VALUING PROCESSES

Martin E. Sandbu

University of Pennsylvania

Conventional economic theory assumes that people care only about ultimate outcomes and are indifferent to the decision and allocation processes by which outcomes are brought about. Building on Sen (1997), I relax this assumption, and investigate the formal and philosophical issues that arise. I extend the formal apparatus of preference theory to analyse how processes may enter preferences, and investigate whether traditional invariance requirements like the Weak Axiom of Revealed Preference are still satisfied in this new setting. I show that it is, provided certain conditions of separability hold, and I discuss the plausibility of these conditions. Further, I argue that processes are often valued in a mode that diverges from the conventional modes of *instrumental* and *intrinsic/independent* valuation. I introduce the notion of *dependent non-instrumental valuation*, and show how processes could *depend* on their instrumental function for their value – making their value dependent – and yet *derive* their value from something else – making it non-instrumental. Dependent non-instrumental value, I argue, can be explained by symbolic and evidential relations between processes and outcomes.

1. INTRODUCTION

Amartya Sen has recently introduced into preference theory a distinction between *culmination outcomes* – 'that is, only final outcomes without taking

205

any note of the process of getting there' – and *comprehensive outcomes* – whose descriptions include 'the processes through which the culmination outcomes come about' (Sen 1999: 27). When this distinction is appreciated, Sen shows that the formal theory of choice and preference that is at the foundation of modern economics can be employed to investigate the way processes are valued, and how the role of processes in people's preferences modifies the standard analysis that is based on culmination outcomes only. The purpose of this essay is to highlight and pursue some challenges that the distinction raises.

In the first part of the paper, I address the formalization of preferences over comprehensive outcomes. Sen restricts his formal analysis to making preferences over culmination outcomes contingent on choice processes, although a more general formalization would be to make comprehensive outcomes themselves the objects of preference. I undertake this extension in Section 2, where I show that the general formalization encompasses Sen's as a special case, but also brings out useful distinctions between different ways choice processes may matter. In particular, it allows us to see that preferences over comprehensive outcomes *per se* are not incompatible with conventional axioms of choice invariance, such as the Weak Axiom of Revealed Preference (WARP). Instead, I show in Section 3 that what such axioms rule out is a form of *nonseparability* in the valuation of choice processes, and discuss the conditions under which it seems reasonable to impose separability.

The second half of the essay moves from formal concerns to the substantive content of preferences defined over comprehensive outcomes. In Section 4, I describe three modes of valuation. In addition to the standard modes of independent valuation and instrumental valuation, I suggest that actions and choice processes have value that is *dependent* without being *instrumental*. That is to say, their value depends on their standing in a certain relation to culmination outcomes, but is not reducible to their causal efficacy in bringing about valued culmination outcomes. Non-instrumental dependent valuation is puzzling, as it allows for preference patterns that seem self-defeating (Section 5). In Section 6, I show how non-instrumental dependent valuation can be explained through the symbolic function of choice acts. This symbolic function, I suggest, is also the reason why preferences over comprehensive outcomes tend so often to reflect social norms and moral principles. Section 7 concludes.

2. REFINING SEN

It may be useful to start by clarifying what is meant by 'preference' in this paper, since the formal apparatus of preference relations and choice functions lends itself to many interpretations. Formally speaking, 'preference' is a binary relation, which is usually assumed to be reflexive

and acyclic, and typically also complete and transitive.[1] The preference relation is frequently interpreted as ranking states by the amount of *well-being* a person enjoys in them, and it is also frequently used to describe a person's *choices* over states (as in the concept of 'revealed preference' – where preference is a ranking, assumed to be complete, revealed by choices, assumed to be optimal with respect to that ranking). For our investigation, it is natural to return to the most intuitive notion of preference – where 'preferring' something means to assess it as better, to consider it of greater value. The preference relation, then, can be read as 'at least as good/valuable as. . .'. This sense of preference is not entirely divorced from the other meanings. One reason to prefer one object to another is that it produces more well-being. Similarly, in normal situations, it is rational to act so as to promote more highly valued states of affairs, so choices will reflect preferences – but note that this correspondence is different from the 'revealed preference' approach, which takes choice as primitive and constructs a relation from it. In what follows, preference is the primitive concept, and I interpret the term as a *betterness ranking* or *value ranking* of objects of value (Sen 1997 calls this a 'reflective preference'). We shall consider the analytics of such value rankings and the choice functions generated by them.

There is a limited sense in which the notion of preference I am using here can be said to represent a person's value system. The sense is this: a preference relation can capture a person's complete view as to which actions or states of affairs are 'better' or 'superior' to other ones, 'all things considered' (that is, all things she deems relevant for assessing the relative value of actions or states of affairs). That is not to say that preference theory can capture her complete evaluative view *tout court* – what we might call her moral outlook. This is because a moral outlook also contains *reasons* for why some actions or states of affairs are better than others; and on these reasons, the binary relations that constitute formal preference theory are silent.

For example, a deontological view may hold that dishonesty is wrong in itself because it does not sufficiently respect the rational autonomy of others. Other things being equal, it makes sense to say that honest actions are morally 'superior' or 'better' than dishonest ones. This, of course, should not be taken to mean that dishonest actions are wrong *because* they are worse than honest ones; that (from the point of view of the deontological theory I am describing) would be getting things the wrong way round. The point is simply that any ethical outlook will *inter alia* entail some value ranking of actions or states of affairs, and that

---

[1] In the formal analysis in the appendix, I treat preferences as transitive but not necessarily complete.

ranking can be represented by formal preference theory, provided the right accommodations are made.

It is one such accommodation that is the subject of this essay. Most people's preferences – and most evaluative theories – are not defined exclusively over culmination outcomes. Typically, the *process* by which the outcomes come about enters preferences as an irreducibly separate factor. One important process concern, and the focus of Sen (1997), is the *act of choice* – it may matter who chooses an outcome, and what alternatives are available. But even in the absence of choice, there can be variations in process which are not a matter of indifference. Think, for example, of the infinity of probability distributions that can define a lottery – in none of them does anyone choose the outcome, but they need surely not be equally fair or equally advantageous and therefore some may be preferred to others. When I refer generically to 'choice processes', I shall understand that term also to include non-choice process factors.

Consider three examples of how choice processes may enter preferences (the first and the third are also discussed by Sen 1997):

Example 1: Choosing fruit
The host of a garden party hands a fruit tray to a guest, who can choose between a mango (m_1) and two apples (a_1, a_2). The guest in general prefers mangos to apples, but does not want to pick the last mango, and hence chooses an apple. Yet he might 'be very pleased if someone else were to 'force' that last mango' on him (Sen 1997: 747). That is, he prefers apples to mangos when choosing them himself, but he prefers mangos chosen by someone else to apples chosen by someone else. In this case, his preferences over the fruit vary depending on who makes the choice. Similarly, they could vary depending on the choice set: if there were two mangos in the tray, he might prefer one of the mangos.

Example 2: Winning fairly
A runner is considering whether to cheat in a race. He in general prefers winning to losing, and he knows that he can improve his chance of winning if he cheats (and there is no risk of being found out). Yet because he prefers winning fairly to winning unfairly, he prefers not to cheat.

