

John O'Neill

Self-love, Self-interest and the Rational Economic Agent

Abstract: Hume has a special place in justifications of claims made for rational choice theory to offer a unified language and explanatory framework for the social sciences. He is invoked in support of the assumptions characterising the instrumental rationality of agents and the constancy of their motivations across different institutional settings. This paper explores the problems with the expansionary aims of rational choice theory through criticism of these appeals to Hume. First, Hume does not assume constancy. Moreover, Hume's sensitivity to the relationships between institutional setting and individual motivation owes something to the relative transparency of the plural language of vices and virtues that he employs. Second, rational choice theory's minimal modification of Hume's account of the relation of reason and the passions through the introduction of constraints of consistency on preferences is unstable.

1. Rational Choice Theory and Universal Economics

Rational choice theory under the particular interpretation it receives in neo-classical micro-economics¹ has ambitions to unify the social sciences: "the economic approach provides a valuable unified framework for understanding all human behaviour" (Becker 1976, 14). At least two claims are made on its behalf. The first is that it provides a unified language for the social sciences. The concepts of neo-classical economic theory are claimed to be universal in application:

"There is only one social science. What gives economics its imperialist invasive power is that our analytical categories—scarcity, cost, preferences, opportunities etc.—are truly universal in applicability." (Hirschleifer 1985, 66)

¹ This is not the only interpretation of the formal apparatus of rational choice theory (O'Neill 1998a). I leave aside others here and in the following where I refer to rational choice theory I refer to it under this interpretation.

The second and stronger claim is that rational choice theory under its neo-classical interpretation offers not just a unifying language but also explanatory unity to the social sciences. The assumptions that are employed to explain the behaviour of actors in the market can be generalised to explain human behaviour in any social setting.

The claim that rational choice theory offers explanatory unity is based on three groups of assumptions. First, methodological individualism, the assumption that all social phenomena are explicable in terms that only involve reference to the properties and behaviour of individuals. Second, the assumptions used about the instrumental rationality of the agent in the market who aims to maximise expected utility: the agent is rational in the sense of aiming, under the constraint of some finite budget, to maximise preference satisfaction given complete and consistent assignments of probabilities to outcomes of actions and given a particular preference structure—preferences are complete, reflexive, and transitive. Third, the assumption that the agent has, in addition to minimal virtues of instrumental rationality, the virtue of constancy: the same assumptions can be employed to explain the behaviour of an actor in different roles: specifically, the axioms that characterise the agent as an actor in markets apply equally to all other institutional settings. The scope of economic theory is thus quite general. In principle the behaviour of individuals in any social setting, be it politics, the family, the scientific community, or any other association, can be explained by the same set of assumptions. The assumptions that are taken to have this universal application here can themselves be understood minimally—to refer only to those describing instrumental rationality—or more strongly to include assumptions that the agent is self-interested. The use of the stronger assumptions concerning self-interest is perhaps best exemplified by public choice theory. If individuals act as rational self-interested agents in the marketplace, “the inference should be that they will also act similarly in other and nonmarket behavioural settings” (Buchanan 1972, 22). Against the ‘conventional wisdom’ which is taken to hold “that the market is made up of private citizens trying to benefit themselves, but that government is concerned with something called the public interest” (Tullock 1970, v), the public choice theorist claims that the same assumptions about the agent should apply in both. Specifically, the axioms that characterise the rational self-interested agent in economic life should be taken to apply also to the explanation of the behaviour of bureaucrat and politician in their political activities.

Hume has a special place in justifications of the expansionary project for rational choice theory. He does so for two reasons. In the first place he is the theorist who is normally invoked in justifying the second set of assumptions in rational choice theory concerning the instrumental rationality of agents. Rational choice theory is taken to start from the Humean assumption that reason

only serves the passions (Hume 1978, II.ii.3). Passions or preferences cannot be themselves either rational or irrational. While this is the starting point, rational choice theory does modify the assumption, by adding constraints of consistency on preferences. Most rational choice theorists are 'moderate' rather than 'extreme' Humeans (Broome 1993). Secondly, Hume is also taken to provide support for the third assumption, that of constancy. Appeal is made to Hume's general claim that humans are moved by certain "constant and universal principles of human nature" (Hume 1975, 65).² Indeed the more specific claim of public choice theory, that in politics every man pursues private interests, is sometimes traced back to Hume:

"Political writers have established it as a maxim, that, in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed to be a knave, and to have no other end, in all his actions, than private interest." ('Of the Independency of Parliament', Hume 1985, 42)³

In this paper I want to dispute these appeals to Hume as a foundation for rational choice theory and its expansionary ambitions. I do not do so primarily for reasons of scholarship, to set the historical record right, nor still to defend Hume. It is rather that the way the Humean justifications fail points to problems with the assumptions and unifying aims of rational choice theory under its neo-classical interpretation.

In section 2 I consider the claim that Hume offers grounds for the assumption of constancy. While Hume did hold that certain dispositions were constant and universal features of the human condition—"ambition, avarice, self-love, vanity, friendship, generosity, public spirit" (Hume 1975, 65)—he goes on in the following passages of the *Enquiry* to qualify that claim: "We must not, however, expect that this uniformity of human actions should be carried to such a length as that all men, in the same circumstances, will always act in exactly the same manner. Such a uniformity in every particular, is found in no part of nature" (Hume 1975, 65).⁴ In particular, Hume's political economy does not support the assumption of constancy which is required to found the expansionary aims of economics, i.e. the assumption that the postulates employed to explain the behaviour of actors in markets can be generalised to explain behaviour in non-market contexts. Indeed it would be surprising if it did, since, with other classical political economists, a major

² See for example Dunleavy 1991, 99–100.

³ The appeal to Hume as founding politics upon self-interest is found for example in Buchanan 1962, 314–5.

