Does Optimization Imply Rationality?*

March 1998

Philippe MONGIN

THEMA, Centre National de la Recherche Scientifique et Université de Cergy-Pontoise,

33 boulevard du Port, F-95000 Cergy

Abstract. The relations between rationality and optimization have been widely discussed in the wake of Herbert Simon's work, with the common conclusion that the rationality concept does not imply the optimization principle. The paper is partly concerned with adding evidence for this view, but its main, more challenging objective is to question the converse implication from optimization to rationality, which is accepted even by bounded rationality theorists. We discuss three topics in succession: (1) rationally defensible cyclical choices, (2) the revealed preference theory of optimization, and (3) the infinite regress of optimization. We conclude that (1) and (2) provide evidence only for the weak thesis that rationality does not imply optimization. But (3) is seen to deliver a significant argument for the strong thesis that optimization does not imply rationality.

1. INTRODUCTION

1.1 Aims and strategy of this paper

The following model of rationality is pervasive in economics and widespread elsewhere in the social sciences. A rational person's preferences are represented by a real-valued objective function, and his or her choices are the values of the instrument variable that maximize this function over the set of available options. This modelling is justified, at least implicitly, by the normative claim that rational individuals always maximize their preferences within their feasibility constraints, or briefly put, that rationality implies optimization. Common though it has become, this view conflicts with the suggestion of ordinary language that rationality is a broad and ill-defined concept. Rational individuals, it is often said, are those who choose in a way that is "appropriate" to the conditions of their choice; or, following another suggestion, those who act "on good

* This essay has evolved from a translation of an earlier French paper. I am grateful to Ann Broome and Richard Bradley for having prepared this translation. I am also particularly grateful to John Broome, Dan Hausman's and Wlodek Rabinowicz, whose comments have influenced the present version.

reasons".[1] These familiar definitions suggest that the technical notion of optimization cannot simply follow from the ordinary concept of rationality.

But the ordinary concept does appear to be at least compatible with a maximizing interpretation. Making the best choice is one way of choosing "appropriately"; the optimality property of a solution is a "good reason" for choosing it. Accordingly, some writers have claimed that optimization provides a modelling of rational choice, albeit not the only one. In his collection *Models of Bounded Rationality* (1983) Herbert Simon broadly subscribes to this position, which I shall refer to as the *conciliatory account*. Simon documents mechanisms or methods of 'bounded rationality' which, as he understands them, are particular cases of rationality; the mechanisms or methods of optimization (relabelled by him 'absolute rationality') are another particular case. Although he has occasionally been misunderstood on this, Simon does not a priori discard the optimizing model of choice. Mongin (1984) follows the conciliatory account by emphasizing that there are two levels of analysis of the rationality concept: on the one hand, the generic level at which rational choice is loosely defined in terms of "appropriateness" (or possibly in terms of "good reasons") and, on the other, the specific level at which specific models (such as those of "bounded" and "absolute" rationality) are introduced and can be compared with each other. Each specific definition aims at expressing the generic notion of rationality better than its rivals, and it is the social scientist's task to assess these conflicting claims by paying attention to the particular circumstances of choice. Simon's contention that an absolute rationality strategy might turn out to be less rational than a 'satisficing' or bounded rationality strategy goes hand in hand with the following methodological stance: which model is the more relevant of the two entirely depends on the particular cognitive circumstances. Simon's position, and the conciliatory account more generally, are highly flexible. This account does have a polemical import nonetheless: it denies the received view among economists that rationality implies optimization. However, it does not deny the significance of optimization as a possible rendering of rationality.

This paper will add further evidence against the economists' received view (in 2.5 and 3 below) but the gist of the argument is to question the conciliatory account itself. I want to take issue with the seemingly unproblematic assertion that if rationality does not imply the optimization condition, at least the converse holds. Accordingly, I have collected here various arguments or constructions of rational choice theory which seem to suggest that under some relevant choice circumstances, to optimize is not rational. At least, this is the initial suggestion. Whether the analysis will eventually confirm it is very much the object of this paper.

The first topic is mostly borrowed from the psychology of decision and concerns intransitive choices. After several writers in the field, I argue that cyclical choices are normatively acceptable in certain choice

[1] Definitions of individual rationality in classical philosophy refer to reason (as a faculty) than to reasons (as motives), though Kant, for example, linked the two themes. The idea of rationality as appropriate choice, or instrumental rationality, which irrigates today's social sciences, has become influential more recently, in particular since Max Weber (see Weber, 1949). But most of the current notions of instrumental rationality, be they optimizing or not, can be traced back to Aristotle's admirable discussion of *proairesis* in the *Nichomachean Ethics*.

circumstances. But I will add that this argument does *not* have the the effect of excluding optimization from the area of rational choice, even for the range of circumstances it applies to (i.e., when the agent chooses between multidimensional objects). Hence, the first group of objections in this paper do not lead to answering its general question in the negative. By briefly reviewing, and dismissing, the famous *money pump* argument, I intend to show that, symmetrically, the defenders of optimization cannot draw much comfort from an normative analysis of cyclical choices.

The second topic is borrowed from pure preference theory, where optimization has been redefined in terms of properties of the agent's choices. Revealed preference theory, as it is has been called, is a methodologically contentious part of economics, but I do not aim at reviving the classic objections it has raised. Rather, taking its axioms at their face value, I classify them according to their normative strength and emphasize the cognitive difficulties of the choice operations to which optimization is equated. This discussion is primarily directed against the claim that rationality implies optimization, not the converse. Indeed, the suggestion was strongly made in some of the revealed preference literature, that the latter provides a *justification* to the optimization principle. Even if the discussion does not yet provide an argument against the rationality of optimization, it already points towards its central weakness - i.e., the unbalance between the internal costs of optimization and what it achieves in terms of the initial objective.

How to analyze this unbalance is the third and last topic of the paper. I mention the classic point in practical philosophy, which was revived by Ryle (1949), that any rational decision criterion implies an infinite regression of decisions. I will suggest that this difficulty is best seen as a consistency problem for the theorist (rather than the agent) and that it becomes most acute whenever rational choices are construed as optimizing ones. I have based my discussion on a particular model of a firm's decision, which is adapted from the more abstract framework in Mongin and Walliser (1987). Admittedly, the conclusions are sensitive to the underlying assumptions, but it does seem that - at long last - the infinite regression objection gives a precise content to the claim that under certain choice circumstances, it is *positively irrational* to optimize. Remarkably, this critical point follows from pursuing Herbert Simon's initial objection against the absolute rationality model that the latter does not take into account the internal costs of the decision-maker's optimization. But Simon stopped at the conciliatory account, even if his pathbreaking analysis had a much stronger potential.[2]

1.2 Some background distinctions

There are at least two ways in which 'optimization' can be discussed, i.e., as a normative or prescriptive principle applicable to the agent, or as a modelling principle intended for the theorist. Contrary to the more

[2] Related to the objective of this paper is a book by Michael Slote, *Beyond Optimizing*(1989). This work also attempts to go beyond Simon's conclusion by arguing that in some circumstances, it is irrational to optimize. However Slote's strategy to establish this point is different from mine; in particular, he does not investigate the infinite regression problem.

elegant phrasing, "Is it rational to optimize?", the title of this article equivocates between these two meanings. The ambiguity is part of its subject matter, as I hope the comparison between the three topics will make clear. As long as one highlights situations pathological for the agent's optimizing choices - such as cyclical choices -, or even when one exposes the heavy cognitive requirements they impose on him - as in the reinterpretation below of revealed preference theory -, one lays oneself open to the following defensive move. (It is implicit in much of theoretical economics.) It consists in replacing the *agent-relative* normative interpretation by an *theorist-relative* modelling principle such as: 'to each act that seems intuitively rational, it is possible and desirable to attach an optimizing description'. The strength of the infinite regression objection is that, if properly formulated, it also hits this seemingly cautious reformulation of optimization. More generally, I do not think that a successful critique of optimization can be put forward if it just addresses the normative issues involved in the agent-relative version, and none of the methodological issues involved in the theorist-relative one.

The initial notion of optimization is agent-relative, and perhaps best given in terms of the following basic microeconomics theorem. If an agent's preferences can be represented as a weak ordering, that is, a reflexive, transitive and complete binary relation, and this ordering is continuous in a technically appropriate sense, then his preferences can also be represented by a continuous utility function (Debreu, 1959; for a good treatment see also Malinvaud, 1971, pp. 18-20). The theorem leads to the conclusion that on compact sets of alternatives it is possible to maximize the numerical representation and, hence, the agent's preferences.

The result does not indicate that the agent is *effectively* a maximizer. The supplementary statement that he is - a statement which is more informal than the preceding one - is generally taken for granted among economists. When they write that an individual is 'endowed' with a utility function that can be maximized on the domain of choice, they typically also imply that, at the moment of choice, this individual will select one of the maximizing values of the 'instrument' variable. (Economists are often worried about which optimal solution is selected if there are more than one, but this is again a different issue.) By and large, the only problem for optimization that is recognized by economists concerns the existence of a numerical representation, not its use by the agent. This is unduly restrictive. The statement that the agent is effectively a maximizer when maximization is possible is not an analytical statement.[3] Symmetrically, it would be unduly restrictive to assess optimization only in terms of effective maximization, while taking for granted the conditions stated by

[3] Among methodologists a somewhat analogous problem has been discussed in connection with the "rationality principle". Popper defines the principle as follows: "when we have built our model, our situation, we suppose only one thing, namely that the actors act within the framework of the theory, where they "infer the consequences" from what is implicit in the situation (1967, p. 144). This deliberately trivializing formulation pushes the difficulties of the principle back onto the situation relevant to the choice of the agent. One may then ask whether a principle of rationality that is apparently so empty of substance is not *analytic*. This implication is usually rejected by Popper's followers (e.g., Watkins, 1970). They claim that the principle is *synthetic* - while this is is difficult to ascertain since it is very rarely violated in reality. This discussion assumes that the "situation" is given to the agent in the same way, roughly, as economists assume that a utility function is given to him. In both cases it is a supplementary principle - the agent "infers the consequences" of the "situation", the agent maximises the function attributed to him - which makes the principle (respectively: of rationality, of optimization) operational. The supplementary principle seems both logically indispensable and less problematic than the initial assumption (of a given situation or utility function).

the existence theorem. There are two sides the optimization coin. One group of arguments in the paper are clearly on one side of the distinction between the conditions of maximization and effective maximization: intransitivite choices call into question the crucial transitivity condition. But the other arguments are ambiguous, and relate to the two sides.