Example 3: Voice and participation
Many people value 'having a say' in the decisions that affect them. Various forms of participation, such as voting, or being allowed to express one's opinion, are valued over and above their (often minimal) instrumental effects. Participatory decision processes, then, are preferred not just because they lead to preferred culmination outcomes, but partly in their own right. Thus, Lind and Tyler (1988) find that people report more satisfaction from 'voice' than can be attributed to its contribution to more satisfactory outcomes. Similarly, many people prefer to vote in elections rather than abstain even though the causal effect is negligible.

Clearly, the valuation of participatory processes is more complex than simple instrumental valuation.

These three examples illustrate different ways in which the choice process matters. 'Choosing fruit' reflects that the choice process can affect preferences over culmination outcomes. 'Winning fairly' illustrates that the process itself can be an object of preference – the runner sees the fairness of the race as something that is valuable independently of his chances of victory. 'Voice and participation' concerns actions whose value is not independent of their outcomes (voting in rigged elections does not have the value of voting in clean elections, on which more below) yet cannot be reduced to the value of the consequences. Throughout this essay, I will probe these examples to investigate different aspects of how choice processes are valued.

'Choosing fruit' informs Sen's formal analysis. He models preferences with *chooser-* and *menu-dependent preference relations*. In Sen's framework, preference relations are defined over culmination outcomes, as in conventional preference theory, but the ordering of culmination outcomes depends on the identity of the chooser and the set he or she is choosing from. Formally, preference relations are denoted by $R_i^{j,S}$, and culmination outcomes by x, y *etc*. In this formalization, i indexes the person whose preferences are being studied, j the person choosing the culmination outcome and S the set of culmination outcomes from which the choice is made.

As Sen shows, his formalization accommodates preferences such as those in Example 1. If we label the choice sets $S = \{m_1, a_1, a_2\}$ and $T = \{m_1, m_2, a_1, a_2\}$, and index the guest and the host by g and h, respectively, then the preference pattern can be stated as follows:

$$a_k R_g^{g,S} m_1, \quad \forall k \in \{1, 2\}$$

(the guest prefers either apple to the single mango when choosing from S)

$$m_1 R_g^{h,S} a_k, \quad \forall k \in \{1, 2\}$$

(he prefers the single mango to either apple when the host chooses from S)
and

$$m_l R_g^{g,T} a_k, \quad \forall k, l \in \{1, 2\}$$

(he prefers either mango to either apple when choosing from T)

This formalization, however, cannot represent a preference of the sort illustrated in 'Winning fairly', where the process itself (a fair race or a cheating race) is an object of preference. Surprisingly, Sen does not offer a formalization of preferences defined directly over comprehensive

outcomes rather than culmination outcomes.[2] The rest of this section aims to show how the distinction between culmination and comprehensive outcomes can be more richly explored with such a formalization instead of Sen's.

Denote a comprehensive outcome by (x,a), where the first argument is the constituent culmination outcome, and the second argument a description of the process that brings the culmination outcome about. We may then use the normal notation of preference as a reflexive and transitive binary relation R, so that $(x,a)R(y,b)$ is to be read 'culmination outcome x, brought about through process a, is preferred to culmination outcome y, brought about through process b.' As is customary, we shall denote by P and I the asymmetric (strict preference) and symmetric (indifference) factors of R, respectively. There is no presumption here that R is necessarily complete. As I shall discuss presently, the degree of completeness of R is interestingly related to the ways in which choice processes matter.

It is easy to see that this more general formalization can encompass the one proposed by Sen. Note first that the general description $a,b\ldots$ of the choice processes in the present notation can index the identity of the chooser and the menu of choice, as well as any other relevant aspect of the process. Then the relation R induces Sen's process-specific preference relations over culmination outcomes as follows.

For all a, x, y such that (x,a) and (y,a) feasible:

$$(x, a)R(y, a) \text{ if and only if } x R^a y$$

where a indexes the chooser and the set of alternatives from which x or y are picked (and where we have omitted the index for the 'owner' of the preferences). The formulation in terms of R replicates all the pairwise rankings given by R^a, and may also rank pairs of comprehensive outcomes that differ in their process component (whether or not they also differ in their culmination outcome component). If the relation R is undefined for all such pairs, then the two formulations are isomorphic. If it is not, however, then R has the advantage that it can represent such preferences as those of the runner in 'Winning fairly'. In that example, the runner strictly prefers realizing some culmination outcome (winning) in certain ways (by competing fairly) rather than others (by cheating). Denoting the culmination outcome of winning by w and the two 'processes' – the fair race and the cheating race – by f and c, his preference is:

$$(w, f)P(w, c)$$

---

[2] His own examples suggest variations that need to be modelled with a more general approach. In 'Choosing fruit', for instance, it is natural to understand the guest as having a strict preference for getting the mango through the host's choice over picking the mango himself. Such a preference ranking of outcomes brought about by *different* processes is not compatible with Sen's formalization.

Examples 1 and 2 show that there are two different ways in which preferences could be sensitive to the choice process. In 'Choosing fruit', the process matters because the preference relation is not *separable* in the process component of the comprehensive outcomes it ranks.

Definition: Process-separability
R is *process-separable* if, for all culmination outcomes x,y and choice processes a, b:

  $(x,a)R(y,a)$ if and only if $(x,b)R(y,b)$

When the relevant aspects of the process are the identity of the chooser and the choice set, process-separability is equivalent to what Sen calls *chooser*- and *menu-independence*.

  In 'Winning fairly', the process matters because the preference relation is not *indifferent* to how a culmination outcome is produced.

Definition: Process-indifference
R is *process-indifferent* if, for all culmination outcomes x and choice processes a, b:

  $(x,a)R(x,b)$

  Conventional economic theory implicitly assumes that preferences are both process-separable and process-indifferent. In Sen's formalization, however, process-indifference is undefined, since preferences there are only defined for a *given* choice process. This is not a shortcoming insofar as Sen's focus is on preferences that violate the common axioms of 'consistency of choice', and as I suggest in the next section, what these formal axioms rule out is in fact process-separability, not process-indifference. Yet once we recognize that people may have preferences defined over comprehensive outcomes, surely we should develop formalizations that also allow such preferences as the runner's in 'Winning fairly'.

  By transitivity, process-indifference implies process-separability (a short proof is given in the appendix), but the converse does not hold: preferences can be process-separable without being process-indifferent. This can be shown with the aid of a few counterexamples, which illustrate interesting differences between the ways in which process-indifference can fail to hold. Consider first the case of a *strict* preference for reaching an outcome through one process rather than another – for example the runner's $(w,f)P(w,c)$. Such a process-nonindifferent preference will be process-separable if the ranking of comprehensive outcomes consists of a 'chain' of identical orderings of culmination outcomes, each conditional on a given choice process. The chained conditional orderings may or may not overlap. Consider the runner who prefers fair races to cheating, but who would rather win in a cheating race than lose (we denote this culmination

outcome by *l*) in a fair one:

$$(w, f)P(w, c)P(l, f)P(l, c)$$

[*Overlapping identical rankings of culmination outcomes (OIR)*]

In this case, *R* is process-separable (the runner prefers winning to losing, for any given process) but not process-indifferent. Similarly, process-separability but not process-indifference would hold in the case of the runner who would rather lose a fair race than win through cheating:

$$(w, f)P(l, f)P(w, c)P(l, c)$$

[*Juxtaposed identical rankings of culmination outcomes (JIR)*]

Here the chained orderings are juxtaposed, and do not overlap.