⁴ The point runs through Hume's *Histories*: "Each century has its peculiar mode in conducting business; and men, guided more by custom than by reason, follow without enquiry, the manners which are prevalent in their own time." (Hume 1983, Volume II, 86. See also 'Of National Characters', Hume 1985)

explanatory problem he was concerned to address was the specification of the conditions in which commercial society could flourish in the first place.

Hume's political economy illustrates a more general point that has implication for questions about the adequacy of the language of rational choice theory. Not only is Hume's account of the motivational structures of individuals in different commercial and non-commercial societies more nuanced than the rational choice portrait would suggest, his work illustrates a shift in the language of economic theory, a shift in which the eighteenth century political economy was of particular significance. Hume's sensitivity to the relationships between institutional setting and individual motivation owes something to the relative transparency of the plural language of vices and virtues that he employs which is lost in the transformation into the more austere language of self-interest, egoism and utility-maximisation in which modern economic theory is written. In section 2 I will argue that the very language in which the discussion of rational choice theory is framed lends undeserved plausibility to its unifying claims. The claim that modern economic theory offers the proper unifying language for the social sciences is mistaken.

In section 3 I consider more briefly the Humean basis for rational choice assumptions about instrumental rationality. Here I suggest that rational choice theory's minimal modification of Hume's account of the relation of reason and the passions through the introduction of constraints of consistency on preferences is unstable. The difficulties point to problems with the basic assumptions of 'rational choice theory' as it is conceived in neo-classical economic theory.

2. From Augustinian to Neo-Classical Economics

"Thus by means and help of . . . commerce all necessities for this life are supplied without intermixing charity with it. So that in states where charity has no admittance, because true religion is banished there, men do not cease to live with as much peace, safety, and comfort, as if they lived in a Republic of Saints." (Nicole 1997, 3)

"There is by now a large and fairly imposing line of economists from Adam Smith to the present who have sought to show that a decentralised economy motivated by self interest and guided by price signals would be compatible with a coherent disposition of economic resources that could be regarded, in a well defined sense, as superior to a large class of possible alternative dispositions." (Arrow/Hahn 1971, vi-vii)

Contrast and compare the following two sets of questions:

1. Is there a universal tendency to avarice in all persons?
2. If avarice is a universal disposition in all persons can a good social order be constructed?
3. Is avarice conducive to individual well-being?
- 1*. Is there a universal tendency in all persons to maximise their own expected utility?
- 2*. If there is a universal disposition to the maximisation of personal utility could a good social order be constructed?
- 3*. Is the pursuit of the maximisation of personal expected utility conducive to individual well-being?

The first set of questions has a long history in the Augustinian tradition of political thought (Deane 1963). To those question the Augustinian answers are roughly as follows: to the third question, "no",—avarice, the insatiable appetite for worldly things is a source of human unhappiness; to the first question, "yes" for we are fallen creatures; and to the second question also a qualified "yes", an earthly political order of sinners is possible and indeed has to constructed on the basis of the assumption that we are fallen beings. The second set of questions has a much shorter history. They are not questions that economists have asked 'ever since Adam Smith': Smith asked different questions. They rather define part of the project of modern welfare economics in the neo-classical tradition and within that tradition the answer to all three questions is, with some qualifications, "yes": the positive answer to the first question is an axiom defining the rational economic agent; the positive answer to the second is a basic theorem of welfare economics—ideal markets issue in Pareto optimal outcomes; a positive answer to the third question is more contentious—there are economists who want to distinguish 'utility' and 'welfare'⁵—but in most economic textbooks on the new welfare economics it is taken to be true in virtue of the fact the terms "utility" and "well-being" are interchangeable so that to maximise personal utility is to maximise well-being.

⁵ Samuelson, for example, interprets the shift to a formal definition of utility as "a steady removal of moral, utilitarian, welfare connotations from the concept" (Samuelson 1938, 344). Hence his comment: "any connection between utility ... and any welfare concept is disavowed." (Samuelson 1937, 161) Broome makes similar divorce of the concepts without the commitment to positivism (Broome 1991, 65). If that line is taken then the answer to our question 3* should not be a straightforward "yes"—no immediate conclusions about welfare follow from the results of modern economics about the relation between ideal markets and utility.

There are parallels between the questions, and in their more popular expositions neo-classical economists sometimes shift into more Augustinian turns of phrase. Not only are there parallels between the questions, there is an historical story to be told as to how the second set of questions developed in part from the first: the Augustinian passage from the seventeenth century Jansenist, Pierre Nicole, that opens this section, is ancestor to the passage from Arrow and Hahn that follows. It is a story to which I will return briefly later. However, while there are parallels and historical connections, there is a clear shift in the language in which the questions are put and answered. It is that shift in the vocabulary of the debate that will concern me here. My concern in examining this shift, I should state at the outset, is not to endorse either the Augustinian questions or answers. I would not want to do either. It is rather to show how the shift in language gives undeserved plausibility to contentious explanatory and normative claims in social choice theory.

What is the nature of these changes in language? Some of the changes involved occurred only in this century. The concept of 'utility' in economic theory has over the last hundred years shifted from a substantive to formal interpretation. In the last century and early in this, the term "utility" carried a commitment to a substantive account of welfare, defined in terms of subjective states. In Jevons utility is defined in hedonistic terms (Jevons 1970, 91 and *passim*), and in Marshall and Pigou it still refers to some psychological state, an 'affection of the mind': the strength of preferences for a good, measured by a person's willingness to pay for their satisfaction at the margin, is an indirect measure of subjective states (Marshall 1962, III.iii; 1; Pigou 1920, 24). In the modern textbook the relation between preferences and subjective states disappears.⁶ The term "utility" is defined in terms of a preference ordering. The account is purely formal in the sense that it says nothing about the content of preferences only their structure. It makes no assumptions about the constituents of human wellbeing. On this account of utility, there are no grounds for any assumptions about the egoism or altruism of agents, or their proclivity to avarice or generosity. That individuals are concerned to satisfy a consistent set of preferences under budget constraints does not imply that agents are egoists in any strong sense of the term. It all depends what preferences they have. As Hahn and Hollis put, the definition of utility rules out neither Ghengis Kahn nor the saint (Hahn/Hollis 1979, 4).