Among the conditions stated in the microeconomics theorem that guarantees existence, I will immediately dispose of continuity. It does raise intriguing questions, but at least in the theory of choice under certainty, to which I limit myself here, continuity is very generally seen as a mere technical requirement.[4] Evidence for this is provided by the textbook discussions of the lexicographic model of choice, whereby the agent maximizes a discontinuous index. Standard texts in microeconomics (e.g., Malinvaud, l971, pp. 18-20) do not suggest that it is irrational to maximize a lexicographic ordering instead of a scalar function. The implicit understanding is that consumer theory uses the latter kind of representation for convenience, not for substantial reasons. There are occasional dissenters, however. Harsanyi (1976, p. 93) - though not in so many words - suggests that ordering partial criteria hierarchically is only one stage towards defining the optimizing rationality model. In his opinion as I reconstruct it, this model should not count as fully formed until it allows for smooth trade-offs between the different dimensions of utility - that is to say, unless the agent is endowed with a continuous utility function. This view is interesting but to discuss it any further would take us away from the main point. So I will endorse the position that continuity is part of the *microeconomic* modelling, but not of the general notion of optimization,[5] and henceforth deal nearly exclusively with the remaining two conditions on the preference relation.

1. 3 A warning on method

The intuitive notions of individual rationality which are to be confronted with more technical conditions are not susceptible to preliminary definition except for the very general statements mentioned in 1.1. This difficulty is of a familiar kind, and like others in rational choice theory, I will turn it around, without disposing of it entirely. The point is to reach a reflective equilibrium, i.e., to progressively adjust abstract conceptions and examples to each other. This method works in a relatively simple way on the examples of intransitive choices. Here, it is just a question of confronting the, as yet insufficiently determined, notion of rationality with particular circumstances in which the meanings of this word can be more precisely appreciated. When one compares optimization with other formal conditions as in the second and third topics, the reflective equilibrium method interacts with the no less classical analysis of concepts in terms of necessary and sufficient conditions. At this stage it is not only a question of balancing partial examples against a vague theory, but also of evaluating the normative plausibility of a choice method in terms of equivalent (or at least necessary) conditions

[4] Expected utility theory has a continuity axiom which, by contrast, is sometimes seen as significantly loaded. For an early discussion, see Marschak (1950).
[5] This position is not unlike Becker's, who distinguishes between rationality in general and microeconomic rationality.

that are somehow more interpretable.

2. Reasonable Intransitivities of Choice

The psychology of decision and decision theory literatures abound with examples of cyclical choices. Some of theme appear to be rationally defensible, a point well emphasized for instance in Anand's (1987) and Bar-Hillel and Margalit's (1988) surveys. If one accepts the view that choices unproblematically reflect preferences, it should follow that in the circumstances spelled out by these examples, the agent's preference is both rational and incompatible with optimization (since it violates transitivity). From a potentially large collection, I have selected three cases which make it already possible to draw some conclusions.

2.1. The choice of spouse experiment

I start with the classic study by May (1954) on choosing a spouse. The experimenter classified the objects of choice in three dimensions - intelligence, beauty, fortune - each measured on a distinct qualitative scale. The subjects were faced with successive binary choices. May obtained 62 replies, of which 17 demonstrated a Condorcet cycle and 21 amounted to applying a lexicographic rule. The other 24 were based on coherent trade-offs between the three dimensions. Only the latter conformed to the microeconomic theory of optimization, in the sense defined above. One of the first of its kind, the experiment appears crude by contemporary standards but remains instructive nevertheless. Besides establishing that lexicographic choice behaviour is widespread, it showed that the 17 cyclical subjects were coherent in one respect - i.e., they had consistently applied the majority rule to the three dimensions. This suggests justifying cyclical choices by analogy with what can be said for deliberative assemblies. The cycles are arguably the consequence (inevitable and even possibly assumed to be so by the agent himself) of adopting a decision rule which conforms to otherwise impeccable properties such as anonymity, neutrality, etc. The analogy is quite simple to defend in the context of the particular study, because May had imposed the splitting up of the object into intelligence, beauty, and fortune. However, May's 'spouse' might be nothing but an artifact of the questionnaire. The study does not contain any evidence to reject this suggestion.

This is a common problem with 'reasonable' intransitivities. All seemingly convincing examples I know involve objects of choice that are identified by the experimenter with vectors of characteristics. Individual choices are then often rationalized as in May, i.e., by analogy with social choices. Characteristics play the role of individuals, and their aggregation obeys the alledgedly compelling, but logically overdetermined constraints which lead to Arrow's impossibility theorem, or some related impossibility result. However, the obvious difference with social choice is that the individual is observable and relevant in a way characteristics are not. Their role in explaining choices is a significant empirical hypothesis. At the normative level too, characteristics raise a problem. If the splitting up of the object of choice, as understood by the observer, is not part of the

agent's preference judgments, it is unclear why it should be relevant to the rationality or otherwise of his choices. This objection will have to be addressed also vis-à-vis the next two examples, which similarly involve choosing between multidimensional objects.

2.2 Intransitive Indifference

Contemporary decision theorists have widely come to recognize that cyclical choices might result from perceptual effects (e.g., Fishburn, 1970a). Suppose that the agent only notices temperature differences of 3 degrees; he will then identify 17°c. with 19°c., 19°c. with 21°c, but not 17°c. and 21°c.. If he is asked to choose between three baths having temperature 17°c., 19°c., and 21°c., he will express indifference between successive baths, and strict preference among the extreme ones, so that, clearly, his choices cannot reflect a transitive preference. On the other hand, they cannot be deemed irrational.

Impressed by this simple fact, a number of decision theorists regard as normatively compelling the transitivity of only the strict preference relation. A weak preference relation that satisfies the latter property, but not necessarily full transitivity, is called quasi-transitive. This weakening is no doubt a departure from the standard economist's conception, but it preserves a sufficiently clear notion of optimality; for a 'maximal element' to exist on a finite choice set only the acyclicity of strict preference is needed, and this is an even weaker property than quasi-transitivity.[6]

Luce's semi-ordering (1956) constitutes a less radical departure from orthodoxy than quasi-transitivity. (While remaining compatible with intransitive indifference, a semi-ordering has implications that a quasi-transitive ordering lacks. For example: if an individual is indifferent between a and b and strictly prefers b to c, he does not strictly prefer c to a..) Starting with Luce's semi-ordering, Tversky (1969) has devised the lexicographic semi-ordering , which contrary to the latter, does not necessarily satisfy acyclicity. Essentially, it is a lexicographic preference relation such that the first dimension satisfies the semi-ordering axioms, and can thus possibly satisfy intransitive indifference; the dimensions after the first might, if so desired, obey the ordering axioms. The following example shows a lexicographic semi-ordering at work. An employer favours intelligence over years of experience, but to measure this intangible quality he has to make do with a scale that permits 3 unit errors. So, the strict preference of the employer will produce cycles like:

$$(115,7) > (117,3) > (120,0) > (115,7).$$

The nicety of Tversky's concept is that it leads to cyclical choices by combining two ingredients neither of which, when taken in isolation, would conflict with acyclicity. Since intransitive indifference cannot be said to be irrational, the conclusion that the employer's choices are not irrational follows as soon as one grants that lexicographic preferences does not invove any irrationality per se - an common step to take as we mentioned in

[6] By a 'maximal element' I mean one to which no element in the choice set is strictly preferred. A preference relation is said to be acyclic if, when an individual strictly prefers a_1 to a_2, …, a_{i-1} to a_i, …, a_{n-1} to a_n, he does not strictly prefer a_n to a_1, irrespective of the length n of the chain of strict preferences. See, e.g., Sen (1970).

1.2 .

Like May's, this example is open to the charge that the employer's preferences are just assumed to be structured in terms of the given dimensions. Notice, however, the following difference. May started from observed choices and offered to rationalize them in terms of a normatively defensible non-transitive preference - an a posteriori reasoning. Tversky started from a normatively defensible non-transitive preference and showed that it sometimes entailed cyclical choices - an a priori reasoning. The next example to come is also of the a priori type.

2.3 The horserace example

In Blyth's (1972) horserace example, the subject is asked to choose among successive pairs of bets on horses A, B and C, respectively. The finishing position of each horse depends on whether the going is hard, soft or heavy, and this in turn depends on an unknown state of nature. If it rains (= S_1), horse A beats B and C, and B beats C; if it does not rain, but the weather is wet (= S_2), horse B beats C and A, and C beats A; if the weather is dry (=S_3), horse C beats A and B, and A beats B. Accordingly, when S_1, it is better to bet on A than B, and on B than C; when S_2, it is better to bet on B than C, and on C than A; when S_3, it is better to bet on C than A, and on A than B . Blyth claims that the ex ante preferences resulting from these data should conform to the rule:

"the agent prefers to bet on X than Y iff the probability that X beats Y is greater than 1/2."

But this apparently plausible rule leads to cyclical choices, as is easily verified in the particular case where $p(S_1)= p(S_2)= p(S_3)= 1/3$. The agent then bets on B against C, on C against A, and on A against B.

