Why Rationality May Be a Consequence of Hume's Theory of Choice

Marc-Arthur Diaye, André Lapidus

HAL Id: hal-00343872

https://paris1.hal.science/hal-00343872

Submitted on 2 Dec 2008

Marc-Arthur Diaye[*]
André Lapidus[**]

Commenting on our paper published in the same issue of this journal, Robert Sugden argued that there is "some evidence […] in support of the reading of Hume's theory of choice as non-rational". Although the author's argument is supposed to support his view, already expressed in a previous contribution (Sugden 1991), and to challenge our own interpretation, we far from disagree with several important features of his analysis. But this last needs to be complemented, in such a way that rationality indeed appears to be a possible consequence of Hume's dynamics of passions.

1. Is it Humean to choose without preferring?

According to Robert Sugden, in the works of David Hume, rationality should be better apprehended in terms of choice functions rather than in terms of preferences. He hence refers to a choice function for which the rationality criterion involves "contraction" and "expansion" consistency conditions, rather than properties of the preference relation. However, this approach raises two objections, which rest on standard choice theory **(a)**, and on Hume's theory of desire and will **(b)**.

(a). It could be shown that, when the domain of choice F = {S: S $\subseteq$ X \ $\varnothing$} is abstract, Sugden's "contraction" and "expansion" consistency conditions, also known, following Amartya Sen, as the α and β properties, are equivalent to the weak congruence axiom (see Sen 1971: 314), so that the choice function C(S) defined over F is rationalisable by a binary relation of preference R defined over the set of choice X, such that $\forall$ S $\in$ F and $\forall$ x, y $\in$ S, x $\in$ C(S) $\Leftrightarrow$ xRy. Now, this conception fits in exactly with one of the three interpretations of rationality through which we have analysed Hume's theory of choice: rationality as *compatibility* between preferences and choice – which, of course, does not involve the other two interpretations, *completeness* or *transitivity* of the preferences. Whereas preferences are formally dismissed, the way rationality is approached thus amounts to reintegrating statements which concern them.

[*] University of Evry Val d'Essonne (*EPEE*) and *TEAM* - boulevard François Mitterrand - 91000 Evry - France. E-mail: marc-arthur.diaye@eco.univ-evry.fr.

[**] University of Paris I Panthéon-Sorbonne and *PHARE* - 106-112, boulevard de l'Hôpital - 75647 Paris Cedex 13 - France. E-mail: lapidus@univ-paris1.fr.

(b). R. Sugden doubts that Hume uses preferences "in the same way that modern theorists do", but he seems to have fewer doubts that desires in the *Treatise* can be interpreted as a choice function C(.) from the domain of choice F into the set of non-empty subsets of the set of choice X. Still, as a consequence of Hume considering *desire* and *will* the last step before the carrying out of action, this position seems difficult to maintain.

It is well-known that, according to Hume, desire is related to pleasure, not in the sense that the intensity of an *impression* of pleasure is reflected by the intensity of the desire for an object, but because the intensity of the *idea of this impression* (which Hume calls a "belief"; see, for example, *THN I*: 96[1]) is expressed in a desire. Desires hence appear at first to be evaluative rather than comparative. But leaving aside all other considerations, it is trivial to conclude that if desire provides an *evaluation* which is a transformation of the intensity of pleasure, it also allows a *comparison*, through a binary relation that we can acknowledge as a relation of preference. In our paper, we have thus represented desires by preference relations.

Still closer to action, the will ("the *internal impression we feel, and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind*"; *THN II*: 399) does not express anything other than the consequence of comparison that desires draw in a context of choice. We have then considered that preferences and choice respectively refer to desire and will, and that for each context S, choices C(S) are locally rationalisable by a local relation of preference R_S. Since it remains local (that is, the preference relations might be contradictory between two different local contexts of choice, S_1 and S_2), this result should of course not be overestimated: from a Humean point of view, it only means that when facing such context S, we always choose what we prefer within this context.

Admitting that a theory of choice may be derived from Hume's theory of action as a special case, both standard choice theory and developments from the *Treatise* or the *Dissertation* devoted to desire and will give sense to the idea that a Humean choice process is grounded on choice functions *and* on preference relations.

2. A Humean restatement of Sugden's rationality criterion

There is therefore no major objection to reformulating Sugden's question in his comment: does Hume's theory of choice imply the existence of a choice function rationalisable by a preference relation? Answering such a question first led Sugden to bring to the fore a methodological criterion: rationality is disconfirmed when its causal explanation is rejected as a justification – that is, when there is no possibility of rescuing the theory by individuating the objects more finely. This criterion is most interesting, but rather puzzling, for it might be argued that a Humean approach leaves no room for such a separation between "causal explanation" and "justification". Yet it might also be considered to

---

[1] References to David Hume's works are given as follows: *THN x = Treatise of Human Nature*, Book x; *ATHN = Abstract of a Treatise of Human Nature*; *EPM = Enquiry Concerning the Principles of Morals*; *DP = Dissertation on the Passions*.

be a non-Humean equivalent to a condition of non-revision of preferences in the dynamics of passions **(c)** as will be shown more formally **(d)**.