Much economics does assume, however, that individuals are 'self-interested' or 'egoistic'. The notion of individuals maximising their utility retains its Benthamite self-seeking sense. However, of significance here is a second and earlier shift in terms of the debate between Augustinian questions and their modern counterparts, namely the language of egoism and self-interest itself. An under-discussed feature of discussions of egoism and altruism is very recent origin in

⁶ I discuss the shift in detail in O'Neill 1998b, ch. 3.

European languages of many of the concepts we use for self-interested dispositions and behaviour (Price 1988). The term "altruism" was coined by Comte in the mid-nineteenth century. The term "egoism" along with many other concepts to describe self-seeking dispositions emerged in the mid-eighteenth. While the concepts of private as against public 'interests' is older, the concept of 'self-interest' in its modern usage to describe economic advantage is likewise one that emerges in the seventeenth and eighteenth centuries (Hirschman 1977, especially 31–42). Now clearly, the late arrival of the terms does not necessarily indicate the late arrival of the dispositions they attempt to capture: much of the behaviour that we now call 'egoistic' and 'self-interested' has a much longer history. However, what does change are the terms of description. There has been a shift away from the use of specific virtue and vice terms such as avarice, ambition, pride, envy, vanity, vainglory, conceit and the like in which the political economy of the seventeenth and eighteenth century is still often framed, to a more sober language of 'interests', 'preferences' and 'utility' in which the language of economics and rational choice in the nineteenth and twentieth is almost entirely framed.

Eighteenth century economics lies in the cusp of the two languages. Much eighteenth century economic writing is still largely written in the language of vice and virtue. In Mandeville's work for example it is the vices such as avarice, prodigality, pride, envy, vanity, folly and fickleness which have unintended beneficial consequences (Mandeville 1988, volume I, 25 and 100ff). Mandeville directly inherits the Augustinian themes of the Jansenists that fallen men without charity and driven only by self-love can live together in peace. The claim that commerce is a key institution through which peace is possible is already to be found in Jansenist texts. It is evident in Nicole's restatement of the Augustinian claim that opens this section of the paper. However it is not only in writers who write directly under Augustinian influence like Mandeville that the language of virtue and vice persists. It still pervades the work of the classical economists such as Hume and Smith, who were critical of Mandeville.⁷ In the work of Hume and Smith the language of the interests and the languages of passions is used interchangeably. Consider the characteristics that Hume takes to be universals of human nature: "Ambition, avarice, self-love, vanity, friendship, generosity, public spirit" (Hume 1975, VIII, 65). The term "self-interest" is used interchangeably with two different terms on this list—"avarice" and "self-love"—and each use carries a distinct sense.

Hume often uses self-interest interchangeably with avarice, the love of gain or avidity, "the interested affection", most notably in Book III of *A Treatise of Human Nature*. Thus understood, self-interest is central to Hume's account

⁷ For the criticisms see Smith, 1982 VII.ii.4; 'Of Refinements in the Arts', Hume 1985, 280.

of the conditions for commercial society. The passion of avarice, avidity, the love of gain is for Hume the foundation of commercial society: "it is more easy to account for the rise and progress of commerce in any kingdom than it is for that of learning ... Avarice, or the desire for gain, is a universal passion, which operates at all times, in all places, and upon all persons: But curiosity, or the love of knowledge, has a very limited influence ..." ('On the Rise and Progress of the Arts and Sciences', Hume 1985, 113) The problem of avarice for Hume is that it is both universal and destructive of society. "This avidity alone, of acquiring goods and possessions for ourselves and our nearest friends, is insatiable, perpetual, universal and directly destructive of society." (Hume 1978, III.ii.2) As such it must be controlled, but neither our benevolence for strangers, nor more importantly any of the other passions—vanity, pity, envy, revenge etc.—are able to restrain it, the former because it is too weak, the latter because they are not constant. The only constant passion that is able to restrain the love of gain is that passion itself: "There is no passion, therefore capable of controlling the interested affection, but the very affection itself, by an alteration of its direction ... [T]he passion of self-interest ... itself alone restrains it." (Hume 1978, III.ii.2) The self-restraint that the passion of self-interest exerts upon itself is the condition for the rules of justice and commerce. In this context "the passion of self-interest" is being used in a narrow sense to refer specifically to the love of gain in contrast with the other passions.

The contrast between the specific passion of self-interest and the other passions takes on explanatory significance in Hume's political, economic and historical writing. A central problem that Hume like other classical economic theorists was concerned to address was the specification of the conditions for the existence of commercial societies, both objective and subjective. In the *Treatise* Hume approaches the problem by way consideration of the institutional principles for the 'stabilisation' and 'transference' of property and the possibility of mutual promising through which the passion of self-interest is redirected to control itself (Hume 1978, III.ii.3–5). However, in other of Hume's writing it is the contrast between, the passion of self-interest and the others that comes to the fore as part of his sociology of the conditions for the emergence of commercial society. Consider Hume's remarks about the incompatibility of absolute monarchy and commerce:

"Commerce, therefore, in my opinion, is apt to decay in absolute governments, not because it is there less secure, but because it is less honourable. A subordination of ranks is absolutely necessary to the support of monarchy. Birth, titles, and place, must be honoured above industry and riches. And while these notions prevail, all the considerable traders will be tempted to throw up their commerce, in order to purchase some of those employments,

to which privileges and honours are annexed." ('Of Civil Liberty', Hume 1985, 93)