When probabilities are equal, Blyth's bettor in effect applies the simple majority rule three times, which takes us back nicely to the choice of spouse experiment. While it shares with it this curious analogy, Blyth's example seems to be more convincing than May's at least in the following respect: the multidimensional structuring of the object is now unproblematic. The distinction between states of nature is 'objective' in a sense that the distinction between psychological dimensions was not. Commenting on Blyth, Bar-Hillel and Margalit go so far as to write that in this case, contrary to May's, "the cyclicity of choice is in the external world" (1988, p. 132). I agree with their essential point although I do object to their formulation, which might obscure the role of psychological assumptions in the case at hand. (The assumption that the agent bases his choices on considering the objective states is easy enough to accept, but it is a psychological assumption nonetheless.)

Thus, we seem to have at last come across a non-artificial problem for optimization. This is what Blyth would like his reader to conclude. To assess his contention,[7] I think it is crucial to distinguish between two classes of decision problems. When faced with a race between two horses (problem 1), Blyth's rule of decision says, very plausibly, that the individual should bet on the horse with the higher chance of finishing the first. Each of the three two-horse contests will elicit a different answer from the individual, hence a cycle of bets, but no paradox:

[7] In this paragraph I am indebted to a conversation with John Broome.

there is nothing strange in the fact that different decision problems lead to different choices. Now, one should resist the temptation of dealing with choices in distinct two-horse races as if they were pairwise choices in a three-horse contest (problem 2). When the three horses are running together, to bet on A rather than B means something different than in the other case - to wit, it means, to bet that A, rather than B, will finish first of A, B and C. Although the decision problem is not the same, Blyth's rule recommends again that one should bet on X rather than Y if X has the greater probability of finishing the first of the two. This is absurd. One should now, of course, bet on X rather than Y if X has the greater probability of finishing the first of the three. When the probabilities are equal, this leads to declare oneself indifferent between betting on B or C, on C or A, and on A or B.

A rational individual always applies the same rule, which is to select the horse with the highest probability of winning, where the meaning of 'winning' is fixed by the nature of the decision problem (the contest). I suspect that Blyth has insisted on his curious rule because he has overlooked the difference between two kinds of decision problems, but this is just an interpretation. At any rate, most commentators have violently disagreed with his rule.[8] The seemingly striking counterexample to optimization has failed its promise.

2.4. Assessment of the three examples

The failure of Blyth's paradox sheds some light on May's. The latter example reveals cyclical bevahiour in precisely those circumstances - pairwise choices among at least three objects given at a time - in which I just said a rational horse-race bettor should not exhibit a cycle of strict preferences. (The rational horse-race bettor's choices might cycle but this will happpen only when he is indifferent between the three bets.) What is then the conceptual point underlying this difference? As opposed to the spouse example, the horse-race example assumes that there are given numerical exchange rates (in the particular instance, probabilities) to the dimensions. So the difference between the two boils down to this: to follow Blyth's rule would be to discard existing numerical information on the characteristics, while to choose spouses cyclically reflects the agent's failure to generate this information - a failure which may be explicable in the circumstances. If I believe that beauty, fortune and intelligence are like the independent constituents of an ideal spouse and are truly incommensurable qualities, what would be irrational for me, after all, would be to renege on my cyclical choices just to restore coherence. Like a member of a political assembly, I might well foresee that my choices will soon become cyclical , but there is nothing I can do about it.[9]

Granting the already made objection about the artificiality of dimensions, May's and Tversky's are the only significant examples of this section. They share a common pattern. In each, the agent is faced with two vectors of characteristics $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, which we may write as the two lines of a matrix, and the

[8] Compare with the commentaries, most of them critical, at the end of Blyth's article, in particular those by Good and Winckler (Lindley *et al.*, 1972, p.375).
[9] Roemer makes a related argument in defence of May's cyclical choosers.

characteristics i are unproblematically evaluated by means of a particular numerical scale u_i. A decision rule amounts to a particular mode of aggregating the information given by all the $u_i(x_i)$, $u_i(y_i)$. Both May and Tversky assume that *vertical* comparisons come first: the differences $u_i(x_i)-u_i(y_i)$ are computed before aggregating across the dimensions. Classical decision theory (of which expected utility theory is a particular case) assumes that *horizontal* comparisons are made first: one first aggregates across the i for each given alternative, and a difference is taken only at the end. By construction, horizontal aggregation cannot lead to intransitivities, while vertical aggregation generally does.[10] This said, there is a point to be made in favour of the vertical method.

Firstly, in the 'Paretian' case when one of the two alternatives dominates the other with respect to all characteristics, the method can conveniently be applied by simple inspection of the matrix. Secondly, and relatedly, it starts by comparing quantities that are already commensurable, thus postponing the more problematic step of aggregating those expressed in different units. Intuitively, this is a wise procedure because at least in the Paretian case, the last step becomes dispensable, while in the general case, the way of performing it might depend on the choice circumstances. Sometimes, it will be possible and not too costly to define a complete system of exchange rates between dimensions. Sometimes however, this is impossible (because dimensions appear to be incommensurable), or impractical (typically, when there are too many dimensions). Then, various second-best rules (such as the majority rule or comparisons along only a few prioritized dimensions) come into play. The vertical method of aggregation is compatible with any of these resolutions; it has a flexibility that the horizontal method lacks. Thirdly, and more tentatively, the vertical method seems to be the closer of the two to the deliberative mode of decision-making. To deliberate about a course of action is to examine arguments for and against it in turn, and suspend judgment until sufficiently many significant arguments have been considered. Here, the course of action is to choose X rather than Y, and the relevant arguments are the differences in each scale $u_i(.)$. Judgment is passed only after all (or sufficiently many) of the differences have been calculated.[11]

The following distinction will perhaps help one to assess the normative strength of the best examples in this section. In general, I shall say that in a given situation, a failure of optimization is *strongly* rational if it appears to be entailed by rationality as intuitively understood, and that it is *weakly* rational (or reasonable) if it appears only to be compatible with it. In other words, given the choice situation, a strongly rational counterexample is entailed by all pretheoretic notions of rationality that come to the mind in this situation, while a weakly rational one is entailed by at least one, but perhaps no more than one, of those pretheoretic notions. Are the intransitivities analyzed in this section weakly or strongly rational? The discussion of the last paragraph - about the advantages of the vertical method of aggregation - suggests that intransitive choice behaviour is at

[10] Tversky (1969) has further investigated the formal disanalogies between the two modes of aggregation. In the field of risky choice, regret theory (Loomes and Sugden, 1982) is another application of the vertical method, leading again to intransitivities. Fishburn (1982, 1988) has extensively investigated intransitive variants of expected utility theory. Like continuity, the transitivity axiom is seen as less compelling in the field of risky choice than in the field of choice under certainty, a disanalogy which would deserve further investigation.

[11] The first and second points of this paragraph are very much Tversky's (1969, 1972a and b).

least weakly rational in a wide range of situations.[12] Can one go beyond this conclusion? Certainly not on the basis of the arguments made thus far. To claim that certain intransitive behaviour is strongly rational would require one to show that transitive behaviour would not even be weakly rational under the same circumstances, and that is more than we can hope to establish, given the available evidence and arguments.

2.5 A word on the 'money pump' argument

For the sake of comparison, a word might be said about the opposite, more popular strategy of arguing that intransitivities are not even weakly rational in certain relevant circumstances. Best known in this area is the so-called money pump argument, inititiated by Davidson, McKinsey and Suppes (1955). It aims at showing that an agent entertaining cyclical strict preferences incurs the risk of being ruined. Here is one illustration: The head of an American university department offers a prospective employee three options: a = a top-level post at $50,000 a year; b = an intermediate-level post at $55,000 a year; c = a low-level post at $60,000 a year. The candidate prefers a to b because more prestige compensates for less money, and b to c for the same reason; but he also prefers c to a because the financial gain compensates for the loss of prestige. In each case the preferences are strict. The head of department, who is not overly scrupulous, propositions the candidate as follows in an attempt to bankrupt him: 'I see you prefer b to c, so I'll let you have b in exchange for $25'. The candidate quickly hands over the cash. 'I believe I am right in thinking that you would prefer a to b. Never mind, just give me $25'. No sooner said than done. 'It would appear that you prefer c to a ..." and so on. Hence the term 'money pump'. An individual having cyclical preferences makes himself a prey to exploitation by others. Alledgedly, this conclusion establishes that cyclical preferences are irrational (not even weakly rational, in the above terminology). Notice the a priori structure of the argument: like Tversky's and Blyth's, it goes from assumed preferences to entailed cyclical choices, but this time, of course, with a view of disqualifying the assumed preferences.

Davidson, McKinsey and Suppes's scenario, when analyzed, is seen to rest on several tacit suppositions. The most obvious one is that the candidate places a monetary value on the replacement of one option by another at each of the three stages of the cycle. This assumption is in the spirit of standard microeconomics, and can be formalized by saying that the agent has a complete and continuous preference relation over a set of alternatives made up of various combinations of professional positions and money amounts. Such a formalization points towards a first weakness in the argument for acyclicity - it depends on accepting other axioms in the first place. Thus, even a defender of the argument has to concede that it does not applies universally; it applies only to those circumstances in which the commensurability of money and the swap of alternatives can plausibly be assumed.

[12] By saying this, I do not want to imply that intransitivity is somehow desirable. At best, intransitivities are the inevitable consequences of a method of decision which is *itself* defensible.

There are other more subtle assumptions on the choice sets at each stage of the argument, and this leads to identifying a second weakness. Now, suppose that at the stage corresponding to the first payout, the choice was between a, b and c; there would be no reason for the candidate to hand over anything in order to get b since a is strictly preferred to b. So, at this stage, the choice must be betweenb and c. By the same reasoning the second and third stages can lead to a pay out only if the choice is between a and b , and c and a , respectively. But, even granting an initial choice set {b,c}, it is not clear why the candidate would want to pay to get b. Suppose for example that the candidate anticipates one step (and only one step) ahead of his current position. Then, initially, he would know that he would next drop b in favour of a which he actually prefers, and accordingly decide to make only one payment, namely the second. But through the same anticipatory reasoning, when the candidate is faced with {a,b}, the second payment becomes worthless to him. Similarly with the third payment when he is confronted with {c,a}. He ends up paying out nothing at all.