The second case where process-indifference may fail is when *R* is incomplete, so that (some) comprehensive outcomes consisting of different choice processes are unranked *vis-à-vis* each other. For example, the runner may prefer winning to losing, no matter whether the race is fair, but not have any preference comparing different types of races:

$$(w, f)P(l, f) \text{ and } (w, c)P(l, c),$$

but no other pairs of comprehensive outcomes are ranked

[*Disjoint identical rankings of culmination outcomes (DIR)*]

This preference pattern is process-separable because, given the process, the runner always prefers winning to losing. But it is not process-indifferent, since (*w,c*) and (*w,f*) are not ranked *vis-à-vis* each other and nor are (*l,c*) and (*l,f*).

The case of strict preference among processes is substantively the more interesting case of nonindifference. It is undeniably an important feature of many people's preferences; so is process-nonseparability. Many examples of these phenomena reflect a concern for norms and principles, and often express deeply internalized values. I return below to the relationship between preferences over choice processes and commitment to principles below. For now it suffices to say that it seems desirable to allow for such preferences in social analysis, and therefore to extend Sen's formalization of preference sensitivity to choice processes along the lines followed here.

3. SEPARABILITY

Sen (1997) presents several theorems on the relationship between the process-sensitivity of *preference* and the process-sensitivity of *choice*. In particular, he shows how a maximiser who has menu-dependent preferences will violate most of the common axioms of choice 'consistency', such as the Weak Axiom of Revealed Preference (WARP). Since his analysis is limited to just one kind of sensitivity to process (nonseparability), it is

useful to ask how his discussion can be extended to the richer structure introduced in the previous section. What, in particular, can we say about choice over different processes?

First, we note that there are many choice processes over which we cannot logically make a choice. Consider a different variant of 'Choosing fruit' in which another guest has awoken my romantic interest, and I would prefer to be given the mango by her rather than by the host. Yet being an incorrigible mango lover, I would prefer being given the mango by the host to being given the apple by my romantic interest. But this is not a preference I can possibly express by choice, since I cannot control what each person will in fact choose. (I am assuming here that my preference is for my romantic interest to *freely* choose to give me the mango – if I could control her 'choice', that would no longer be the comprehensive outcome I most prefer.) The same is true for menu-dependent alternatives – it is logically impossible, for instance, to face a choice between 'a mango when only one mango and two apples are available' and 'a mango when two mangos and two apples are available'.[3] More generally, because processes and culmination outcomes are causally linked, there could be comprehensive outcomes that are logically possible, and over which I may have strict preferences, but which are causally impossible (or impossible to guarantee), so that I cannot face a choice between them. Some strict preferences over choice processes are therefore 'non-practical preferences' (Broome 1993).

Surely, however, there are *some* preferences over different choice processes (comprehensive outcomes with different process components) that are 'practical'. How does standard axiomatic choice theory apply to agents who act[4] on such preferences? Sen (1997) in his investigation of chooser- and menu-dependent preferences identifies conditions under which they generate choice functions that violate standard axioms of choice invariance.[5] We can now extend his analysis to the more general formal

---

[3] Of course, a person could be faced with a choice between being given two different choice sets to choose from. But that would be a choice between 'a choice of one piece of fruit among one mango and two apples' and 'a choice of one piece of fruit among two mangos and two apples', which is a different choice to the one described above. See Broome (1993) for a longer argument to the same effect.

[4] I shall take acting on a certain preference to mean optimizing or maximizing behavior with respect to that preference. An optimizer chooses the available option(s) that is (are) weakly preferred to all others; a maximiser chooses the available option(s) which no other option is strictly preferred to. The difference is important when preferences are incomplete; see Sen (1997) and the appendix.

[5] We note that Sen's menu-dependence is equivalent to the notion of process-nonseparability when the choice process is fully described by the set of alternatives. Whenever *menu*-dependence generates 'inconsistent' choices, therefore, *process*-nonseparability does so *a fortiori*, since it can change the relative ranking of two culmination outcomes even when they are picked from the *same* choice set (if some other relevant process aspect changes).

structure presented here, focusing on WARP as the best known and most widely used of these axioms. WARP requires that if from some choice set which contains alternatives x and y, x is chosen, then whenever x and y are available, if y is chosen, x must also be chosen.[6] The idea is that the first choice 'reveals' x as 'at least as good as' y, and that that all other choices should be consistent with that revealed weak preference.

It is straightforward to see that process-nonseparable preferences can generate choices that violate WARP. Consider an agent with complete preferences over comprehensive outcomes. If his preferences are not fully process-separable, then there exist some culmination outcomes x,y and some choice processes a,b such that $(x,a)R(y,a)$ but $(y,b)P(x,b)$. If the agent has to choose from the set $\{x,y\}$, he will choose x under process a, but choose y and reject x under process b; precisely what is ruled out by WARP. As the examples in the previous section show, there need not be anything odd about such choices, nor are they uncommon. Indeed the experimental literature of the past two decades has solidly demonstrated that choice reversals occur even within identical sets of culmination outcomes, and that they are systematically related to the choice process.[7] There is a large class of common choices, then, that violate WARP due to process-nonseparable preferences.

The point here is that WARP is not inimical to preferences over comprehensive outcomes *per se*; but that it requires some consistency of preferences across choice processes. The appendix gives formal derivations of the relationships between WARP and conditions on preferences. I mention here the results that hold in the 'normal case' of agents with complete and transitive preferences.[8] In the normal case, process-separability (in practical preferences) is a necessary but not a sufficient

Sen shows that menu-dependent preferences are compatible with a 'menu-independent choice function', defined as a choice function that can be rationalized by *some* menu-independent 'revealed' preference. This, in turn, means that choice functions generated by menu-dependent preferences need not violate two weak axioms of choice invariance: 'contraction consistency' or property α and 'expansion consistency' or property γ. If choice processes are described by more than just the choice menus, however, this result need no longer hold.

[6] I use the formulation given by Mas-Colell *et al.* (1995). Sen (1971) defines WARP slightly differently (calling the version I use the 'Weak Congruence Axiom') and gives choice- and preference-theoretic formalizations of other axioms of choice consistency. WARP was introduced by Samuelson (1938).

[7] See, for example, Prasnikar and Roth (1992), Blount (1995), Brandts and Sola (2001), Güth *et al.* (2001), Andreoni *et al.* (2002), Falk *et al.* (2003), and Sandbu (forthcoming).

[8] Most of the statements that follow hold up (but only for 'practical' preferences) also when preferences are incomplete, if choice behaviour is optimizing. In the case of incomplete preferences and maximizing choice behaviour, the results must be modified as shown in the appendix. For ease of exposition in Sections 4, 5, and 6, I continue to focus on process-separability and complete preferences there. Similar arguments can be constructed in terms of the stronger condition of Weak Inter-Process Consistency (defined below) for the case of incomplete preferences. See the appendix for details.

condition for WARP to hold. The preference pattern JIR shows why it is not sufficient: holding the process constant, the runner prefers winning to losing, but he would chose losing a fair race over winning an unfair race. Although these preferences are process-separable, they generate choices that violate WARP.