In traditional societies, 'birth, titles and place' are defined as the object of one's interest: to sacrifice them for money would be a sign of vulgarity. At the limit the prince, driven by the desire for glory, could throw away all goods for its sake. The theme runs through Hume's histories. Typical is his censure of Richard I: "Elated with hopes of fame, which in that age attended no wars but those against the infidels, he was blind to every other consideration; and when some of his wiser ministers objected to this dissipation of the revenue and power of the crown, he replied, that he would sell London itself could he find a purchaser." (Hume 1983, volume I, 380) In conditions of absolute monarchy, the passion for fame could subdue the passion of self-interest. Successful commercial societies require not just the redirection of the 'interested affection', but also conditions in which it is not subdued by countervailing passions.⁸

Hume uses the term "self-interest" not only to refer to the specific passion of avarice, but also in a more general sense interchangeably with the concept of self-love.⁹ The use is common in the eighteenth century. Butler for example uses the phrase "self-love and interestedness" understood "to be an affection to ourselves, a regard to our private good" (Butler 1983, IV.11). The concept of self-love is the one single older term that might be thought to do some of the work of the more recent term egoism. Indeed it is not uncommon to find commentators using the term "egoism" in discussions of classical texts that employ equivalents of the English term "self-love" and employing the distinction between egoism and altruism to describe the contrast drawn between self-love and benevolence. What is meant by "self-love"? In both classical philosophical and biblical texts the concept of self-love is used in both a narrow negative sense and a wide non-negative sense. Thus Aristotle claims that while it can be to describe a narrow pursuit of "the biggest share of money, honours and bodily pleasures", and is commonly thus understood, proper self-love incorporates wider attachments: the virtuous are proper self-lovers (Aristotle 1985, IX). In Biblical texts, the concept is sometimes likewise used in a narrow negative sense to describe various vices of character—covetousness, pride, an absence of filial obedience etc. (*Timothy* 2, 3)—and sometimes in a neutral or positive sense, most notably in the commandment "love thy neighbour as thyself" (*Leviticus*, XIX, 18; *Luke* X, 27; *Matthew*, XIX, 19; XXII, 39). The distinction between improper self-love in the negative sense and true self-love runs through Augustine's texts.¹⁰ The negative use is that which is

⁸ See also for example Hume's comments on the effect of commerce on making "the love of gain prevail over the love of pleasure" ('Of Interest', Hume 1985, 301).

⁹ See for example Hume 1975, 174, 178.

¹⁰ For a detailed discussion see O'Donovan, 1980. See also Aquinas: "Love of self is

taken up in the Augustinian writings of Jansenists like Nicole, where self-love (*l'amour-propre*) in a narrow sense is taken to underlie commercial society.

In eighteenth century writers the distinction between the two forms of self-love is sometimes marked by the contrast between 'immoderate' or 'excessive' self-love and 'reasonable' self-love (Butler 1983, II.9). The distinction is central for example to Butler's influential arguments for the compatibility of self-love and benevolence. A dispassionate and reasonable understanding of our interests requires that we should not be over-concerned with ourselves: "*Disengagement* is absolutely necessary to enjoyment; and a person may have so steady and fixed an eye upon his own interest, whatever he places it in, as may hinder him from *attending* to many gratifications within his reach, which others have their minds free and open to." (Butler 1983, IV.8) Thus 'immoderate self-love', love of self to the exclusion of other affections, is contrary to a persons private good. "Immoderate self love does very ill consult its own interests; and how much soever a paradox it should appear, it is certainly true that even from self-love we should endeavour to get over all inordinate regard to and consideration of ourselves." (Butler 1983, IV.8) The paradox allows Butler to conclude that the call of benevolence, acting for the sake of others, is not contrary to self-love: "there is no peculiar contrariety between self-love and benevolence, no greater competition between these than between any other particular affections and self-love." (Butler 1983, IV.12) Implicit in the last remark is Butler's more general point that self-love is a second-order affection that presupposes more specific passions and interests to give it content: "The very idea of an interested pursuit necessarily presupposes particular passions or appetites; since the very idea of interest or happiness consists in this, that an appetite or affection enjoys its object. It is not because we love ourselves that we find delight in such and such objects, but because we have particular affections towards them. Take away these affections, and you leave self-love absolutely nothing at all to employ itself about ..." (Butler 1983, Preface.37)

The uses of "self-love" in both narrower and broader senses are also acknowledged by Hume. Thus he notes, with criticism, its narrow and pejorative use: "Avarice, ambition, vanity and all passions vulgarly, though improperly, comprised under the denomination of *self-love* ..." (Hume 1975, 221) In contrast to that account, Hume echoes Butler's claim that self-love in its proper sense is a second-order affection that is parasitic on first-order passions. Thus, in the course of his rejection of the claim that all human action is motivated by self-love in the second appendix to the *Enquiries* Hume writes:

"If I have no vanity, I take no delight in praise; if I be void of ambition, power gives me no enjoyment: if I be not angry, the

common to all in one way; in another way it is proper to the good; in a third way it is proper to the wicked." (Aquinas 1952, II.II.25.7)

punishment of an adversary is totally indifferent to me. In all these cases there is a passion which points immediately to the object, and constitutes it our good or happiness; as there are other secondary passions which afterwards arise and pursue it as part of our happiness, when once it is constituted such by our original affections. Were there no appetite of any kind antecedent to self-love, that propensity could scarcely ever exert itself ..." (Hume 1975, 253)

Since, self-love in this general sense is understood to be a secondary affection that is derivative upon the distinct first order passions, first order affections for other persons cannot be said to be itself derivative on the second order passion of self-love: "Now where is the difficulty in conceiving, that this may likewise be the case with benevolence and friendship, and that, from the original frame of our temper, we may feel a desire for another's happiness or good, which by means of that affection, becomes our own good, and is afterwards pursued from the combined motives of benevolence and self-enjoyment?" (Hume 1975, 254)