The foregoing counterargument is based on a perhaps little plausible hypothesis of myopic anticipation, but nothing in the initial wording of the money pump argument excludes this possibility. Besides, it seems as if to extend the candidate's time-horizon further - covering more than one step - would make any pay out rather strange, because the candidate would then see the money pump at work. Commentators - notably Schwartz (1986), who wrote in detail about this issue - usually agree that the argument does not hold unless one assumes that the agent has no anticipation whatsoever about the future.[13] This supposition is scarcely plausible to begin with, and, as the cycles are reproduced it becomes even less so. One might perhaps concede that the candidate would lose a few $25 bills but not that he would bankrupt himself!

Other objections have been raised against the money pump argument. Virtually all those who have examined it end up claiming that is inconclusive.[14] It is safe to say that other things being equal, acyclical preferences do not lead the candidate to bankruptcy, but of course, he argument did not aim at establishing this trite point. It aimed at showing that acyclical preferences are the only ones which preserve the candidate from bankruptcy, i.e., in the above terminology, that in the given circumstances, acyclical preferences are strongly rational. The failure to establish this exactly parallels the failure to establish that under certain circumstances, cyclical preferences are strongly rational. [15]

3. Optimization from the Perspective of Revealed Preference Theory

Revealed preference theory analyzes the relationships between a) the properties of individual choices when these

---

[13] The analysis of the money pump argument has recently been pursued more thoroughly in terms of dynamic rationality principles, see MacClennen (1990) and Rabinowicz (1995).

[14] Besides Schwartz (1986, p. 128-131) see, among others, Anand (1987, 1993), Bar-Hillel and Margalit (1988), Schick (1986) and Sugden (1985). An earlier article of Schwartz was revealingly called "Rationality and the Myth of the Maximum" (1972).

[15] There have been other suggestions less famous that the money pump argument to show that transitivity is a strongly rational property. An argument of this kind by Tullock (1964) is often cited, but is unanimously considered defective. The reader will find other references and arguments in Fishburn (1991). He concludes that "reasonable people sometimes violate transitivity and may have good reasons for doing so" (p. 131).

choices bear on subsets of the full set of available alternatives, and b) the existence and properties of a binary preference relation on the full set. In its original version, due to Samuelson (1938), the theory applied only to the neo-classical consumer, which imposes a special structure on the set of alternatives and on the set of choice subsets. In order to meet the requirements of social choice theory a variation that is both more elementary and conceptually more general was devised. This version, which is the only one needed here, does not restrict alternatives and describes choices just by simple set-theoretic properties. I will rely more particularly on Richter (1966, 1971) and Sen (1970, 1971). Their results make it clear how the two axioms of optimization, namely the transitivity and completeness of the binary preference relation, can be broken down axiomatically into conditions put on the so-called choice function.

A review of these classic results will serve two purposes here. First, the results have been used to justify optimization in terms of alledgedly 'natural' equivalent properties. This justification strategy underlies much of the abstract literature on revealed preference (though not the initial Samuelsonian version, which is meant to be descriptive). To illustrate, we cite at some length Sen's early work, where it is put to use explicitly. So one aim of this section is to assess a prima facie significant argument made in favour of the claim that rationality implies optimization. I will dismiss the argument, a conclusion which connects with a second purpose of this section: it also serves to emphasize the *cognitive costs* of the mental operations by which an agent maximizes a transitive and complete preference relation. The theorems of revealed preference theory are one way of introducing this issue, even if few writers have discussed it from this perspective - Plott (1973) being one of the exceptions. The theme of cognitive costs is central to the claim that not only does not rationality imply optimization but even the converse may not hold.

3.1 Those Famous Conditions Alpha and Beta.

Formally, we are given a non-empty set X of unspecified objects of choice, a non-empty family Σ of subsets of X, not including $\emptyset$, and finally a function h: $\Sigma \rightarrow 2^X \setminus \{\emptyset\}$ satisfying $h(S) \subseteq S$, $\forall S \in \Sigma$. Thus h picks out a subset of each set of options S. For this reason it is called a*choice function*. The condition $h(S) \neq \emptyset$ (when $S = \emptyset$ is excluded) is universally accepted in the theory, but is not conceptually vacuous.[16] The technical question is, what conditions on h are equivalent to the property that 'h arises from optimization', or, more explicitly, that 'there exists a binary relation such that the agent maximizes in each choice situation S '. This property is defined formally as follows: there exists a reflexive and complete binary relation, R, such that:

$\forall S \in \Sigma$, $h(S) = \{x \in S | \forall y \in S\ , xRy\}$ (= the set of the best elements in S according to R).

Consider first the particular case where the choice function is complete and single-valued, that is, $\Sigma = 2^X \setminus \{\emptyset\}$ and $\forall S \in \Sigma$, $h(S)$ is a singleton. Then h arises from optimization if and only if h satisfies property α, which is

[16] The value $h(S) = \emptyset$ can be taken to mean that the objects in S are non-comparable. To exclude this value, while taking h to be completely defined, is thus equivalent to excluding non-comparability.

defined as:

$\forall S, S' \in \Sigma$ ($S \subseteq S'$ and $x \in h(S')$ ® S) => $x \in h(S)$.

'If the world champion in a particular discipline is a Pakistani, he must also be the Pakistani champion' (Sen, 1970, p.17). This condition has often been stated as a minimum rationality requirement in the social choice and game theory literatures. Consider now the more general case where the choice function is still complete but not necessarily single-valued, i.e., we just require that $\Sigma = 2^{x} \setminus \{\emptyset\}$. Then h arises from optimization if and only if it satisfies properties α and β. Property β is defined thus:

$\forall S, S' \in \Sigma$ ($S \subseteq S'$ and $h(S')$ ® $S \neq \emptyset$) => ($y \in h(S)$ => $y \in h(S')$).

'If a Pakistani is world champion, then all the Pakistani champions are world champions' (*ibid.*). The conjunction of α and β , that is to say:

$\forall S, S' \in \Sigma$ ($S \subseteq S'$ and $h(S')$ ® $S \neq \emptyset$) => ($h(S')$ ® $S = h(S)$)

is sometimes labelled *independence of irrelevant alternatives* (but it is not the same as Arrow's condition, as was pointed out by Karni and Schmeidler, 1976). Others call it the strong axiom of preference.

Conditions α and β are most famous among those stated by revealed preference theory. Being apparently such weak rationality conditions, they provide interesting support to the orthodox economist's contention that rationality implies optimization. A first objection to this claim is that they do not have the same normative force. Condition α says: 'If one discards an option after comparing it with others in some subset, then it will not be retained when one compares it with a new subset of options containing the former subset'. Condition β says: 'If one selects two options after comparing them with others, and one still selects the first after comparing it with a subset of options containing the previous options (hence in particular the second), one will retain the second option along with the first'. If choices occur sequentially, α says that the selection process is not creative, while β says that it is not destructive. The first requirement is more basic than the second. Consider a sequential choice process which satisfies α and suppose that the number of champions selected at each stage is bounded above, so that β may be violated. I cannot see why this process should be excluded on rationality grounds, although it could be, it seems, on equity grounds. In the absence of any distinguishing information, it is arguably not *equitable* to oust a player at the expense of an equal, so that β seems to apply with equal force than does α. When the 'champions' are replaced by alternatives of choices, this symmetry principle does not apply. Sen's 'champions' metaphor might cause the significant difference between α and β to be overlooked because of its peculiar normative connotations. Notice in passing that the following point has emerged again: individual choice theory does not allow for exactly the same formal considerations as social choice theory. But in the context of intransitive choices, the claim that the two theories are disanalogous (since characteristics are not individuals) had the effect of weakening a criticism levelled against optimization; whereas, this time, the claim that they are disanalogous (since alternatives are not individuals) runs against a defense of optimization.

A second line of argument involves distinguishing α from α and β taken together in terms of the memory and computation requirements implicit in either axiomatization. If $X = \{a,b,c\}$ and the choice function h is defined on the pairs as follows: $h(\{a,b\}) = \{a,b\}$, $h(\{a,c\}) = \{c\}$, $h(\{b,c\}) = \{b,c\}$, α is compatible with three solutions for $h(\{a,b,c\})$, i.e., $\{b\}$, $\{b,c\}$ and $\{c\}$. By contrast, α and β taken together entail the unique

solution $\{b,c\}$. Consider now the following simple sequential choice procedure: the agent takes an arbitrary pair, retains only one best element in the choice made from this pair, forms another pair with the remaining element, and again retains only one best element. This is an algorithm for computing $h(\{a,b,c\})$, and it is cost-effective. It saves memory space since from the pair, say, $\{a,b\}$, only a or b is retained as a result of the choice made on this pair. The algorithm also economizes on calculations: at the next stage after $\{a,b\}$, the agent will compare c and a, or c and b, but not both. Obviously, the procedure satisfies α, but violates β, and thus the optimization model as it has just been axiomatized. If the procedure is now modified to satisfy β, the memory space will have to be expanded, and more calculations to be performed. For larger sets than the three-element one mentioned here, the difference in total operating costs can be considerable. This negative consideration would have to be balanced against the claim that β is a rationality requirement as strong as α, supposing that this claim were made.

3.3 Completeness

The previous discussion was confined to those choice functions which are totally defined, i.e., when $\Sigma = 2^X \backslash \{\emptyset\}$. To stop at this case is tantamount to taking for granted one of the two axioms of optimization that are at stake in this paper, i.e., completeness. Does revealed preference theory provide a normative argument for this axiom, as it does for optimization once completeness is granted? I claim not. To establish this point, I must review two further equivalence results of the theory. Let us say that the function h defined on an arbitrary set Σ of nonempty subsets is *binary* if some binary relation R exists, such that:

$\forall S \in \Sigma$, $h(S) = \{x \in S | \forall y \in S, xRy\}$.