Recalling the difference between process-separability and process-*indifference*, we may ask whether WARP also requires process-indifference. It turns out that WARP does *not* require process-indifference. The preference pattern labelled OIR in Section 2 shows why. The runner's preferences (he always prefers winning to losing) generate identical choices over culmination outcomes no matter what the process is, and so it must satisfy the standard axioms of choice invariance – yet the preferences in OIR are not process-indifferent. Process-indifference is, however, a *sufficient* condition for WARP to be satisfied.

Indeed in the normal case, the requirement that choices satisfy WARP is exactly equivalent to a condition on (practical) preferences that is weaker than process-indifference, but stronger than process-separability. The necessary and sufficient condition for WARP to be satisfied is:

Definition: Weak Inter-Process Consistency (*WIPC*). For all culmination outcomes x,y and choice processes a, b, c, d (not necessarily distinct):

If $(x,a)P(y,b)$, then not–$[(y,c)R(x,d)]$

which is fulfilled when identical subrankings of culmination outcomes conditional on a process overlap so that same culmination outcomes are 'stacked' next to each other in the overall preference order (as in OIR). This condition is 'in between' process-indifference and process-separability in that it is entailed by the former and entails the latter.

In the whole discussion up to now, we have treated the axioms of choice invariance as imposed on choices between *culmination* outcomes. One might retort that axioms such as WARP are intended to be imposed on the 'real' objects of preference, which are comprehensive outcomes, not culmination outcomes. But that would be a mistake. The axioms are only useful insofar as they allow for the possibility that certain alternatives are seen as 'the same' even if they figure in different situations (in particular in different choice sets). WARP and other axioms are rendered vacuous if they are only 'imposed' on choices between objects so individuated that the axioms cannot be violated (Broome 1993; Sen 1993, 1997). Echoing Sen (1997), '[t]he kinds of influences considered here suggest the need for limiting the domain of applicability' of the axioms, not for applying the axioms to ever more finely individuated objects.

This poses the question of *how* to 'limit the domain of applicability' of the axioms. When should we impose them, and when should they be jettisoned? And what reasons can we give for the decision? It will not do simply to say that people as a matter of fact have preferences

that violate WARP. The appeal of the axioms of choice invariance is not just their usefulness for modelling behaviour, but their plausibility as characteristics of rational choice. Indeed WARP is often taken as an axiom of *rationality*. How to 'limit its applicability' ought therefore to be supported by a normative argument, not just an empirical observation that people do not behave that way. Whenever we think WARP should not be applied we should be able to explain why it is not reasonable to impose its requirements on the choices of rational persons. This is best done by first explaining why WARP may seem reasonable in the first place, or more precisely and in light of the argument up until now, why it may seem reasonable to impose process-separability on preferences. Why and when is process-nonseparability objectionable?

Consider what is often taken to be a paradigmatic case of irrational preferences: the sunk cost fallacy.[9] I submit that what makes caring about sunk costs a fallacy is that it involves an objectionable type of nonseparability. Take an agent who attributes value to one single thing, namely monetary profits.[10] Now the ranking of currently available actions by profitability depends only on their ranking by gains and losses to be made in the future as a result of the current choice. More specifically, if starting the new project x will bring in more profits from today than will continuing the ongoing project y, then regardless of the losses and profits I have made up until now, switching to project x will lead to a correspondingly higher level of *total* accumulated profits than will sticking with project y.[11] Since ultimate profits are just the sum of the gains or losses made 'along the way', they are by their nature process-separable. Let us call them *inherently* process-separable. If now the path to the outcome – described by the net irrecoverable losses they have caused me so far, i.e. the sunk costs – influences my preference ranking of the culmination outcomes that are currently available to me, then my preferences are *non*separable in the process. For someone who finds value *only* in profits (such as the stylized profit-maximizing firm in standard producer theory), this is

---

[9]   The thought that it is irrational to let one's choices be determined by sunk costs is cemented in folk wisdom, which exhorts us to let 'bygones be bygones', or 'not to cry over spilt milk.' That it is irrational, of course, does not mean that the tendency to care about sunk costs is not useful. 'We can knowingly employ our tendency to take sunk costs seriously as a means of increasing our future rewards. If this tendency is irrational, it can be rationally utilized to check and overcome another irrationality' (Nozick 1993). If I think it would be good for me to attend a number of classical music concerts this year, and know that on the night of a concert I will not be motivated to leave the comfort of my house, I can buy tickets in advance, anticipating that I will not want to have wasted the money and so will overcome the temptation to stay at home.

[10] The text discusses the case of a profit-maximizing agent, but the fallacy applies more generally to other denominations of net benefits or gains, so long as they are separable in past losses like monetary profits.

[11] This claim presupposes the empirical assumption that the profits made or lost in a past project do not affect the determinants of future projects' profitability *differentially*.

irrational. A profit-maximiser whose preference over currently available projects is in the least sensitive to the amount of money sunk in them commits a fallacy by failing to adhere to the inherent separability of ultimate profits, which by hypothesis are the source of any value the projects may have. The appropriate generalization of the sunk cost fallacy is this: It is irrational to violate process-separability when (all) the objects that have value are inherently separable with respect to choice processes.

An unconditional insistence on process-separability would seem to presuppose that *only* inherently process-separable objects could carry value. That is an implication of imposing process-separability that we cannot tolerate. We should limit the applicability of axioms that entail process-separability, therefore, when it is reasonable to allow as sources or carriers of value things that are not inherently process-separable. Sensitivity to sunk costs, to return to our example, is a fallacy when we *assume* a (unique) object of value that happens to be inherently separable in sunk costs. Ultimate profit is indeed separable in past losses; in fact, it is separable with respect to profit streams within any single time period. So *if* we assume from the outset that only profits have value, it is irrational for preferences over profit-making projects to be nonseparable in sunk costs. If we relax that assumption, however, we see that sunk cost sensitivity could achieve *other* valuable things, even as it involves 'throwing good money after bad' – and these other valuable things might not be inherently separable in sunk costs.[12]

It would seem straightforward to show that there are things it is reasonable to value which are not inherently process-separable. That, after all, is what all the examples we have discussed attempt to do. However, things are not as simple as they appear at first glance. Nonseparable valuation of processes, I shall now argue, is difficult to accommodate in terms of the two standard ways something can be valued – independently or instrumentally – and must be understood as a *third* mode of valuation.