Hume's work stands prior to a shift in the language of political economy from a language that employs specific virtue and vice terms, such as avarice, ambition, vanity, along with the more general concept of self-love, to a language exclusively of self-interest, egoism and the maximisation of expected utility in which virtue and vice terms largely disappear. Is there anything of significance in this shift in language? Does it matter? One possible answer is that it does not. Price (1988), for example, argues that it is simply a paradox that the recent terms which describe universal self-seeking dispositions have arrived so late in history. Had the terms had a longer history earlier theorists such as Machiavelli would have agreed with the assertion that human action had an egoistic or selfish character. This response is not I think entirely satisfactory. Whether or not the shift in language does mark a change in dispositions I leave aside here. Even if there is only a shift in language, not in real dispositions, the shift matters. Whatever other advantages the more austere language of modern economics might have, the shift involves a loss of distinctions and sensitivity to motivational shifts across different institutions. In doing so it gives a spurious plausibility to the assumption of constancy of motivations across institutions.

It is notable in this regard that Hume makes a poor starting point for claims about constancy. The claim that Hume offers grounds for the claim that assumptions about the behaviour of agents in markets can be generalised elsewhere, including politics, is false. While Hume did hold that certain dispositions were constant and universal features of the human condition, including avarice or the love of gain, these universal characteristics are plural and different institutional contexts both foster and rely upon distinct human passions.

Hume's political economy indeed illustrates how the universal principles and the particularities of culture are supposed to meet. Because he is concerned to specify the motivational conditions for the existence of commercial society, he is less inclined than modern neo-classical economists to assume that the motivations exhibited in commercial society can be extended to explain all behaviour. It is only where riches, rather than honour become objects of esteem that commercial institutions can flourish. Interestingly, if Hume here anticipates any recent economic theory, it is not public choice theory but rather the institutional economics of Polanyi: "The outstanding discovery of recent historical and anthropological research is that man's economy, as a rule, is submerged in his social relationships. He does not act so as to safeguard his individual interest in the possession of material goods; he acts so as to safeguard his social standing, his social claims, his social assets. He values material goods only in so far as they serve this end." (Polanyi 1957, 46) The emergence of market economies requires that money-making is a source of social standing.¹¹ The point echoes Hume's claim that while the original motive for riches may lie in the direct pleasures they might bring "the possessor has also a secondary satisfaction in riches arising from the love and esteem he acquires by them ... This secondary satisfaction or vanity becomes one of the principal recommendations of riches, and is the chief reason, why we either desire them for ourselves, or esteem them in others." (Hume 1978, II.ii.5)

Hume's account of the motivational structures of individuals in different commercial and non-commercial societies is then not that of the public choice theorist. Indeed, nowhere is his departure from public choice clearer than in the very passages about knavery in politics that are taken to anticipate public choice themes. The point about knavery in politics is simply misread if it is understood as an early statement of the public choice principles. Indeed it is almost the opposite. Where the public choice theorist sees the problem as one of consistently applying a picture of a self-seeking individual to the public domain, against those who would see the public domain as one of benign actors, Hume's problem is the reverse: why it is that people who are honourable in pursuit of their private life cease to be so in the public domain. His point is one that applies solely to the political and concerns the behaviour of men when they act in parties, such that the countervailing check of personal honour is absent: "men are generally more honest in their private than their public capacity, and will go greater lengths to serve a party, than when their own private interest is alone concerned. Honour is a great check upon mankind: But where a considerable body of men act together, this check is, in a great measure, removed ..." ('Of the Independency of Parliament', Hume

¹¹ Polanyi 1957, 68ff. At the same time, Polanyi is mistaken when he claims that this observation is inconsistent with the views of Adam Smith (Ibid. 43–45). Smith, even more than Hume, argues that social standing, vanity, is the primary source for the acquisition of wealth (Smith 1982, IV.1.8; 1981, III.iv.17).

1985, 43) Hume's claim about knavery depends upon variation in motivational structures in different institutional settings, not their constancy.

My point here is not that all of Hume's claims in political economy and theory are true. I believe many are false. For example, his claims about the universality of avarice are at best exaggerated. My point is rather that Hume's work shows a sensitivity to the relationships between institutional setting and motivation, and that this sensitivity owes something to the relative transparency of the plural language of vices and virtues that he employs which is lost in the transformation into the language of self-interest, egoism and utility-maximisation. The former renders explicit distinctions that are fudged in the language of self-interest, egoism and altruism in which the discussion of rational choice theory is framed. Public choice arguments in particular commit fallacies of equivocation that translation into the older language makes transparent.

The claim that assumptions about self-interested motivations of agents in market can be generalised to all social spheres draws some of its plausibility from an equivocation about 'self-interest'. In explanations of the behaviour of actors in the market it is used in a narrow sense to refer to specific motivations that drive accumulation—the passion for gain or avarice. The assumption of self-interest in general equilibrium theory amounts to an assumption of non-satiation: given a choice between a smaller and larger bundle of commodities, the agent always chooses the larger (Arrow/Hahn 1971, 78; Haussman 1992). Even if this claim is true of market actors, it is far from clear that it can be generalised to other spheres. Indeed, once the utility functions assumed in different social settings are unpacked public choice accounts are far more traditional than is normally supposed. In the political domain, for example, the individual is taken to be driven by ambition, not avarice. Bureaucrats aim to maximise the size of their bureau budget, since that is correlated with their utility understood in terms of personal promotion and advancement (Niskanen 1973, 22–3). Politicians offer goods to different interest groups since they aim at power. In other spheres such as science, the utility function is defined in terms of the desire for recognition, and so on. As the institution changes so does the utility function that is invoked.¹² Reference to 'self-interest' or 'maximisation of expected utility' disguises the diversity of motivations and possible conflicts between them that are explicit when the language of avarice and ambition is employed. Vice and virtue terms describe dispositions of character to involve specific desires and feelings towards particular objects: ambition, for power; avarice, for material gain; vanity, for recognition or glory. As such the older terms like "avarice", "vanity", "ambition", and "pride" describe dispositions of character and are clearly associated with different institutional settings. Hence, they render clearer conflicts between

¹² See O'Neill 1994 for further discussion.

different motivations that are associated with institutional conflicts. Classical economic theory of the kind found in Hume is in consequence more sensitive than modern neo-classical theory to conflicts in motivation across institutions.