Then, the first result of interest tells us that h is binary if and only if it satisfies property α^+:

$\forall x \in X$, $\forall S \in \Sigma$ ($x \in S$ and [$\forall u \in S$, $\exists S' \in \Sigma$: $u \in S'$, and $x \in h(S'')$]) => $x \in h(S)$.

This property is logically stronger than α, and not comparable with α and β together. It is often formulated in terms of the so-called *revealed preference* relations. Define: xVy ('x is directly revealed as preferred to y') if there is S such that $x \in h(S')$ and $y \in S$. Then α^+ becomes:

$\forall x \in X$, $\forall S \in \Sigma$ ($x \in S$ and [$\forall u \in S$, xVu]) => $x \in h(S)$.

In words, x is directly revealed as preferred to y if there is a choice subset (possibly different from the initial subset) from which x is chosen when x and y can be. Condition α^+ thus imposes a form of coherence on the individual's multiple choices and does have some normative force. When he compares x and y in S, this comparison must implicitly take into account that the same elements might have already been compared in some other S'. The trouble is that α^+ still does *not* ensure that h arises from optimization.

There is a need for a stronger notion of revealed preference than V, and the early developments of revealed preference theory demonstrate that it was not obvious how to formulate it exactly. Define: xWy ('x is indirectly revealed as preferred to y') if $x^0, x^1, \ldots x^n$ exist, such that

$x^0 = x$, $x^n = y$ and $x^0 V x^1 V \ldots x^{n-1} V x^n$.

In words, x is indirectly revealed as preferred to y if there exists a sequence of sets $S_1,\ldots,S_{n-1}$, and of options x^1,

..., x^{n-1} chosen from these sets, such that x is directly revealed as preferred to x^1, x^{n-1} directly revealed as preferred to y, and each intermediate x^i directly revealed as preferred to the following one. The conclusion now is that h arises from optimization if and only if h satisfies *property* κ (called *congruence* by Richter, 1966):

$\forall x, y \in X, \forall S \in \Sigma\ (x \in h(S), y \in S$ and $yWx) \Rightarrow y \in h(S)$.

(This second result of interest is a set-theoretic trivialization of the theorem in consumer theory which Samuelson groped for, and which was finally proved in the 50's). By comparison with α^+, which it implies, κ strongly reinforces the constraint imposed on multiple choices. The choice between any two elements of S, x and y, is now determined by what earlier choices revealed *not only directly, but also indirectly* - through comparisons between successive choices made in a sequence. Mathematically, W is the transitive closure of V.

On one reading, κ is just another coherence condition imposed on the decision-maker. I do not think that that this interpretation is very plausible. Consider again $X = \{a,b,c\}$ and assume now that:

$h(\{a,b\}) = \{a,b\}$, $h(\{a,c\})$ = not defined, $h(\{b,c\}) = \{b,c\}$.

The only solution conforming to κ for $h(X)$ is:

$\{x \in X | \forall y \in X, xWy\} = \{a,b,c\}$.

That a belongs to $h(X)$ follows because the missing information on $h(\{a,c\})$ has been replaced by a calculation: the theorist has constructed the transitive closure W of the V relation, i.e., $aVbVc \equiv aWc$ and $cVbVa \equiv cWa$. Emphatically, it is the decision theorist who makes the calculation and *decides* that a belongs to $h(X)$. There is no corresponding choice on the agent's part. Should he nonetheless repeat the theorist's inference, which in the particular instance, means treating a, b and c completely symmetrically? There is at least one typical situation in which he should resist the inference - i.e., when he is truly unable to establish a comparison between a and b. Briefly put, the transitive closure W is not endowed with any meaning in terms of the agent's activities or proclivities. I do emphasize the distinction between V and W, or conditions α^+ and γ: A rational agent is in some sense committed by the earlier choices he made, but I cannot see the sense in which he should be committed by the inferences that the observer draw from these choices.

I do not expand on the cognitive costs implied by κ, precisely because I cannot see how to interpret this condition from the agent's point of view. The criticism just presented is coherent with the claim sometimes made even among classical theorists that this axiom does not enjoy the same normative status as transitivity.[17]

3.3 A Counterargument

At this juncture, a defender of optimization could resort to a curious defense strategy for optimization, which was suggested - perhaps in passing - by Sen (1971, in 1982, pp. 48-49). It consists in arguing that one of the initial special cases - that of a complete choice function - is, the appearances notwithstanding, the relevant one to

[17] See for instance Aumann: "of all the axioms of utility theory, that of completeness is perhaps the most debatable" (1962, p. 446). Luce and Raiffa (1957, p. 25) have also criticised this axiom.

consider. The gist of this argument is to justify optimization not in terms of κ, but of the milder axioms α and β. Why, Sen asks, should one restrict attention to a particular family Σ of subsets? The nature of the choice problem - for instance, in consumer theory - might entail fixing Σ - in that example, Σ is the set of all "budget triangles". But at the level of a general argument about rationality, no selection criterion presents itself. In the absence of a reason for selecting *a* particular subset, it is appropriate, Sen claims, to adopt the set of *all* subsets of X.[18]

A curious feature of this argument is that it raises *at the metalevel of the theorist* a classic problem that revealed preference theory raises at the level of the agent, i.e., the confounding of non-comparability with indifference.[19] The theory captures the agent's indifference by allowing for multi-valued $h(S)$. There are two possible ways of capturing non-comparability: either one states that $h(S) = \emptyset$ for some S , or one allows that h is not defined for all S. For technical reasons, the latter modelling has prevailed over the former. Now, consider Sen's revealed preference theorist, who does not restrict the subsets relevant to the agent. By recommending that the theorist should reason on the set of *all* subsets, Sen implicitly proposes a selection function for this theorist, and in effect - I am pursuing here the basic analogy provided by revealed preference theory itself - he analyzes the theorist's indeterminacy as a case of *indifference* between all subsets. But indifference is only one of two possible analyses of the lack of choice, and in the particular instance, the less plausible of the two. Rather than being indifferent between the subsets, the theorist has failed to establish relevance comparisons between them. If the latter, not the former, is the case, it is right to investigate the properties of choice functions for arbitrary domains Σ. I do not want to suggest that Sen himself would lay much emphasis on this little piece of argument I tried to disentangle.

3.4 Optimization and Path-independence

Another attempt to justify optimization can be made by appealing to an alternative axiomatization in terms of path independence. Informally, this property says that the choice finally made from the whole set X should not depend on the path taken through the set Σ of choice subsets. Its first technical formulations are pre-war in origin, when microeconomists - following Pareto's lead - were investigating the integrability of demand functions. Abstracting again from the framework of consumer theory, we get set-theoretic definitions of path independence, such as Plott's (1973):

$$(IP)\quad \forall S, S' \in \Sigma,\ h(h(S) \copyright h(S'')) = h(S \copyright S'').$$

Plott's axiom ensures that the choices made in any set T coincide with the result of the two-step choice defined

---

[18] "Why therefore restrict the domain of an axiom to [Σ] and not to [$2^X \setminus \{\emptyset\}$] when (a) the satisfaction of [the axioms] is not possible either in the case of [Σ] or in that of [$2^X \setminus \{\emptyset\}$], and (b) there is no a priori reason to expect that the axiom is valid on [Σ] but not on [the complement of Σ]" (Sen, 1982, p.48) Condition (a) need not retain us here; it refers to the fact that there are typically too many subsets for Σ to be observable.

[19] In a different (and more critical) piece on revealed preference theory Sen (1982) has usefully discussed this problem. It connects with the philosophical conundrum of Buridan's ass (on which, see Rescher, 1982).

by first choosing from within two subsets S and S', which together make up T, and, second, choosing between the alternatives thus selected. This should hold whatever the splitting of T into S and S', so that the axiom can indeed be taken to state independence of path. It is easy to see that if h arises from optimization, it satisfies *IP*, but an example in Plott (1973, p. 1081) shows that the converse does not hold. At least, *IP* implies α (this follows from another remark of Plott's (1973, p. 1087)), and this implication is strict. It thus follows that in terms of logical implication, path independence lies between optimization and condition α. So the alternative justification fails, even if it goes some way beyond the first stage of the previous argument for optimization.

Intuitively, *IP* is an attractive stopping place to formally define a rationality concept implied by, but not implying, the optimization condition. This move would be a very clear instantiation of what I have called the conciliatory account in section 1. If *IP* or similar conditions were not met, the individual's final choice could change even though the choice circumstances remain the same. Rationality, whatever it exactly means, appears to imply that choices should remain invariant with respect to, at least, the *objective* circumstances. Then, these circumstances being fixed, the outcome of a deliberation should not depend on the way it is conducted.[20] However, there is another, quite opposite intuition about path independence and rationality. A rational choice, I mentioned at the beginning, is a choice made for good reasons. A well-conducted deliberation is by itself a good reason for the choice it results in. But crucially, the concept of a well-conducted deliberation does *not* involve that of a uniquely determined conclusion; that is, the external circumstances being fixed, another well-conducted deliberation could possibly lead to a different conclusion. It is the properties of the *procedure*, not of its outcome, that are referred to by the adjective 'well-conducted'. Simon's later work (e.g., 1976) usefully defends such a view of rationality, which he calls the 'procedural' view, by contradistinction with the 'substantive' view, which underlies (among others) the optimizing model and is outcome-oriented.[21] The two conceptions depart from each other in the way they deal with the agent's internal choice circumstances, such as his memory or his computation abilities. Typically, the 'substantial' theorist either does not take internal circumstances into account, or somehow includes them into the description of either the available means or the external circumstances. By contrast, a 'procedural' theorist deal with internal circumstances as being distinctive factors of the choice.[22] I do not want to expand here on Simon's distinction but just to use it to clarify the conflicting rationality intuitions surrounding Plott's *IP*. The property this axiom formalizes is no more than weakly rational (since there is a no less attractive intuition of rationality which does *not* warrant path independence). And anyhow, as already said, *IP* does not imply that the choice function results from optimization.