4. THREE WAYS OF VALUING CHOICE PROCESSES

Preferring one thing over another is to attribute more value to it, to judge it to be better in some relevant evaluative dimension. Preferences

---

[12] It may, for example, meet a concern for integrity. If I have put many resources and effort into a project, I may feel a special responsibility to 'see the project through', even if it would no longer be worthwhile for someone who had not invested the same effort (and even if I acknowledge this). See Williams (1973) for an insightful discussion of personal integrity. Robert Nozick points out the usefulness of such 'path-dependent' preferences, stressing that they are important components of our personal identity: 'We do *not* treat our past commitments to others as of no account except insofar as they affect our future returns. . . and we do *not* treat the past efforts we have devoted to ongoing projects of work or of life as of no account (except insofar as this makes their continuance more likely to bring benefits than other freshly started projects would). Such projects help to define our sense of ourselves and of our lives' (Nozick 1993: 22, original italics).

over comprehensive outcomes attribute value to processes, not just to culmination outcomes. Now there are many reasons why processes may figure in the valuation of states of affairs. Sen (1997) mentions four possible accounts in his discussion of 'Choosing fruit': Reputation and indirect effects (picking the last mango will make people hostile to me in the future); social commitment and moral imperatives (it is rude to take the last mango); direct welfare effects (the way people look at me when I take the last mango makes me uncomfortable); and conventional rule-following (taking the last mango is just not done). Clearly, one of the main differences between these explanations is whether the choice process (the act of choice) is valued in its own right or according to its connection with valuable culmination outcomes. The first explanation, for instance, is straightforwardly instrumental: choice acts have (more or less) value because they have indirect effects on (more or less valuable) outcomes.

The formal apparatus we have explored in the previous sections can handle all of these explanations, so the nuances are inconsequential for purely formal purposes. The source of the processes' values, however, determines what the formalizations are useful for, in at least two senses. The first concerns how we *interpret* preferences over comprehensive outcomes and violations of process-indifference or process-separability. If the value of processes is merely instrumental, then formalizations like Sen's or the one proposed here should be seen as no more than 'reduced-form' simplifications or shorthand renditions of preferences that are 'really' defined over culmination outcomes after all. This is not necessarily a shortcoming: simplifications are always necessary, and proceeding *as if* people value processes directly may be fruitful. Still, the explanatory status of formal preference relations is different if they are seen as shorthand rather than representing people's actual preferences. A second concern is that their *normative* status is different. When a normative theory of social choice that gives some weight to satisfying the preferences of individuals must determine how to value social decision processes, it will surely matter whether individuals value them in their own right or merely as instruments to valuable outcomes. For these reasons, it is important to investigate what I shall call the *mode of valuation* that applies to choice processes.

In the preceding paragraphs, I have not challenged the conventional partition of modes of valuation into intrinsic and instrumental value. This dichotomy, however, is too impoverished for a full understanding of the ways processes (and indeed outcomes) may be valued. To see the point, we may start by considering Raz's (1986) three-fold distinction between 'instrumental', 'intrinsic', and 'ultimate' value. Raz treats instrumental and intrinsic value in the conventional fashion: 'Having intrinsic value is being valuable even apart from one's instrumental value.' But, he goes on to say, 'not everything which is intrinsically valuable is also of ultimate value.' (Raz 1986: 177). The relationship between a man and his dog, according to

Raz, could be intrinsically valuable, not just instrumentally valuable: 'Its value is not just that of a cause of a feeling of security and comfort in the man. Such feelings may be produced by tranquillizers. The relationship is not valued just as a tranquillizer. Its value is in its being a constitutive part of a valuable form of life. Those who share these views believe that the existence of the dog is intrinsically valuable.' Still, if the relationship made the man miserable, it would not have the same value. So while the dog is not valued *according to* its contribution to the man's well-being (and so the value is not instrumental), its value *depends*[13] on the man's well-being.

The lesson to take from Raz's two dichotomies is that the value of something can depend on something else without the valuation being instrumental. Let us call value that does not depend on anything else *independent* value (Raz's 'ultimate' value). Value that is not independent is *dependent*, but it could be either *non-instrumental* ('intrinsic' in Raz's terminology) or *instrumental*. Thus we have three modes of valuation: independent value, instrumental (and dependent) value, and dependent non-instrumental value. I shall suggest that many of the most interesting examples of preferences over comprehensive outcomes value processes in the third mode.

Let us first, however, examine process-indifference and process-separability in the light of the first two modes of valuation. An *independently* valued object has value *in vacuo*, by itself, without any connection with valued objects external to itself. An action or choice process, even while being a cause of valued ends, may itself be a valued end. The value attributed to it when considered in the absence of any consequences it may have – its value *in vacuo* – is its independent value. Independent valuation, clearly, is one reason why process-indifference may fail. The value of a comprehensive outcome – let us call this the 'comprehensive value' for short – reflects the independent values of its constituent choice process and culmination outcome. So the independent values of different choice processes each contribute something to the comprehensive values, and may therefore influence the ranking of comprehensive outcomes that contain them, even with an unchanged culmination outcome.

The two 'chained' preference patterns we proposed to formalize 'Winning fairly' in Section 2 – OIR and JIR – illustrate independent valuation of processes. They involve identical subrankings of culmination outcomes (winning or losing) one of which is above the other in the overall ranking over comprehensive outcomes, depending on the process.

---

[13] Raz says 'the intrinsic value of the dog is not ultimate for it *derives* from the dog's contribution to the well-being of the man, [which] is here taken as the ultimate value' (my italics). I prefer to say that the value 'depends' on the dog's contribution to the well-being of the man, for a reason that will be elaborated on below: a valued object could depend for its value on one thing, but derive its value from another.

A natural way of interpreting this is that the more valuable process 'lifts' the ordering of comprehensive outcomes that contain it above the ordering of comprehensive outcomes that contain the less valued process. If the independent value of a fair race is only slightly more than that of an unfair race, then it does not outweigh the lower independent value of losing relative to winning, and the overall ranking will be overlapping (OIR). If it is sufficiently higher, however, it can outweigh the disvalue of losing, so that the whole subranking involving the fair race will be lifted above the one involving the unfair race. We then get preferences that are juxtaposed (JIR) rather than overlapping. This shows how independent valuation can account for process-nonindifference.

It does not seem, however, that independent valuation of processes can explain process-nonseparability. For suppose that the entire value of choice processes and of culmination outcomes is independent value. The comprehensive value is then an aggregate of the two independent values. For this aggregation to exhibit nonseparability, either the contribution of the choice process's value to the comprehensive value depends on which culmination outcome it is paired with, or *vice versa*. But this contradicts the hypothesis that both processes and culmination outcomes have only independent value. So we cannot attribute process-separability to independent valuation of processes.

In fact, many valued processes do not have independent value. They are not valued *in vacuo*, independently of their outcomes and consequences. On the contrary, they are valued *qua* paths to outcomes. Consider example 3, the preference for participation in decision processes. We often prefer culmination outcomes to come about through processes that afford us 'voice' – *i.e.* an opportunity to express our concerns in decision processes that affect us. Yet the value of voice is not independent. We put no value on voice if it is clear to us that it has no causal relationship with the outcomes whatsoever. What is valued is voice, but only voice that is part of the causal path to the culmination outcome. The same seems true for other participatory decision procedures, such as voting. The value of the act of voting is contingent on its instrumental function, and we do not attribute any value to participating in rigged elections.