(c). Following Hume, since choice, as a special case of action, is determined by our passionate investment directed towards an object – a *desire* – concerning which we have a *belief* that it is able to cause the satisfaction of this desire, its justification (the couple desire-belief) is itself the causal explanation. On second thoughts, however, Sugden's methodology might be considered a non-Humean echo to the conditions for a revision of what Hume called "our general resolutions" (*EPM*: 239), in the context of a passionate process. Such a revision in our general resolutions is a possible consequence of the "double relation" (*THN II*: 282-4; *DP*: 144-5), of ideas, and of impressions (R. Sugden favours the relation of ideas; but this could be extended without difficulty) which constitutes, beside external perceptions, the endogenous engine of the Humean dynamics of passions.

This dynamics of passions concerns the transformation of a mental state, characterised by some desires and will, into another mental state, characterised by possibly different desires and will. We have recalled above why we have, on the one hand, linked desires and will, and on the other hand, preferences and choice. However, R. Sugden's doubt that Hume uses preferences in the same way as "modern theorists" is well-founded. But the difference with "modern theorists" does not lie in preferences as such, which should not be rejected from the analysis, but with the set over which they are defined. In keeping with Hume, objects cannot be considered, in our minds, pre-existent data over which a binary relation of preference is defined: rather, they are constructed by our desires (see, for instance, Hume's discussion in *THN II*: 414). Other objects could exist for other agents or for an external observer, but for the agent who does not desire them, they simply do not exist. Moreover, when an agent desires such object, bringing this object into existence also entails the *description* under which it is perceived – a description which might differ substantially from that which another agent, or an external observer, might have adopted. This question of description is crucial. What Sugden considers a case of non-rationality corresponds to a passionate process in which the agent's description of objects comes with a change in his or her general resolutions, that is, in his or her preferences. On the contrary, when the agent's description fits into his or her general resolutions, his or her preferences are not submitted to any revision, and his or her choice may be called "rational" according to Sugden's methodology.

(d). Let us hence consider the following two choices:

$\{x\} = C(S_1)$ over $S_1 = \{x, y\}$

$\{y\} = C(S_2)$ over $S_2 = \{x, y, z\}$

These choices clearly violate Sugden's contraction consistency condition (Sen's property α), so that no general preference relation over $X = \{x, y, z\}$ can rationalise the choice function. But, as Sugden acknowledges, rationality could be rescued by "individuating options more finely". A well-known example, drawn from Amartya Sen (1993: 501), might help make this point clearer. Suppose that x, y and z denote three slices of cake, the sizes of which vary in increasing order from x to y and from y to z. We may then encounter two types of situations:

1. Suppose I am well-mannered, and that I do not want to pick the largest slice, though I am very fond of this cake: this *justification* for my choices $C(S_1)$ and $C(S_2)$ meets the alleged *causal explanation* for my behaviour, and, according to Sugden's criterion, I am rational.

2. But instead, if I try to *justify* my choices by explaining that they are *caused* by my attempt to satisfy a mild appetite, or to obtain the largest slice, still according to Sugden's criterion, I am no longer rational.

Now, these situations can be translated in Humean terms as follows:

1'. I am so well-mannered that when the slices of cakes are offered to me, I do not see them under some alleged *natural* description, but under a description determined by my manners and by my hunger, which favours their rank, rather than their size: over S_1' I am faced with the biggest slice of cake (a) and with a smaller (b); over S_2', I have to choose between the biggest slice (a), a smaller (b), and the smallest (c). Under this description, my choices

$\{b\} = C (S_1')$ over $S_1' = \{a, b\}$

$\{b\} = C (S_2')$ over $S_2' = \{a, b, c\}$

are such that they express local preferences over S_1' and S_2' which can be considered to be restrictions over S_1' and S_2' of the same general preference relation R defined over X = $S_1' \cup S_2' = \{a, b, c\}$. These choices are then evidently rational, in the sense that the preference relation R rationalises the choice function C(.). In other terms, my "general resolutions" need not be changed and do not contradict any local preferences, precisely as is the case in our paper, when preferences are not submitted to revision.

2'. My appetite, though moderate, is the only guide for my desires. So that the slices of cakes appear to me under a description which favours their size: over S_1' (resp. over S_2') I have to choose between x and y (resp. x, y, and z) defined as above. But if $\{x\} = C (S_1')$ and $\{y\} = C (S_2')$, my general resolutions have to be changed, as in D-PROC, in our paper, when preferences are revised. For instance, facing S_2', I could say: "Well, after all, these slices are not so big and I am beginning to feel a little hungry. Suppose that S_1' is again offered to me, I now choose $\{y\}$, and evidently not $\{x\}$".

