the whole logic and structure of Hume's own explanation. In the *Treatise* Hume says that the rule of present possession is only a temporary expedient that is agreed to at the first formation of society, and that societies subsequently introduce new rules because of the inconvenience to society in general of confining property distribution to the pattern established at the beginning of social development.

Once again, these other rules – first occupation, prescription, accession and succession are explained by Hume in terms of the principles of public utility and the imagination and there is no mention, or even suggestion, that they result from a process of bargaining. And why, in any case, would rational self-interested bargainers who, according to the Barry/Waldron reading, had grabbed more than they could get by their own labour and good fortune agree to a system based on first occupancy and all the other rules Hume discusses? Barry, in fact, seems to be half hearted about this and conscious of the anachronism involved.

Hume's discussion of present possession in the *Enquiry* is very different from the discussion in the *Treatise*. In the *Enquiry* it is considered more traditionally as a ground for a claim where there is no preceding or stronger claim. It is in relation to this that Hume says that the rules are sometimes no more than analogies – slight connections of the imagination. Present possession is here treated as a rule dreamt up by lawyers faced with the need for a rule where there is no stronger ground for awarding the property right. Hume's explanation of the other rules is also different in the *Enquiry* and the principle of utility is to the forefront of Hume's discussion. The imagination is brought in only where utility recommends more than one possible rule.

The main point to stress, however, is that as with the explanation Hume gives in the *Treatise*, the rules are not presented as the result of a bargain and, moreover, the rules themselves do not reflect any trace of bargaining power. First occupancy, for example reflects hunting ability, but also luck and geographical location. It is just as likely to protect the weak against strong

invaders as it is to justify the actions of the strong and it was potentially, and in some cases actually, the principle legal protection of the American Indians against the conquering Europeans. Hume's conservatism, if that is what we should call it, is not always on the side of the strong and powerful.

### III

Charges of conservatism are, in any case, beside the point. The main force behind Barry's remark that Hume was a better conservative than he was a philosopher is Barry's contention that Hume's mutual advantage theory of justice committed him to the idea that bargaining power determines the terms of justice, and that this inevitably conflicts with Hume's view that the moral approbation of justice requires the adoption of an impartial point of view. Hume's supposed failure to face up to this conflict in his own theory is the basis for Barry's dismissive assessment. However, it is because Hume separated his explanation of the origins of justice from his explanation of the rules of justice that he was able to be as open minded as he was about the rules. The fact that they are merely 'conventions', shaped, in part at least by what we think is in the public interest, can be seen to open the door to reform. In the *Enquiry* he says that the choice of rules can depend on 'the sentiment of private humanity and aversion to private hardship, on positive laws, on precedents, analogies, and very fine connexions and turns of the imagination.'<sup>38</sup> Hume only rules out complete equality and distribution according to virtue or merit. Barry points this out, although he condemns

Hume for not carrying out such a normative evaluation of the standard rules himself.

However, Hume had good philosophical reasons for limiting his speculations about different rules of justice. He was acutely aware, as many are today, that moral disagreement blocks any kind of normative consensus for the design and evaluation of the rules of justice.<sup>39</sup> Moreover, what was at stake in Hume's day was not gradual democratic reform of distributive institutions within a well-ordered society, but violent revolution. Hume's targets were the religious and political 'fanatics' that he thought had been so destructive in Hobbes' century, and the supporters of the different currents of political opposition in his own. And it was not so much the reformist intent of the latter that Hume challenged, but the 'speculative and dogmatic theories' – primarily the theory of the social contract - on which their opposition was based. Hume's objection was not just that these currents of opposition sought in different ways the overthrow of a regime that he thought was fulfilling the proper functions of government, but that any alternative regime that replaced it would be equally vulnerable to the charge that it was not justified by an imagined original agreement.

More important than any of this, however, is the matter of Hume's intentions, with which we began this discussion. Hume's theory, like any theory, can only be defended, or criticised, on its own terms. Hume's theory of justice was not a theory of distributive justice, and while the distinction between explanation and justification in Hume's theory is often overplayed, Hume's primary concern was not to provide normative criteria for the evaluation and reform of society's distributive rules. For that reason alone it

does not measure up to the requirements of a theory of justice for most contemporary theorists. But as a theory that sought to explain both common sense morality and our most fundamental legal institutions it is probably the best theory we have.

<sup>1</sup> Allen Buchanan, Justice as Reciprocity versus Subject-Centred Justice, in *Philosophy and Public Affairs*, vol. 19. no. 3 (Summer. 1990), p. 229.

<sup>2</sup> David Hume, *An Enquiry Concerning the Principle of Morals*, ed. L.A. Selby-Bigge, 3<sup>rd</sup> ed. (Oxford, Clarendon Press, 1975) Section III, pp 190-191.

<sup>3</sup> Martha C. Nussbaum, *Frontiers of Justice*, (Cambridge Mass., Harvard University Press, 2006) p. 46.

<sup>4</sup> Ibid., p. 48 and p. 49.

<sup>5</sup> Ibid., p. 61.

<sup>6</sup> Buchanan, *Justice as Reciprocity versus Subject-Centred Justice*.