There are some processes whose meaning is so bound up with the goal they aim for (their 'purpose') that we do not even consider them the same processes when we contemplate them *in vacuo*. One such process is the journeying involved in pilgrimage. The proximate aim of a pilgrimage is to arrive at a holy site and carry out rites of religious devotion (the ultimate aim presumably being to fulfil one's duty or be brought closer to God). The length and arduousness of the journey towards that goal, although making it less likely or more costly to achieve the valuable culmination outcome, may *enhance* the value of the comprehensive outcome. The value of the means (the journey) may even exceed the value of the culmination

outcome (the arrival and worship at the holy site) or compensate for a failure to achieve the latter. Clearly, the journey has value – but this value is not independent of the outcome. A pilgrimage must have a destination. Random itineration with no determined goal could never be a pilgrimage; the very notion of an aimless pilgrimage is unintelligible (except perhaps in a metaphorical sense). The value of the journey-as-pilgrimage, then, depends on its intended consequence – the intentionality is part of what makes it valuable.[14]

In each of these examples, the value of the process is contingent on its connection with valuable outcomes: the influence of one's expression, the counting of one's vote, and the destination of the pilgrimage. Its value (or part of it) is dependent, not independent. What is the nature of this dependence? The conventional case of dependent valuation is *instrumental* valuation, in which the instrumentally valued object depends for its value and derives its value from the same thing; namely, its (expected) causal consequences.[15] The instrumental value of a process or action is nothing more or less than the expected value of its outcomes; we might say that the action 'borrows' the (expected) value. The 'borrowing' metaphor is an apt one, since it captures why an action that is *only* instrumentally valued cannot contribute anything to the value of the comprehensive outcome that includes it. Once the value of the culmination outcome has been taken into account, it would be double-counting to let the value of the choice process contribute anything more, since that value by hypothesis is just the 'borrowed' from the culmination outcome.[16] So if actions and processes had only instrumental value, preferences would have to be

---

[14] This example is discussed by Hirschman (2002[1982]: 88) who also mentions that modern sports fans practice their fandom in ways quite similar to mediaeval pilgrims – the further away their home team is playing, the greater the value in travelling to support it. See also Jon Elster (1983: Ch. III) for examples of 'states that are essentially by-products'. Certain benefits of certain activities (for example the satisfaction derived from political participation) are only psychologically available if they are *not* the aimed-for goal of the activity in question. As with pilgrimage or fandom, the value of the by-product may exceed the value of the aimed-for goal.

[15] Processes could have both independent and instrumental value at the same time. Consider a preference for honesty: people seem to prefer achieving the same culmination outcome through more rather than less honest actions. Such a preference is in part due to the instrumental effects of the actions: I expect to get a worse car from a dishonest car dealer than from an honest one. Most of us, however, have an independent dislike of deception – even when I do not think I could have got a better deal from a more honest car dealer, or from the same dealer had he told the truth, I still dislike being lied to. This is confirmed by experimental studies. Brandts and Charness (1999) find that subjects are willing to reduce their payoffs in order to punish other subjects who have lied to them, even when the lying did not affect their own payoffs in any way.

[16] More generally, a choice process could lead to any one of several possible culmination outcomes, each with some probability, and its instrumental value would be the *expected* value of its consequences. If the choice process is only instrumentally valuable, then the

process-indifferent (and by implication process-separable), and the choice axioms that entail process-separability would seem warranted.

This would be an incorrect account of the valuation of voice and participation. These processes are neither independently nor instrumentally valued. They depend for their value on their causal connection with the culmination outcomes they aim for. But their value cannot be *derived* from the outcomes they produc, at least not exclusively. Their value cannot be derived from them because it is not commensurate with the process's instrumental efficacy, which is often minimal, such as in the case of voting. Moreover, there exist alternatives to voting that are equally or more efficacious in carrying out the instrumental function, such as vote-trading, and yet people are not indifferent between voting and instrumentally equivalent vote-trading. (They might say, perhaps, that to consider the act of voting a commodity that can be traded is to miss what the value of voting consists in.) In such preferences, processes have dependent non-instrumental value. The existence of such preferences has been documented by social psychologists. Lind and Tyler (1988), for example, find the subjective satisfaction people report to derive from voice to be higher than what is attributable to the causal effect of voice in bringing about a more satisfactory outcome. However, if subjects believe that there is no effect at all – if inclusion of voice in the decision process is seen as a 'sham' – then the reported satisfaction derived from voice disappears and may even turn negative (relative to a no-voice benchmark).

It is this dependent, non-instrumental mode of valuation we need in order to account for process-nonseparability. This third mode of valuation, however, is puzzling. There is a tension between a process value depending on its causal relation with culmination outcomes, and yet being different from the causally expected value of those outcomes. The next section explores this tension.

5. SELF-DEFEATING VALUATION

The previous argument suggests that we can limit the application of WARP and similar axioms to cases where processes are valued either independently or instrumentally. Nonseparability should not be ruled out when the third mode of dependent non-instrumental valuation applies. But this mode is puzzling, at two levels. First, it seems mysterious that the value of an action or process could be dependent on its having a certain instrumental function, yet be different from its instrumental value. Where, we want to ask, does this 'extra' value come from? How could the presence of a certain process affect the value of a comprehensive outcome,

value of the comprehensive outcome that contains it must simply be the expected value of the lottery over possible culmination outcomes.

when that process's own value is contingent on the culmination outcomes it causes? Second, it seems to allow for a particularly disturbing type of nonseparability. Suppose the culmination outcome x is independently more valuable than culmination outcome y. Nonseparability in the valuation of choice processes means that there is some choice process a such that $(y,a)R(x,a)$. Because of choice process a, the comprehensive outcome comprising the independently *less* valuable culmination outcome is ranked above the comprehensive outcome comprising the independently *more* valuable culmination outcome. But if the value of a is dependent on its instrumental function rather than independently valuable, then this is a paradox. The value of the choice process is conditional on its instrumental relation with independently valuable culmination outcomes, and yet it causes an independently less valued culmination outcome to be ranked more highly when 'joined' to it in the comprehensive outcome. The paradox has a logical and a practical aspect. It would seem to be a logically inconsistent value system that makes the value of an action contingent on its instrumental effect and at the same time allows that value to outweigh the values of the culmination outcomes it leads to. Such a value system would also seem to be practically self-defeating when acted on. A person who maximizes or optimizes with respect to such rankings over comprehensive outcomes will fail to realize the (independently) most highly valued culmination outcomes, on which the value of the process component is assumed to be contingent.

Return to the example of voting. We have said that the knowledge, or at least the belief, that one's vote is *counted* is crucial for the act of voting to be valued. So the value of voting depends on its instrumental function. But its value is clearly not equal to its (typically minimal) instrumental efficacy. Most people are unimpressed by the fact that their vote has a miniscule chance of being pivotal, and even if they think the instrumental effect is non-negligible, they mostly choose not to engage in vote-trading.[17] So the value of the choice act of voting for one's preferred candidate is dependent but non-instrumental. Now if this value affects preferences in a nonseparable manner, then the voter may end up ranking

---

[17] The instrumental efficacy of my participation in the ballot is identical to that of convincing (or paying) somebody else to vote for my candidate who would not otherwise vote, and almost identical to that of convincing (or paying) somebody else not to vote who would otherwise vote for the opposing candidate. And if a vote for my candidate is more likely to be pivotal in a different constituency than my own, then the instrumental efficacy of engaging in vote-trading across constituencies is *greater* than that of voting for my own candidate: I can increase the probability of my preferred electoral outcome if I trade my vote against someone else's more pivotal vote for my candidate in another constituency. This was the reasoning behind the Internet-based vote-trading scheme whereby Greens in Florida would vote for Senator Albert Gore in return for Democrats in less pivotal states voting for Ralph Nader in the 2000 US presidential election.

the comprehensive outcome (my preferred candidate loses; I voted for my preferred candidate) above the comprehensive outcome (my preferred candidate wins; I traded my vote). The choice act, whose value is supposed to depend on its instrumentality in affecting the outcome of the election, makes the independently *less* valued culmination outcome come out *higher* in the ranking over comprehensive outcomes. This, I have argued, is logically contradictory and practically self-defeating.[18] If non-instrumental dependent value is to limit the normative reach of WARP, this contradiction has to be resolved.