The plausibility of the public choice claim for the universality of self-interest as a motivation across different institutions draws much of its strength on an equivocation between the narrow sense of the term as it is used in explanations of market behaviour and a wider sense in which it appears in the basic axioms of neo-classical theory. It is possible for the defender of the claim to retreat to a general sense of 'self-interest' or 'the maximisation of expected utility' in which it is understood as the pursuit of preference-satisfaction, given some constraints of consistency on preferences. However, thus understood, the claim that the assumption of self-interest can be generalised to explain behaviour in different spheres may well be true but it begs the explanatory questions. Self-interest in this general sense has the same logical attributes that Butler and Hume note of 'self-love' understood in its broad sense. The concept refers to a second-order motivation that is parasitic on prior first-order motivations. Without some specification what persons take to be in their interests or what their preferences are the concept is empty: "Were there no appetite of any kind antecedent to self-love, that propensity could scarcely ever exert itself ..." Appeal to 'self-interest' without reference to the prior motivation will be explanatorily inert. Once these are specified appeal to self-interest begs the explanatory questions. In different institutional contexts an individual's conceptions of their interests will shift. In particular in non-market institutional spheres, interests and preferences will differ from those expressed in markets. Where markets are introduced into non-market spheres the consequence will often be a shift in persons' conceptions of their interest. The shift is evident not just in the transition from pre-commercial to commercial society, but also in the shifting boundaries of markets within modern market economies. For example, the interests of the scientist in institutions that follow the norms of Mertonian science will not be the same as those of a scientist in institutions governed by commercial norms: a self-interested act in the former—publishing a novel result in peer-reviewed journal—might be an altruistic act in the latter in which a patent rather than publication is the object of an individual's interests.¹³ Constancy of the kind that public choice theory assumes cannot be sustained.

¹³ For a more detailed discussion see O'Neill 1990; 1992; 1998b, chs. 11–12.

3. Can Choices be Inconsistent?

"Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them ... A passion ... contains not any representative quality, which renders it a copy of any other existence or modification ... 'Tis impossible, therefore, that this passion can be oppose'd by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, consider'd as copies, with those objects, which they represent." (Hume 1978, II.iii.3)

It might be objected with some justice that the arguments of the last section, while they might raise difficulties for public choice theory, do not touch the core of rational choice theory. The core of rational choice theory is not the idea of self-interest or egoism, but that of rational choice. The justification of rational choice typically appeals to a modified Humeanism. Where Hume allows preferences to be guided by reason in just two ways—in ascertaining the existence of the object of the preference and in instrumentally finding the means to preference satisfaction¹⁴—the rational choice theorist adds a third—that of consistency. The theory is built upon the notion of rational choice. Preferences, understood as dispositions to choose, must be consistent. One basic constraint of consistency is taken to be that of transitivity. Preferences must be transitive. If an individual prefers *a* to *b* and *b* to *c* then transitivity entails they prefer *a* to *c*. Since preferences are understood as dispositions to choose—preferences are revealed in choice—the constraint of consistency is taken normally to rule out intransitive cycles of choice: if an individual given a choice from a set $\{a, b\}$ would choose *a*, and given a set $\{b, c\}$ would choose *b*, then transitivity enjoins that given $\{c, a\}$ a rational agent would choose

¹⁴ Hence Hume's well-known continuation of the passage that opens this section: "What may at first occur on this head, is, that as nothing can be contrary to truth or reason, except what has a reference to it, and as the judgments of our understanding only have this reference, it must follow, that passions can be contrary to reason only so far as they are *accompany'd* with some judgment or opinion. According to this principle, which is so obvious and natural, 'tis only in two senses, that any affection can be call'd unreasonable. First, When a passion, such as hope or fear, grief or joy, despair or security, is founded on the supposition or the existence of objects, which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the design'd end, and deceive ourselves in our judgment of causes and effects. Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger ... 'Tis as little contrary to reason to prefer even my own acknowledge'd lesser good to my greater, and have a more ardent affection for the former than the latter ... In short, a passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment." (Hume 1978, II.iii.3)

a. The assumption thus stated has been the object of a number of putative counter-examples. To follow are some brief examples:¹⁵

1. Multi-criteria Choices

I am applying for three jobs *A*, *B*, and *C*, and I am comparing them across three value dimensions, *u*, the intrinsic interest of the job itself, *v*, the location of the job, and *w*, the friendliness of the colleagues. Suppose the following value-assignments hold:

	<i>u</i>	<i>v</i>	<i>w</i>
<i>A</i>	1st	3rd	2nd
<i>B</i>	2nd	1st	3rd
<i>C</i>	3rd	2nd	1st

If I apply the majority decision rule—given a choice, choose the job that is the best on most dimensions—choices will be intransitive. Offered a choice between *A* and *B* I would choose *A*; offered a choice between *B* and *C* I would choose *B*; offered a choice between *C* and *A* I would choose *C*.

2. Regret

Judy is offered a choice of birthday treats. Given a choice between going to the city (*C*) and a garden party (*GP*), she would choose the city—the garden party would be more fun but it would be disastrous if it was wet, and so she plumps for the visit to the city which is unaffected by the weather. Given the choice of a house party (*HP*) and going to the city, she chooses the house party. Neither will be affected by the weather and the party more fun. However, when given the choice of house party and garden party she chooses the garden party. While it may be disastrous if it rains, she knows she will deeply regret having chosen the house party if the sun is shining. She knows she just couldn't bear the possible regret and so given the third choice she decides to minimise expected regret and chooses the garden-party. Her choices form an intransitive cycle: *C* from {*C*, *GP*}; *HP* from {*C*, *HP*}; *GP* from {*GP*, *HP*}¹⁶

¹⁵ For useful overviews of examples see Bar-Hillel/Margalit 1988 and Anand 1993, ch.4.