3. Rationality as a possible outcome of choice

Up until this point, our disagreement with Robert Sugden should not be overstated, and neither the elimination of preferences nor the introduction of a non-Humean methodological criterion of rationality seriously threaten the kind of lesson that, as economists, we can draw from Hume's theory of passions. Accordingly, his sturdy opposition, expressed in the very title of his paper – "rationality *may not be* a consequence…" – which aims at closing the path that we have cautiously opened in our own title – "rationality *may be* one consequence…" – should rest on other grounds. These last concern **(e)** the interpretation of Hume, and **(f)** the place of rationality in the theory of choice.

(e). Sugden's interpretation of Hume appears to follow a well established tradition, which favoured what was called "radical scepticism", and laid emphasis on the impossibility of granting rational foundations to causal inferences. On the contrary, our reading of Hume is closer to a revival of naturalistic interpretations, of which N. Kemp Smith (1941) was a most remarkable forerunner. This reading insists on Hume's project of a "science of man", explained at length in the introduction of the *Treatise*. Far from being purely academic, such a difference has repercussions on the way we understand the changes generated by Hume's double relation.

Applying his own criterion for testing rationality, Sugden only regards the possibility that the working of the double relation (or, why not, of external perceptions) transforms the agent's desire, hence making

choices non-rational, like in situation 2 above. But Hume never claimed that the engine of the dynamics of passion that he tried to explain would lead to a sort of mental chaos: preferences do not simply change along the passionate process, they are revised and adapted. We have therefore tried to grasp Hume's few intuitions concerning this transformation. And it is on the basis of these intuitions that we have shown in proposition 1 that the updating process displayed in D-PROC led at any step to a rational choice, in the sense of a compatibility between choices and general preferences – that is, which would fulfil Sugden's conditions of contraction and expansion. In the same way, we have drawn on Hume's suggestion of the existence of some kind of stabilization of the emotions so as to show (proposition 2) that D-PROC led to complete and transitive preferences.

Of course, Hume never explicitly assumed something like preordered preferences. But his depiction of a standard functioning of our mind brings about rationality as an outcome of this functioning.

(f). Obviously, Sugden's view of the place of rationality differs from ours. In several important papers previously published, he sees rationality as a property of some specific *situations*, which is not shared by other situations. For instance, referring to Sen's well-known sporting metaphor which illustrates condition α ("if the word champion in some game is a Pakistani, then he must also be the champion in Pakistan" (Sen 1970: 17)), he argued that in other situations, like in interactive sporting championships, minimal consistency requirements[1] no longer prevail (Sugden 1985: 169-71). A consequence is that, although rationality does not become meaningless, there is no evidence that it should keep its privileged position as a standard in the assessment of choice. In contrast, we view rationality as a property which may be displayed by our *minds* in the process of building our world. This clearly excludes any fetishistic approach to rationality. But since it is linked to some regularity in the functioning of our minds, it also constitutes a reference to appreciate any departure from this regularity.

One needs no more than an emphasis laid on the positive aspects of Hume's theory of passion, along with a conception of rationality as an expression of our mind to open the path to this modest statement: rationality *may be* a consequence of a Humean theory of choice.

References

David Hume, 1739-1740. *A Treatise of Human Nature : Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*, edited by L.A. Selby-Bigge [1888] and P.H. Nidditch [1978], Oxford: Clarendon Press, 1978.

--------- , 1751. *Enquiry Concerning the Principles of Morals*, in D. Hume, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by L.A. Selby-Bigge [1894, 1902] and P.H. Nidditch [1975], Oxford: Oxford University Press, 1975.

---

[1] This attitude is the occasion of a shift in the vocabulary in use, where "consistency" stands for "rationality" and "rationality" for "having good reasons". In the same way, see also Sen 1993.

-------- , 1757. *A Dissertation on the Passions*, edited by T.H. Green and T.H. Grose, in *The Philosophical Works of David Hume*, vol. II, [London: 1882], Darmstadt: Scientia Verlag Aalen, 1964.

Norman Kemp Smith, 1941. *The Philosophy of David Hume : A Critical Study of Its Origin and Central Doctrines*, London: Macmillan, 1966.

André Lapidus, 2000. La Rationalité du Choix passionnel: En Quête de l'Héritage de David Hume, *L'Année sociologique*, 50(1).

Amartya K. Sen, 1970. *Collective Choice and Social Welfare*, San Francisco: Holden Day

-------- , 1971. Choice Functions and Revealed Preference, *Review of Economic Studies*, 38(3), July.

-------- , 1993. Internal Consistency of Choice, *Econometrica*, 61(3), May.

Robert Sugden, 1985. Why be Consistent? A Critical Analysis of Consistency Requirements in Choice Theory, *Economica*, 52(206), May.

-------- , 1991. Rational Choice: A Survey of Contributions from Economics and Philosophy, *Economic Journal*, 101(407), July.