6. EXPLAINING NON-INSTRUMENTAL DEPENDENT VALUE

When we place value on participation or voice, what we value is (in part) our *causal role* in producing culmination outcomes. Without that causal role, the process has no value. The puzzle described in the previous section is that *the value of the causal role* of the process is different from the process's *causal value*. So while the value of the process depends on its causal role, it cannot be derived (only) from its causal role. This means that we must distinguish between two ways in which the value of something 'comes from' something else. First, when an object is dependently valued, its value depends or is *contingent* on something other than itself. I shall refer to this as the *condition* of its dependent value. Second, a dependently valued object must *derive* its value from something other than itself. I shall use the term *source* of value for this relationship.

In the cases of independent or instrumental valuation, the condition and the source of value are in fact identical. An independently valued object depends on nothing else and derives its value from nothing else; it serves as the condition and source of its own value. An instrumentally valued object depends for its value on its consequences and derives its value from them; the condition and source of its value are its expected causal effects. When the dependent value of an action or process is non-instrumental, however, it must derive from something other than the consequences. The solution to the threat of contradiction or self-defeatingness therefore lies in explaining how actions like voice and participation, while depending on their instrumental function for their value, derive their value from something else than their consequences.

How does value derivation work? Instrumental valuation – the imputation of value from ends to means – is the standard case of value derivation. Through their causal/instrumental relation to valuable outcomes, actions have causal expected value. Instrumental valuation, in Robert Nozick's (1993) words, involves value 'travelling back' from outcomes to actions through this relation. The causal-instrumental relation, however, is not the only connection actions can have to outcomes. Raz's dog stands in

---

[18] As some of those voters mentioned in footnote 17 came to experience.

a *constitutive* relation with its owner's well-being. Nozick (1993) highlights *evidential* and *symbolic* relations – respectively the relation of 'indicating' that some state of affairs is or will be the case,[19] and the relation of 'symbolizing' or 'standing for' something.[20] Just like value can 'travel back' from outcomes to actions through a causal relation, Nozick proposes, the value of outcomes can also 'travel back' to actions that indicate or symbolize them through the evidential or symbolic relation. By indicating or symbolizing something valuable – in one word, by *representing* it – an action is endowed with 'evidential expected value' or 'symbolic value' over and above its instrumental or independent value. (Of course, the object that is symbolized or indicated could have *negative* value; the action would then be overall less valuable than its instrumental and intrinsic value would suggest.) These kinds of value are dependent – since they depend on the representative relation – but not instrumental – since their value is derived through evidential or symbolic relations, not a causal one.

To someone primarily interested in preference theory as a foundation for economic analysis, the evidential or symbolic value of action may at first sight seem esoteric and of limited relevance for the phenomena economists study. A few examples are enough to show that this would be a mistake. The classic example of evidential value is the Calvinist belief that earthly success was evidence of divine grace – a belief which endowed worldly riches and the actions conducive to them with more value than an otherwise ascetic value system could attribute to them either independently or instrumentally. Such value systems are often thought to have played a profound role in the process of capitalist economic development.[21]

A contemporary example, also a fundamental factor in many economic phenomena, is advertising, which associates the consumption of certain goods with certain desirable lifestyles or popular people. Advertising creates evidentially or symbolically derived value by making the consumption of certain goods mark one as a certain type of person or symbolize certain virtues.[22] It may even be that *any* valuable object can impute its value through evidential or symbolic relations (consider the 'sentimental value' of heirlooms or personal mementos); however, such

---

[19] An action 'indicates' or is evidence of a valuable outcome if the probability of the outcome, conditional on the action being performed, is higher than the probability conditional on non-performance, when the conditional probabilities do not reflect causal influence but inferred likelihood.

[20] The relation of *expressing* something could be seen as a variant or a subcategory of symbolizing. An action can express something to others; it can symbolize something both to others and to oneself.

[21] The locus classicus is Weber's *Protestant Ethic and the Spirit of Capitalism*.

[22] Such as when Coca-Cola encouraged consumers to drink its product with the slogan 'Be sociable' – drinking Coke indicated sociability. This example is cited by Alexander Schuessler in his interesting comparison of the parallel evolution of soft-drinks marketing and presidential election campaigns (Schuessler 2000: Ch. 5).

value derivation seems to play a particularly important role in the in the valuation of choice processes.[23]

We are now in a position to account for the dependent non-instrumental value of voice and participation. I argued above that their value is contingent on the choice act having some causal connection to relevant culmination outcomes. Now there is no reason why they could not at the same time derive symbolic or evidential value from something else than these outcomes. At the same time as being instrumental in achieving some valuable object (and thereby deriving instrumental value), a choice act can symbolize or indicate a *different* valuable object. Further, the choice act may derive *more* value (or disvalue) from the object it represents than from the effect it causes. The participatory acts of voice and voting have some (often minute) instrumental effect on culmination outcomes. Yet in their symbolic and evidential role, they represent something much broader: they represent our autonomy as subjects who can shape our reality, rather than objects whose lives merely happen to them (another way of putting this, perhaps, is that they represent the much larger instrumental value of our causal control over a host of *other* outcomes). These symbolic and evidential relations are what endow such actions with a value far greater than they could derive from their instrumental effects.

The possibility of value derivation through other relations than the causal-instrumental one explains why non-instrumental dependent value need not be self-defeating. In the previous section, we pointed to the possibility that such valuation could place an independently less valuable culmination outcome above a more valuable one in the ranking of comprehensive outcomes. The worry was that this seemed to be logically inconsistent with, or practically undermined by, the source of the dependent value. We can now address this worry by pointing to the targets of the symbolic and evidential relations, rather than the instrumental effects, as the source of the derived value. This account not only explains how the dependent value of a choice process can be different from its instrumental value. It also shows why that value is still contingent on the instrumental function of the process – that is, the instrumentality can be the *condition* of dependent value even if it is not its source. This is because the ability of a choice process to represent certain valuable objects may itself be contingent on its standing in a causal-instrumental relation to certain (other) culmination outcomes. Somewhat uncharitably to the symbolic or evidential relations, we may say that they can be *parasitic* on an instrumental relation. This is simply an empirical statement about the social meanings of actions and processes – what they in fact symbolize or represent – especially when what is symbolized is agency or autonomy. It is not surprising, for example, that voice and voting should fail to

---

[23] In his examination of the 'significance of choice', for example, Scanlon (1986: 178–80) points to the 'expressive' and 'symbolic' roles of choice.

symbolize or indicate autonomy when they lose their causal connection with the events they are supposed to influence (in rigged elections or 'sham' channels of voice). This explains why choice processes are typically assigned dependent rather than independent value.