I conclude this review of revealed preference theory by stressing that it does not provide an argument for optimization. The argument has brought two significant by-products. First, we have just seen that at least on

[20] This line of argument has deep roots in the Classical Rationalists.

[21] Correspondingly, Simon (1976) deemphasizes the contrasts of his earlier papers between 'bounded' and 'absolute' rationality, or 'satisficing' and 'optimizing'. Mongin (1986) argues that the procedural versus substantive distinction is more novel and fundamental than these other, more popular distinctions.

[22] The procedural theorist can accept the principle above that rationality implies that choices should remain invariant with respect to the *objective* choice circumstances, but he regards some of the *internal* choice circumstances as being objective.

one relevant construal, the conciliatory account is not on a safe basis. Second, the issue of internal circumstances should urgently be addressed. Using naive principle of cost-effectiveness - i.e., that memory demands increase strictly with the number of previously selected alternatives, and calculation costs with the number of operations performed - I have tried to discriminate between α and β. Plott (1973) proceeds no differently when he argues for some weak forms of path-independence - and implicitly, against the stronger forms, hence against optimization.[23] The simple point about cognitive costs is this: the agent's incurring extra cognitive costs as a consequence of complying with a particular condition or principle must count negatively when one assesses the overall rationality content of this condition or principle - since, intuitively, rationality implies making good (cost-effective) use of the available means. In the next section, I sharpen this point to argue that optimization can sometimes be not even weakly rational.

4. The Infinite Regress of Optimization

All decision principles lead to an infinite regress. Applying a principle is just another decision to make, and the question arises of whether or not this decision satisfies the given principle. As a particular application of this general problem, we have just seen that to optimize is, implicitly, to *decide* to implement a choice function of a certain kind. Supposing that the cost of each kind of choice functions is known, one needs to ask whether it was optimal to implement a choice function of that particular class, and not one from another class. If one now assumes that to answer this question requires a choice function of higher order, which itself has an implementation cost, a new question arises, and so on *ad infinitum*. To investigate the problem exemplified by this reasoning I will drop the reference to the choice functions of revealed preference theory. They have provided an initial example, but are just an example. It should be clear that any alternative account of optimization is open to an infinite regress objection. A conveniently general formulation is to say that to optimize requires one to run a costly *algorithm*, that to optimally select the costly algorithm requires one to run another costly algorithm, etc. In this section, I will explain, and then illustrate, how this problem can be sharpened into a significant objection against the optimizing model of rationality.

4. 1 Not Every Infinite Regress Is 'Vicious'

I first need to make a distinction between two kinds of infinite regress, and in order to introduce it, I will discuss (and rebut) Ryle's statement of the infinite regress of decision in *The Concept of Mind* (1949). In a well-known passage, he raised the following objection against those conceptions of action which he calls *intellectualist*:

"To put it quite generally, the absurd assumption made by the intellectualist legend is this, that a performance

[23] See also Campbell's (1978) definition of "calculation viability" which follows Plott's intuitions. Some work along the lines of the theory of algorithmic complexity would also clearly be to the point here.

of any sort inherits all its title to intelligence from some anterior internal operation of planning what to do. Now very often we do go through such a process of planning what to do, and, if we are silly, our planning is silly, if shrewd, our planning is shrewd. It is also notoriously possible for us to plan shrewdly and perform stupidly, i.e., to flout our precepts in our practice. By the original argument, therefore, our intellectual planning process must inherit its title to shrewdness from yet another interior process of planning to plan, and this process yet another interior process of planning to plan, and this process could in its turn be silly or shrewd. The regress is infinite, and this reduces to absurdity the theory that for an operation to be intelligent it must be steered by a prior intellectual operation." (Ryle, 1949, pp. 31-2)

Certainly, this is a striking passage[24], but what exactly is its polemical target? All the existing theories of rational choice describe the performance of choice as being steered by some "anterior internal operation of planning", and they assess the rationality of choice in terms of the rationality of this underlying process. So all these theories are - apparently - 'intellectualist' and hence open to Ryle's sweeping objection. It would hit not only the optimization model but also any of the emerging alternative models. But is there an objection after all? I think not - at least, not at the present stage of the reasoning. Ryle evidently believes that in order to reject a theory, it *suffices* to show that it leads to an infinite regress (cf: "the regress is infinite, and this reduces to absurdity the theory"). To see that this cannot be the case, consider the standard assumption in game theory that the rules of the game (i.e., the sets of strategies and the pay-offs) are common knowledge. Definitionally, this means that each player knows the rules of the game, knows that each player knows the rule of the game, and so on. Are we to dismiss the common knowledge assumption as being 'absurd' on the grounds that (for suitably large state spaces) no finite level of mutual knowledge reached between two players can exhaust the content of the assumption? Or, to take an explicitly decision-theoretic counterexample, consider this other familiar notion in game theory - rationalizability. Informally, a strategy s_1 is rationalizable for player A if it is a best reply to some strategy t_1 of player B, which is itself a best reply to some strategy s_2 of A, the latter being itself a best reply to some stategy t_2 of B, and so on, possibly ad infinitum.[25] To paraphrase Ryle, A's planning to play s_1 inherits its shrewdness from another act of planning, this time by B, to play t_2, and "the regress is infinite". Are we to reject rationalizability on this ground? This would be a ridiculous inference, as in the common knowledge case, and here is one quick argument why it is so: there are restatements of both rationalizability and common knowledge which make no reference to infinity.[26]

The situation exemplified by these two concepts is typical. Not every infinite regress is logically vicious, and one way to recognize that is to compare it with restatements of the problem when available. Generally, I suggest

---

[24] Ryle's critique of 'intellectualism' has attracted considerable attention from contemporary philosophers of mind.

[25] The sequence is infinite only if there is an infinite number of strategies, which happens in standard game theory when the rationalizability concept is applied to *mixed* strategies.

[26] Compare Aumann's (1976) sequential definition with the unproblematic one in terms of the meet of the agent's partitions, and compare, more generally, the iterative and circular approaches to common knowledge (Lismont and Mongin, 1994). As to rationalizability, compare for instance Bernheim's (1984) initial definition with a non-iterative restatement in Bernheim (1986).

formalizing infinite regresses as infinite sequences for which some appropriate notion of convergence can be defined. Then, a regress will be said to be 'vicious' and 'harmless' depending on whether the infinite sequence that formalizes it diverges or not. Viewed this way, infinite regress arguments cannot be expected to deliver ready-made refutations against a whole *class* of theories as in Ryle; since the properties of the sequence will typically depend on the particular values of the parameters. The most natural definition of convergence in the context of decision-making is arguably*stationarity*: for some integer n, the level $n + 1$ leads to exactly the same n-level decision, and similarly for the levels after $n + 1$. That is, there exists a logical level such that the decision recommended by the given principle at this level coincides with the decision recommended by the same principle at all subsequent levels. Alternatively and less plausibly, it is conceivable to use the more general notion of *asymptotic convergence* (relative to either the space of level 1 decisions, or, perhaps more conveniently, but with possibly different results, to the space of utility values).

I now move the next issue of how to apply the general method sketched here to optimization.[27]

4. 2. An Example of the Infinite Regress of Optimization.

When applied to optimization, the infinite regression can be destabilizing in two distinct ways. First, an optimal solution in the ordinary sense can be downgraded in favour of a meta-optimal solution which is in turn downgraded, and so on, level $n + 1$ never 'supporting' level n. Second, the optimizing method *itself* can be downgraded in favour of another, for example, the search for a 'satisficing' value, a method which, in turn, is liable to be downgraded, and so on. [28]

I will try to illustrate, by means of an easy calculation, an infinite regression of the former and simpler type (downgrading of the optimal decision but not of the optimizing method itself). Given E, the set (assumed finite) of states of nature, D, the set (also finite) of decisions, and U, the utility function: $E \times D \rightarrow \Re$, classical decision theory says that the agent finds a rule

$r_1^* : E \rightarrow D$

satisfying the optimization condition, that is to say, such that for each e:

(1) $U(e, r_1^*(e)) = \max_d U(e,d) = \max_{r1 \in D^E} U(e, r_1(e))$.

(As usual, D^E denotes the set of all functions from$E$ to D.) Let us now assume (that the theorist attributes to the agent)[29] the following cost function on rules:

$C_1 : E \times D^E \rightarrow \Re$

[27] There are very few systematic discussions of the infinite regress of optimization. To the best of my knowledge, they are those of Göttinger (1982), Mongin (1984), Mongin and Walliser (1988) and Lipman (1991).

[28] More on the latter in Mongin and Walliser (1988). It is, however, much more difficult to tackle than the former.

[29] This, and other parenthetic statements below, are meant to alert the reader to the fact that there is also an interpretation of the formalism going in terms of theoretical steps taken by an ideal observer. More on this interpretation in 4.3.

Then (following the theorist) the agent must find the rule

$r_2^* : E \rightarrow D^E$

which to each e assigns a rule at level 1, $r_2^*(e)$ such that:

(2) $U(e, r_2^*(e)(e)) - C_1(e, r_2^*(e)) = \max_{r1 \in D^E} [U(e, r_1(e)) - C_1(e, r_1)]$.

Comparing (2) and the second equation in (1), there is nothing to ensure that there is e satisfying:

$r_2^*(e) = r_1^*$

or even such that:

$r_2^*(e)(e) = r_1^*(e)$.

And one can a fortiori replace 'there is e' by 'one has for all e' in the preceding sentence. Briefly, meta-optimization can very well contradict optimization.

Let us take a particular application where the agent is a business, assuming that its production function depends only on the factor labour. The state variable e that is relevant to his decision-making is the wage rate, U is the business's profit function. This gives some flesh to the maximization programme (1). To explain (2), let us now assume that the business employs programmers with the job of determining the optimum level of production, and they are paid the same rate e. The difference between (2) and (1) arises because the business is trying to 'internalize' its programming expenses, measured by C_1, which it had not originally taken into account. The result of the second calculation can obviously upset the result of the first.