¹⁶ The example is a variant of one by Tversky reported in Elster 1986, 107–8.

3. Context Dependence

You are a guest at someone's house. After a wonderful meal you wait for the fruit: given a choice between a small apple (*S*) and an orange (*O*), you would choose the orange; given the choice between an orange and a large apple (*L*) you would choose the large apple; but given the choice between a large apple and a small apple, constraints of politeness enter and you choose the smaller apple: *O* from $\{S, O\}$, *L* from $\{L, O\}$ and *S* from $\{S, L\}$.¹⁷

4. Epistemic Dependence

At a hotel breakfast you are offered a choice between porridge (*P*) and scrambled egg (*E*). While well-done scrambled eggs are your perfect start to the day, you choose porridge as you know from bitter experience just how soggy hotel scrambled eggs can be. The waiter takes your order, but then emerges and tells you that there is a special that day—salmon (*S*). Would you prefer salmon to porridge? You are about to choose salmon when the chef comes in embarrassment. Porridge is off—the kitchen is out. So it's a choice of salmon and scrambled eggs. You choose scrambled eggs. Having discovered that the chef can cook salmon, you trust his capacity to cook scrambled eggs. The menu of choices of itself conveys information which effects the choice made: *P*, from $\{P, E\}$; *S* from $\{S, P\}$; *E* from $\{S, E\}$.¹⁸

How damaging are such apparent counter-examples? In all these cases it is possible to save the assumption of transitivity of preferences by redescribing the object of choice in a way that relativises it to the set from which the choice is made: *a*-from-the-set- $\{a, b\}$ is different object of choice from *a*-from-the-set- $\{c, a\}$. If one relativises all descriptions of objects to the choice set from which they are made, any intransitive cycle of choice can be redescribed in ways in which transitivity is not violated (Anand 1993, 103ff). As a general strategy this has the problem of rendering the transitivity axiom trivially true and no longer of normative force: no choices could ever violate it. However in at least some of the cases above one can see independent rationale for why the objects of choice might be described differently as the set from which they are chosen changes. Consider the final three cases. In the second, *HP*-from- $\{GP, HP\}$, since it involves risk of the state of affairs of having chosen a party in the house regretting you didn't have it in the garden, does describe a different state of affairs from *HP*-from- $\{HP, C\}$ in which the additional feature of risk of regret is not present. In the third, in one case one prefers a-small-apple-and-being-polite to a large-apple-and-being-impolite, whereas in the other choices the additional feature of impoliteness does not feature: one prefers a

¹⁷ The example is a variant of one in Anand 1993, 67.

¹⁸ The example is based on Sen 1995.

big-apple-and-being-polite to an orange-and-being-polite, and an-orange-and-being-polite over a-small-apple-and-being-polite. Likewise in the fourth case to choose E -from- $\{S, E\}$ is to choose a scrambled egg with information that it is well cooked whereas to choose E -from- $\{P, E\}$ is to choose a scrambled egg without that information.

Redescription might save transitivity of preferences. It does not thereby save transitivity of choices. Descriptions of choice under the standard behaviourist interpretation of neo-classical economics are extensional, not intensional. If you choose an object under one description you have chosen it under every description. Descriptions of preferences are intensional. If you prefer an object under one description it does not follow that you prefer it under another. Insisting that an object or state of affairs is chosen under a particular description, while it may not rule out intransitive cycles of choice, can save the transitivity of preferences, since it is under one description rather than another that an object is preferred. However, the strategy of relativising descriptions to the set from which choice is made also creates a distance between choice and preference. Given the completeness axiom in rational choice theory, preferences can no longer be understood as dispositions to choose. Given the completeness axiom it has to be assumed that an agent has a preference between a -from-the-set- $\{a, b\}$ and a -from-the-set- $\{c, a\}$. However, there can be no disposition to choose between a -from-the-set- $\{a, b\}$ and a -from-the-set- $\{c, a\}$ for the choice is not possible.¹⁹

There is a more general point here about the very notion of consistent choices that the examples raise. What makes the apparent counter-examples to transitivity problematic is that there appears to be no failure of consistency in the reasoning that leads to the intransitive cycle of choices. So in what sense is there an inconsistency here at all? There is as Sen (1995) has noted something odd about the very idea that choices of themselves can be consistent or inconsistent. Acts of choosing themselves cannot be either consistent or inconsistent, since they are not the kinds of things that could be. They have no truth-value. However, this is true not just of choices but also of preferences if they are understood in a Humean fashion. On the Humean view to have a reason for an action is to have a belief and desire, where the desire provides the motivational push, the belief provides guidance to the appropriate object. Beliefs alone have no motivating force, the desires no representational content. For Hume passions or preferences cannot be unreasonable because they do not represent states of the world and hence cannot be true or false: "Tis impossible, therefore, that ... passion can be opposed by, or be contradictory to truth and reason." For the same reason, given a Humean view, neither can they be consistent or inconsistent. Only judgements or beliefs can be inconsistent,

¹⁹ Sugden 1992. See also Broome's discussion of non-practical preferences in Broome 1993.

not passions or preferences. Only if one takes expressions of preferences to be expressions of judgement, and hence statements with a truth-value can the constraint of consistency apply at all. However, that would be to reject the Humean psychology from which rational choice theory begins. Constraints of consistency on preference cannot be married with a Humean model of the agent (cf. Broome 1993).