<sup>7</sup> Jonathan Wolff, Models of Distributive Justice, in *Empathy and Fairness*, Novartis Foundation 2006. pp. 5-6.

<sup>8</sup> Brian Barry, *Theories of Justice*, vol. I of *A Treatise on Social Justice*, (Berkeley, University of California Press, 1989) p. 162.

<sup>9</sup> Brian Barry, *Justice as Impartiality*, vol. II of *A Treatise on Social Justice* (Oxford, Clarendon Press, 1995) p. 45 footnote.

<sup>10</sup> Ibid., pp. 41-42

<sup>11</sup> Ibid., p. 41.

<sup>12</sup> In the *Treatise of Human Nature* Hume supported his argument that justice is an artificial virtue by claiming that justice is only useful, and hence only arises, in circumstances of moderate scarcity and because human beings have limited generosity. In the *Enquiry* Hume made no direct references to the idea that justice is an artificial virtue but relied on a fuller explication of the

circumstances of justice. He also adds the further condition that justice is only necessary and possible where there is rough equality of strength between individuals.

<sup>13</sup> Ibid., p. 39-40.

<sup>14</sup> Barry, *Theories of Justice*, p.161

<sup>15</sup> Ibid., p. 299.

<sup>16</sup> Jeremy Waldron, The Advantages and Disadvantages of the Humean Theory of Property, in Ellen F. Paul, Miller, Fred., jr., Jeffrey Paul, *Property Rights*, (Cambridge, Cambridge University Press, 1994), p. 86.

<sup>17</sup> Ibid., p. 97.

<sup>18</sup> Ibid., p. 105.

<sup>19</sup> Thomas Pogge, Review of *A Treatise on Social Justice*, Vol. I: *Theories of Justice* by Brian Barry, *Journal of Philosophy*, Vol. 87, No. 7. (Jul., 1990), pp. 375-384.

<sup>20</sup> Arthur Kuflik, Hume on Justice to Animals, Indians, and Women, *Hume Studies*, Vol. XXIV, No.1, April, 1998. p. 55.

<sup>21</sup> Ibid.

<sup>22</sup> Kuflik provides an alternative reading of the passage in question, namely: 'what makes Hume's notions of justice and injustice irrelevant, in our dealings with (most kinds of) animals is the inability of those animals to understand and to follow rules' (p. 61). However, there is no support at all for this reading. In fact, Hume seems to be deliberately differentiating his argument from the one

Kuflik proposes by saying that his imaginary species were rational, although weaker than humans.

<sup>23</sup> Barry, *Theories of Justice*, p.164.

<sup>24</sup> David Gauthier, David Hume, Contractarian, *The Philosophical Review*, Vol.88, No. 1 (January, 1979), p. 17.

<sup>25</sup> *Enquiry*, p. 183.

<sup>26</sup> Gauthier, David Hume, Contractarian p. 18.

<sup>27</sup> Ibid.

<sup>28</sup> For a discussion of this see Stephen Darwall, Motive and Obligation in Hume's Ethics, *Nous*, Vol. 27, No. 4 (December, 1993) pp. 415-448.

<sup>29</sup> *Enquiry*, p. 218.

<sup>30</sup> David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2<sup>nd</sup> ed., (Oxford, The Clarendon Press, 1978) p. 490.

<sup>31</sup> See Christopher W. Morris, Moral Standing and rational-choice contractarianism, in Peter Vallentyne ed., *Contractarianism and Rational Choice*, (New York, Cambridge University Press, 1991). Morris says that this is the case explicitly considered by Hume (p.87).

<sup>32</sup> For a discussion of games of this kind see Robert Sugden, *The Economics of Rights, Co-operation and Welfare*, (Oxford, Basil Blackwell, 1986).

<sup>33</sup> *Treatise*, p. 502.

<sup>34</sup> Now there is evidently a problem with this two stage explanation since the convention, as Hume actually describes it in Section II produces a specific rule, namely: 'to bestow stability on the possession of those external goods,



and leave everyone in the peaceable enjoyment of what he may acquire by his fortune and industry.’ Since we are considering here a simple society of hunters and gatherers, this amounts to the rule of first occupation. This complication in Hume’s text deserves greater attention than we can give it here. Nevertheless, it should be clear that Hume’s official view is that the origins of justice and the origins of the rules of justice are two separate questions, to which he gives separate answers.

<sup>35</sup> Waldron, *The Advantages and Disadvantages of the Humean Theory of Property*, p. 86.

<sup>36</sup> *Treatise*, p. 503.

<sup>37</sup> *Ibid.*, p. 504.

<sup>38</sup> *Enquiry*, p. 310

<sup>39</sup> Andrew Lister has argued that Rawls was wrong to think that Hume’s account of the circumstances of justice included ideological and moral conflict as well as conflicts of material selfish interests. However, the only evidence Lister presents to support this relates to ‘reasonable’ ideological and moral conflict and to Hume’s view that ‘the deep and intractable differences that arise from differences of religious and philosophical doctrine are essentially unreasonable.’ Andrew Lister, *Hume and Rawls on the Circumstances and Priority of Justice.*, *History of Political Thought*, Vol. XXVI. No. 4. Autumn 2005.