What has been said for the first two levels apply to all others. Accordingly, I will define:

- an infinite sequence of sets of rules:

$R_1 = D^E$, $R_2 = D^{E^E} \approx D^{E \times E}$, ...

- a corresponding sequence of cost functions:

$C_1 : E \times R_1 \rightarrow \Re$, $C_2 : E \times R_2 \rightarrow \Re$, ...

The optimization programme of order n generalizes condition (2). The preceding microeconomic example again illustrates that to calculate the optimum level of output (both physical and intellectual) at level n -1 imposes on the business a cost determined by C_{n-1}, a cost which is taken into account only in the programme of order n. No programme automatically takes account of its own cost. The observation that meta-optimization can contradict optimization now applies to the programmes of any order n .

Let us say that an infinite regression *converges for all* $e \in E$ *(for some* $e \in E$) if $n \in N$ exists such that for all $e \in E$ (resp. some $e \in E$):

$r_n(e) = r'_n(e)$.

Various plausible assumptions can be introduced on the programming costs, and some of them quickly lead to the conclusion that infinite regressions typically do not converge. I just mention a particularly easy case.

ASSUMPTION:

$(\forall n)\ (\forall r_n, r'_n \in R_n)\ (\forall e \in E)$

$|\mathrm{Im} r_n| > |\mathrm{Im} r'_n| \Rightarrow C_n(e, r_n) > C_n(e, r'_n)$

(I.e., *ceteris paribus*, a rule costs more to implement the more decisions it includes.)

PROPOSITION: if this assumption holds, then either r_1* is constant, or the infinite regression does not converge on any e.

(The proof is easy and left to the reader.)

In plain words, either optimization establishes from the outset a crude behavioural pattern, or it never converges. Alternative assumptions that are just as easy to formulate would prevent the convergence of infinite regression for at least *some* e.

The little model of this section has a special feature which calls for a comment.[30] The variable e influences *each* level of decision, and not every meta-optimization problem, even of the simpler sort, has this particular structure. For concreteness one can imagine that the same programmers are called upon at each stage to perform a calculation that they had not performed at the previous stage. To generalize beyond this concrete example, one could hypothesize that each level n defines a new category of programmer especially suited to his task, but that the salaries e_n of the different categories are fixed proportionally to one another, so that e acts like a scaling factor. At any rate, the infinite regress would completely change in character if the choice of e did not influence all the costs C_n at the same time.

4.3 Some Counterarguments

From discussing the issue with both economists and philosophers[31] I have found that a defender of optimization is likely to dismiss the infinite regress objection in one of these three ways: a) by downplaying it in relation to the true problem in decision theory, which is, alledgedly, to assess and further refine available decision-making criteria; b) by stating axiomatically that individuals only reason at finite levels of thought - while possibly justifying this by empirical arguments; c) by claiming that the infinite regress affects all principles of decision in the same way and thus does not deliver an interesting argument against optimization. I am going to see whether, and how, the infinite regression objection can be sharpened against these points taken in turn. Thus, I am to make precise the objection not so much in the absolute as dialectically.

Thesis a) is not exclusively put forward by orthodox theorists. It could well be, in effect, the one adopted by Herbert Simon on the infinite regress problem.[32] Despite this prestigious backing, I think it is irrelevant to the topic of this paper. Crucial to the assessment of a rationality principle is what may be called its*reflective coherence*, that is, the ability or otherwise of the principle to be applied both to itself and its external conditions of exercise.[33] To check the reflective coherence of optimizing decisions, one strategy would consist in seeing

[30] Jean-Pierre Dupuy raised this issue during a seminar.

[31] Among the latter, Wlodek Rabinowicz whose comments have been very helpful.

[32] Significantly, it is not Simon, but Winter (1975) who first introduced this problem into the discussion of absolute versus bounded rationality.

[33] Reflective coherence, as I present it here, is part of a concept of rationality applicable to the agent. It has to do with the maxim of checking the consistency of one's principle of action or reasoning with the particular way in which it is used . The next pararagraph discusses reflective coherence as a constraint on the theorist.

whether an explicitly *self-referring* notion of optimization can be constructed, that is to say, as Winter (1975) once suggested, in investigating the idea of 'optimization that takes account of its own cost'.[34] The other method, which I have followed after Mongin and Walliser (1988), and is, I think, more tractable, amounts to studying the convergence properties of infinite sequences. A convergent infinite regress of optimzation in the present framework exactly plays the role of an optimization that takes account of its own cost. Whichever method is chosen, the general point is that the reflective coherence of optimization should be addressed in the context of a paper exploring it as a rationality principle. More deeply, I would distance myself from thesis (a) on the grounds that it downplays the normative component of decision theory, but to argue this in detail seems unnecessary here.

Thesis b) stresses that human individuals cannot cope with an infinite number of reasoning stages, as well as - perhaps - the following more subtle point: beyond a certain reflective level n, they can no longer define his choice objects - here, the algorithms (nor, a fortiori, their costs). This thesis is only impressive by dint of a hidden presumption - that the infinite regression actually sets out actual stages to be gone through by the individual. Before trying to deflect this exceedingly literal interpretation, I will take a shortcut, and argue that regardless of what they mean for *the agent*, infinite regresses should at least be a cause a concern for *the theorist*. The latter should check whether he is not employing an incoherent theory. If he were to use optimization to predict the agent's behaviour and if, by ill luck, the optimizing model at each logical level saw its own predictions contradicted at the next level, this would mean a serious flaw in the theory adhered to. I am arguing again for reflective consistency, but now viewed as a constraint *on the theoretical exercise*. Now, to check whether level n+1 decisions reinforce or, to the contrary, upset those of level n, the theorist must know the nature and costs of higher-level objects (algorithms) and be able, if need be, to carry out a large number of calculations. The structure of the argument is then: either the opponent denies that higher-order costs are meaningful to the theorist, and he is unable to claim that the optimizing model is reflectively ccoherent, or he does not deny that, and has to face the infinite regress objection.[35]

But let us take issue more directly with the thesis underlying b), i.e., that infinite regressions set out actual steps to be taken by the agent. This thesis is formidable only if it impossible for a human to cover an infinite number of logical stages in a finite time, which I strongly disbelieve: the time a single stage takes might be infinitesimal. The elementary model of converging series or sequences could be applied here, just as it served to eliminate the alleged paradoxes of the impossibility of movement. More worrying is the point that the data for higher-order levels may not be defined - a problem which, of course, did not arose with the impossibility-of-movement example. It might be the case that only in particular instances, such as that of 4. 2, is the structure of decision well-defined at *each* level. But here I revert to the point just made: the difficulty is for the defender of optimization, who then fails to make his case.

---

[34] See Winter (1975, pp. 81-5). Unfortunately, his analysis stops at this suggestion.

[35] The orthodox economists who put across the counterargument b) - there are some - seem to lose sight of the *ad hominem* structure of the present discussion. Incidentally, it would be curious to see, for once, a naturalist point like b) rescuing optimization!

The counterargument c) has an unsophisticated variant, which has already been covered. If one believes that all decision principles lead to an infinite regress (an indisputable point), and one also believes, like Ryle, that any infinite regress constitutes *ipso facto* an objection, the conclusion follows that the objection does not help discriminate between particular principles. Here, the conclusion has exactly the cash value of the second part of the antecedent. However, there is a sophisticated version of the counterargument, based on the presumption that the alternative models of decision than optimization would also lead to divergent infinite regresses for a wide class of circumstances. I do not want to deny this statement, but it does not sound like a satisfactory answer to an objection raised against optimization. By contrast, what would provide a decisive answer is a comparison between optimization and some representative bounded rationality principle, say, Simon's satisficing, leading to a proof that generically, any time that the infinite regress of optimization diverges, so does the infinite regress of the alternative model. Disregarding the huge difficulties in formalizing the comparison, it seems unlikely to lead to such a favourable conclusion, and I challenge the classical theorist to establish it.

5. Conclusions

I have tried to question the standard optimizing model of rationality from different angles. I repeat the basic distinction between two critical purposes of the paper. The brief review of the money-pump argument in section 2, and the more original discussion of revealed preference theory in section 3 added further evidence against the (still not uncommon) claim that rationality implied optimization. The main critical purpose, however, was to challenge the converse statement, and the conciliatory account which goes along with it. In a similar paradoxical vein, Fishburn sought to establish 'the irrationality of transitivity in social choice' (1970b). Quite obviously, both here and in Fishburn's attempt, the aim was not to conclude that transivity or optimization was always irrational, only that it sometimes was. To avoid crude misunderstandings, I will once again exploit the social choice analogy and let the reader transfer what Fishburn said of his problem to mine:
"The examples presented show that social transitivity is untenable as a *general desideratum* for social choice functions. This does not say that [the social preference relation] should be intransitive for *every* profile of preference but only that there are [some profiles of preference] for which transitivity [of social preference] should not be required." (p.122, my italics)

Another important distinction cuts between two ways of discussing optimization - either in terms of its underlying conditions, as in sections 2, or directly, as in sections 3 and 4. On one reading,[36] the transitivity and completeness of the preference relation are just prerequisites of optimization, while the latter is only the act of selecting a best element for the preference relation. Or nearly equivalently, transitivity and completeness are conditions for the existence of a utility index, while optimization is just the act of maximizing the index. As I suggested in 1.1 this construal entails too narrow a definition of the word "optimization". Those who express

[36] Suggested to me, in particular, by Dan Hausman.

dissatisfaction with the normative claims of orthodox economists surely do not want to question only the rationality of finding a maximum, but also the very assumptions that make the search for a maximum meaningful. Hence the broader and less precise definition of optimization adopted here.