The problem here is not simply that rational choice theory is not Humean, but that under standard interpretations it may not be coherent. It at least faces a *prima facie* dilemma. Consider the claims of the form: 'I prefer *a* to *b*', 'I prefer *b* to *c*' and 'I prefer *c* to *a*'. If those statements are given a non-cognitivist interpretation—they express attitudes, they do not express judgements—then the statements have no truth-value and cannot be inconsistent. The agent may express desires that conflict with each other in the sense that they cannot all be satisfied, and in certain contexts, say when presented with a Dutch book, they may lead to a dilemmas. However, there is no question of contradiction here. If on the other hand they are taken to be inconsistent, they need to be understood under a cognitivist interpretation. They express judgements say of the relative goodness or expectations of the relative goodness of different items under some set of covering values. Given a cognitivist interpretation of claims of relative goodness, then statements of this form do have a truth-value and can be inconsistent with each other. Given more contentious assumptions, e.g. that the relation 'better than' is logically transitive, then it may follow that intransitive preferences involve inconsistencies. However, once we are in the domain of truth and falsity, not only consistency between preferences, but also the other resources of rational argument can tell for or against preferences. The introduction of consistency is a bridgehead for a wider role for reason. These observations clearly raise large meta-ethical questions. However, whatever meta-ethical direction one follows, rational choice theory on standard interpretations is unsatisfactory. It has to allow either a lesser or greater role for reason than that of consistency.²⁰

Bibliography

- Anand, P. (1993), *Foundations of Rational Choice Under Risk*, Oxford
 Aquinas, T. (1952), *Summa Theologica*, Chicago
 Aristotle (1985), *Nicomachean Ethics*, T. Irwin trans. Indianapolis
 Arrow, K./F. Hahn (1971), *General Competitive Analysis*, Holden-Day
 Bar-Hillel, M./A. Margalit (1988), How Vicious are Cycles of Intransitive Choices?,
 in: *Theory and Decision* 24, 119–145

²⁰ My thanks to Jonathan Aldred, Alan Holland and Jimmy Lenman for conversations and comments.

- Becker, G. (1976), The Economic Approach to Human Behavior, in: *The Economic Approach to Human Behavior*, Chicago
- Broome, J. (1991), *Weighing Goods*, Oxford
- (1993), Can a Humean Be Moderate?, in: R. Frey/C. Morris (eds.), *Value, Welfare and Morality*, Cambridge
- Buchanan, J. (1962), *Calculus of Consent*, Ann Arbor
- (1972), Towards Analysis of Closed Behavioral Systems, in: J. Buchanan/R. Tollison (eds.), *Theory of Public Choice*, Ann Arbor
- Butler, J. (1983), *Five Sermons Preached at the Rolls Chapel*, in: S. Darwall (ed.), Indianapolis
- Deane, H. (1963), *The Political and Social Ideas of St. Augustine*, New York
- Dunleavy, P. (1991), *Democracy, Bureaucracy and Public Choice*, New York
- Elster, J. (1986), The Market and the Forum: Three Varieties of Political Theory, in: J. Elster/A. Hylland (eds.), *Foundations of Social Choice Theory*, Cambridge
- Hahn, F./M. Hollis (eds.) (1979), *Philosophy and Economic Theory*, Oxford
- Haussman, D. (1992), What are General Equilibrium Theories?, in: *Essays on Philosophy and Economic Methodology*, Cambridge
- Hirschleifer, J. (1985), The Expanding Domain of Economics, in: *American Economic Review* 75, 6, 53–68
- Hirschman, A. (1977), *The Passions and the Interests*, Princeton
- Hume, D. (1975), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, Oxford
- (1978), *A Treatise Of Human Nature*, Oxford
- (1983), *The History of England*, Indianapolis
- (1985), *Essays Moral, Political, and Literary*, Indianapolis
- Jevons, W. (1970), *The Theory of Political Economy*, Harmondsworth
- Mandeville, B. (1988), *The Fable of the Bees or Private Vice, Publick Benefits*, Indianapolis
- Marshall, A. (1962), *Principles of Economics*, Eighth Edition London
- Nicole, P. (1997), *Of Charity and Self-Love*, in: E. Hundert (ed.), *The Fable of the Bees and Other Writings*, Indianapolis
- Niskanen, W. (1973), *Bureaucracy: Servant or Master*, London
- O'Donovan, O. (1980), *The Problem of Self-Love in St. Augustine*, New Haven
- O'Neill, J. (1990), Property in Science and the Market, in: *The Monist* 73, 601–620
- (1992), Altruism, Egoism, and the Market, in: *Philosophical Forum* 23, 278–288
- (1994), Preferences, Virtues and Institutions, in: *Analyse & Kritik* 16, 202–216
- (1998a), Against Reductionist Explanations of Human Behaviour: Rational Choice and the Unified Social Science, in: *Proceedings of the Aristotelian Society, Supplementary Volume LXXII*, 173–188
- (1998b) *The Market: Ethics, Information and Politics*, London
- Pigou, A. (1920), *The Economics of Welfare*, London
- Polanyi, K. (1957), *The Great Transformation*, Boston
- Price, R. (1988), Self-Love, 'Egoism' and *Ambizione* in Machiavelli's Thought, in: *History of Political Thought* 9, 237–261
- Samuelson, P. (1937), A Note on Measurement of Utility, in: *Review of Economic Studies* 4, 155–161

- (1938), The Empirical Implications of Utility Analysis, in: *Econometrica* 5, 344–35
- Sen, A. (1995), Is the Idea of Purely Internal Consistency of Choice Bizarre?, in: J. Altham/R. Harrison (eds.), *World, Mind and Ethics*, Cambridge
- Smith, A. (1981), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Indianapolis
- (1982), *The Theory of Moral Sentiments*, Indianapolis
- Sugden, R. (1992), How People Choose, in: S. Hargreaves-Heap/M. Hollis/B. Lyons/R. Sugden/A. Weale (eds.), *The Theory of Choice*, Oxford
- Tullock, G. (1970), *Private Wants, Public Means*, New York