As it turns out, the most damaging examples of intransitive choices are still compatible with the claim that optimization is at least weakly rational, while the assessment of revealed preference conditions also is. The only significant argument against the conciliatory account is the infinite regress of optimization. I emphasize that the latter does not, alas, constitute the matter of an impossibility theorem, only of an *argument*. It is effective only if the burden of proof falls upon the optimizing viewpoint. This makes the whole paper a dialectical exercise, rather than a demonstration. On balance, I hope to leave the reader with an assymetric impression: that the view that optimization implies rationality is less plausible than the contrary view.[37]

To establish that optimization sometimes takes the agent away from rationality, or the social scientist from rational choice theory, would have serious methodological consequences. In the absence of suitable restrictions, the optimizing model could not take advantage any more of such standard construals as Weber's 'ideal-types', Popper's 'situational logic', or Davidson's 'principle of charity'. But when it comes to methodology, other considerations influence the social scientist's judgement. In particular, the analogies between the principle of least action in physics and the maximization hypothesis in microeconomics have been repeatedly stressed, and they were indeed part of some of the founders' heuristics (at least, Pareto and Samuelson). It is possible to argue for the latter in the way done for the former, quite independently of rationality considerations. To wit: maximizing theories derive significant benefits from their generalizability, mathematical simplicity, elegance, and heuristic fruitfulness.[38] Maximizing theories, not only of economists and biologists, but even of some physicists, are thought to be difficult to test, but even this handicap has sometimes be turned into an advantage in the name of the necessary continuity of research programmes. To evaluate these and other related justifications was not part of the normative assessment made here.

REFERENCES

Anand, P.

1987 "Are the Preferences Axioms Really Rational?" *Theory and Decision,* vol. 23, p.189-214.

1993 "The Philosophy of Intransitivite Preference". *Economic Journal*, vol. 103, p.337-346.

Aristotle *The Nichomachean Ethics.*

Aumann, R. J.

[37] There is an interesting dialectical precedent in the philosophy of decision theory, which in some sense sets a standard for papers like the present one - MacClennen's "Sure-thing doubts" (1983). It confronts technical principles (here, von Neumann independence and the "sure-thing principle") with pretheoretic concepts of rationality. McClennen's discussion is not entirely conclusive but oriented nonetheless, and it leaves the reader on the firm impression that one side of the argument is stronger than the other.

[38] Interestingly, the decision theorist Schoemaker's (1991) wide survey of optimization appears to eventually favour a defence in terms physical and biological analogies.

1962 "Utility Theory Without the Completeness Axiom". *Econometrica*, vol. 30, p. 445-462.

1976 "Agreeing to Disagree". Annals of Mathematical Statistics, vol. 4, p.1236-1239.

Bar-Hittel, M. and A. Margalit

1988 "How Vicious Are Cycles of Intransitive Choice?", *Theory and Decision,* vol. 24, p. 119-145.

Bernheim, D.B.

1984 "Rationalizable Strategic Behavior", *Econometrica*, vol.52, p. 1007-1028.

1986 "Axiomatic Characterizations of Rational Choice in Strategic Environments", *Scandinavian Journal of Economics*, vol.88, p.473-488.

Blyth, C. R.

1972 "Some Probability Paradoxes in Choice from Among Random Alternatives". *Journal of the American Statistical Association,* vol. 67, p. 366-373; and "Rejoinder", p. 379-381.

Campbell, D.

1978 "Realization of Choice Functions". *Econometrica*, vol. 46, p. 171-180.

Davidson, D., J. C. C. McKinsey and P. Suppes

1955 "Outline of a Formal Theory of Value, I". *Philosophy of Science,* vol. 22, p. 140-160.

Debreu, G.

1959 *Theory of Value*, Cowles Foundation Monograph, New Haven, Yale University Press.

Fishburn, P. C.

1970a *Utility Theory for Decision Making*. New York, Wiley.

1970b "The Irrationality of Transitivity in Social Choice". *Behavioral Science*, vol 15, p. 119-123.

1982 "Nontransitive Measurable Utility". *Journal of Mathematical Pschology* vol 26, p. 31-67.

1988 *Nonlinear Preference and Utility Theory* . Baltimore, Johns Hopkins University Press.

1991 "Nontransitive Preferences in Decision Theory". *Journal of Risk and Uncertainty*, vol. 4, p. 113-134.

Göttinger, H. W.

1982 "Computational Cost and Bounded Rationality". In W. Stegmüller, W. Balzer and W. Spohn, ed.,*Philosophy of Economics*. Berlin, Springer.

Harsanyi, J. C.

1976 *Essays on Ethics, Social Behavior and Scientific Explanation*. Dordrecht, D. Reidel.

Karni, E. and D. Schmeidler

1976 "Independence of Nonfeasible Alternatives and Independence of Non-optimal Alternatives". *Journal of Economic Theory,* vol. 12, p. 488-493.

Lindley, D. V., I. J. Good, R. L. Winckler and J. W. Pratt

1972 "Comment". *Journal of the American Statistical Association*, vol. 67, p. 373-379

Lipman, B.

1991 "How to Decide How to Decide How to...: Modeling Limited Rationality". *Econometrica*, vol. 59, p. 1105-1125.

Lismont, L. and P. Mongin

1994 "On the Logic of Common Belief and Common Knowledge". *Theory and Decision*, vol. 37, p.75-106.

Loomes, G. and R. Sugden

1982 "Regret Theory", *Economic Journal*, vol.92, p.805-824.

Luce, R. D.

1956 "Semi-order and a Theory of Utility Discrimination". *Econometrica*, vol. 24, p. 178-191.

Luce, R. D. and H. Raiffa

1957 *Games and Decisions*. New York, Wiley.

McClennen, E. F.

1983 "Sure-Thing Doubts". In B. P. Stigum and F. Wenstop, ed., *Foundations of Utility and Risk Theory with Applications*. Dordrecht, D. Reidel, p. 117-136.

1990 *Rationality and Dynamic Choice*. Cambridge, Cambridge University Press.

Malinvaud, E.

1971 *Leçons de théorie microéconomique,* Paris, Dunod.

Marschak, J.

1950 "Rational Behavior, Uncertain Prospects, and Measurable Utility", *Econometrica*, vol.18, p.111-141.

May, K. O.

1954 "Intransitivity, Utility and the Aggregation of Preference Patterns". *Econometrica*, vol. 22, p. 1-13.

Mongin, P.

1984 "Modèle rationnel ou modèle économique de la rationalité?". *Revue économique*, vol. 35, p. 9-64.

1986 "Simon, Stigler et les théories de la rationalité limitée". *Information sur les sciences sociales / Social Science Information*, vol. 25, p. 555-606.

Mongin,P. and B. Walliser

1988 "Infinite Regressions in the Optimizing Theory of Decision". In B. Munier, ed., *Decision, Risk and Rationality*. Dordrecht, D. Reidel, p. 435-457.

Plott, C. R.

1973 "Path Independence, Rationality, and Social Choice". *Econometrica*, vol. 41, p. 1075-1091.

Popper, K. R.

1967 "La rationalité et le statut du principe de rationalité". In E. M. Claassen, ed., *Les fondements philosophiques des systèmes économiques*. Paris, Payot, p. 142-150.

Rabinowicz, W.

1995 "To Have One's Cake and Eat It, Too", *Journal of Philosophy*, p.586-620.

Rescher, N.

1982 "Choice Without Preference : A Study of the History and the Logic of the Problem of 'Buridan's Ass'". In *Essays in Philosophical Analysis,* Lanham, University Press of America, chap. 5

Ryle, G.

1949 *The Concept of Mind*. London, Hutchinson.

Samuelson, P. A.

1938 "A Note on the Pure Theory of Consumer Behavior", *Economica*, vol.5, p. 61-71.

Schick, F.

1986 "Dutch Bookies and Money Pumps", *Journal of Philosophy*, vol.83, p.112-119.

Schoemaker, P. J. H.

1991 "The Quest for Optimality : A Positive Heuristic of Science?" *Behavioral and Brain Sciences,* vol. 14, p 205-215; followed by comments by other authors, p. 215-237; and "Author's Response", p. 237-240.

Schwartz, T.

1972 "Rationality and the Myth of the Maximum". *Noûs*, vol. 6, p. 97-117.

1986 *The Logic of Collective Choice.* New York, Columbia University Press.

Sen, A.

1970 *Collective Choice and Social Welfare*. San Francisco, Holden Day.

1971 "Choice Functions and Revealed Preference". *Review of Economic Studies,* vol. 38, p. 307-317; reprinted in A. Sen, 1982, chap. 1.

1973 "Behaviour and the Concept of Preference". *Economica,* vol. 40, p. 241-259; reprinted in A. Sen, 1982, chap. 2.

1982 *Choice, Welfare and Measurement.* Oxford, Blackwell.

Simon, H. A.

1976 "From Substantive to Procedural Rationality". In S. J. Latsis, ed., *Method and Appraisal in Economic*s. Cambridge, Cambridge University Press. p. 129-148; reprinted in H. A. Simon 1983.

1983 *Models of Bounded Rationality.* Cambridge, MIT Press.

Slote, M.

1989 *Beyond Optimizing. A Study of Rational Choice.* Cambridge. Harvard University Press.

Sugden, R.

1985 "Why be consistent?", *Economica*, vol. 52, p.167-184.

Tullock, G.

1964 "The Irrationality of Intransitivity". *Oxford Economic Papers*, new series, vol. 16, p. 401-406.

Tversky, A.

1969 "Intransitivity of Preferences". *Psychological Review,* vol. 76, p. 31-48.

1972a "Choice by Elimination", *Journal of Mathematical Psychology*, 9, p. 341-347.

1972b "Elimination by Aspects. A theory of Choice", *Psychological Review*, 79, p.281-299.

Watkins, J. W.N.

1970 "Imperfect Rationality", in R. Borger and C. Cioffi, eds, *Explanation in the Behavioural Sciences*, Cambridge, Cambridge University Press.

Weber, M.

1949 *Max Weber on the Methodology of the Social Sciences*, tr. and ed. by E.A. Shils and H.A. Finch, The Free Press of Glencoe, Ill.

Winter, S.

1975 "Optimization and Evolution in the Theory of the Firm". In R. H. Day and T. Groves, ed., *Adaptive Economic Models*. New York, Academic Press, p. 73-118.