The possibility of deriving value through representative relations also suggests why so many examples of preferences over comprehensive outcomes seem to involve moral principles or rule-following behaviour. Principles and rules have the function of grouping actions together in classes. 'By adopting a principle, we make one action stand for many others and thereby we change the utility or disutility of this particular action' (Nozick 1993: 18). Rules and principles, therefore, are particularly well suited to creating representative relations. Principles facilitate symbolic and evidential value derivation at two levels. At one level, a single action comes to represent the entire class of actions covered by the principle (it can both *symbolize* the class and increase the evidential probability – even if not the causal probability – that other actions from the class are performed). Thus, even when the specific action is instrumentally negligible, the other actions it represents may have much higher instrumental (or independent) value. The symbolic and evidential relations created by principles, therefore, may give value to what the classical utilitarians called the *tendency* of a class of actions, as opposed to the actual consequences of a single action.[24] At a deeper level, being the type of person who can act on principles (or on a particular principle) and conform to rules (or certain

[24] It is only when a class of actions has been defined that it is possible to talk of the effects an action *tends* to have. Since rules and principles mark actions as belonging to a certain class of actions (those covered by the rule or principle), they make it possible for an individual action to derive value from the instrumental tendency of its class. Note that my individual action could even represent *others* acting in a certain way, if the principle is a general one – so symbolic or evidential value gives a rationale to the common-sense moral question 'what if everyone did that?' (Nozick 1993 applies this insight to the rationality of cooperating in the Prisoner's Dilemma and being a 'one-boxer' in Newcomb's Problem.) Principled or rule-following behaviour, then, 'enriches' the instrumental value of the actions under the principle through the representative relations between them. This analysis addresses a puzzle in rule-utilitarianism, which can be represented as a preference ranking over comprehensive outcomes, where the culmination outcomes are sums total of utility and the process components are actions. For rule-utilitarianism to be different from act-utilitarianism the rule-utilitarian value ranking must fail to be process-indifferent – there must be instances where an optimific action should not be performed because the general practice of a rule which proscribes it is the optimific practice. But why should a utilitarian obey the rule in instances where it is known that violating it will produce more utility overall (Smart 1956)? Put differently, how can a utilitarian contribute any other value to an action than that action's own utilitarian consequences? Clearly, rule-utilitarianism cannot justify this by appeal to an independent (non-utilitarian) value of actions, since ultimately only utilities should matter in a theory that is recognizably utilitarian. The derivation of value through representative relations seems a more plausible account. Admittedly, the rule-utilitarian would still bear the burden of showing why it makes sense for someone with a utilitarian sensibility to accept that value is derived through other relations than the conventional instrumental one.

particular rules) may itself be valued. If so, then actions that comply with a principle will represent not just the class of actions covered by the principle, but the acting on principle itself.

In fact, whether or not a process-nonseparable pattern of preferences over comprehensive outcomes can be characterized as self-defeating may depend on whether the non-instrumentally valued actions are subsumed under a principle. Consider a pacifist who is deciding whether to participate in an anti-war demonstration that is likely to *raise* the risk of a war breaking out.[25] Being a pacifist, he presumably values the culmination outcome of no war more highly than the culmination outcome of a war, where this preference reflects the independent value he attributes to those culmination outcomes. But the choice here, by hypothesis, is between participating in the demonstration and increasing the likelihood of war (we make the admittedly unrealistic assumption that his participation makes a difference to the culmination outcome) and sitting at home, keeping the risk lower. It would seem that a *principled* pacifist may rationally prefer to participate because that action expresses or symbolizes his pacifist principle. His principles may commit him to fight war through vocal participation in anti-war efforts – a class of action that *tends* to make wars less likely even if this particular demonstration has the opposite effect. By instantiating this principle and representing the whole class of actions whose expected tendency is to promote peace, the act of demonstrating takes on sufficient value to outweigh the expected disvalue of making war more likely. For a *non-principled* pacifist, on the other hand – one who simply prefers culmination outcomes in which there is less war – such a preference over comprehensive outcomes would seem to be self-defeating, because the culmination outcome of lesser independent value (increased likelihood of war) resulting from his participation in the demonstration undermines any dependent value the choice act might at first have seemed to have.[26]

---

[25] Suppose the demonstration will make the foreign party in the conflict more resistant to the demands of the pacifist's government, which consists of non-pacifists who will take that as a reason to declare war.

[26] Our principled valuations are more profoundly bound up with our sense of who we are than other valuations. Unless they reflect blind rule-following, they tend to express what Bernard Williams calls 'commitments.' It is relatively easy to change one's 'mere' preferences, whereas altering one's commitments is a more radical transformation, frequently accompanied by reflective soul-searching or at least a retrospective acknowledgment of having become 'a different person.' On a formal level, the difference can be analysed as the difference between first-order preferences and second-order preferences or meta-preferences (Sen 1974). Principled value judgments will constitute higher-order preferences, while direct preferences that do not reflect principles are lower-order preferences. When I choose between actions *a* and *b*, I may first-order prefer the consequences of action *a* to those of *b* (or even independently prefer performing *a* to performing *b*), but I may second-order 'prefer', i.e. value, actions like *b* (I hold a principle that favours the class of actions to which *b* belongs). Even if I do, on the first order, value action *a* more highly than *b*, action *b represents* the class of actions covered by the second-order preference, and

7. CONCLUSION

Sen's (1997) contribution not only introduced the distinction between comprehensive and culmination outcomes, but also gave a first analysis of why the distinction matters and how it can be treated formally. This essay has attempted to extend both the formal and the substantive aspects of that investigation. On the formal side, it is fruitful to apply the conventional binary preference relation to the domain of comprehensive outcomes. The study of the properties of this relation contributes to a more nuanced understanding of how preferences can be sensitive to choice processes, such as the difference between process-indifference and process-separability. On the substantive side, I argued that failures of process-separability must be understood in terms of a mode of valuation that is distinct from the conventional independent and instrumental modes. This third mode of valuing choice processes – dependent non-instrumental valuation – can be explained by symbolic or evidential relations, but it is a mode of valuation until now largely ignored in economics and other disciplines using its methods. There should be little doubt, however, that these phenomena are real – it is illustrated, for example, by the behavioural economics research programme and its rejection of many of conventional preference theory's predictions – and that the time is ripe for economics to incorporate them in its analyses.

TECHNICAL APPENDIX

This appendix gives short proofs of the formal claims mentioned in the main text. It formally defines several conditions on pairwise preferences, one axiom of choice invariance (WARP), maximizing, and optimizing choice, and the difference between practical and non-practical preferences. It proves a series of relationships between these conditions, in six propositions. The first three analyse relationships between the conditions on preferences only. The last three study relationships between the latter and WARP. The table at the end of the appendix summarizes all the logical relationships.

Definitions

Transitivity (*T*). For all culmination outcomes x, y, z and choice processes a, b, c:

$$[(x, a)R(y, b) \text{ and } (y, b)R(z, c)] \Rightarrow (x, a)R(z, c).$$

(We assume transitivity throughout.)

---

the value it derives from them or from the second-order preference itself may outweigh the first-order preference. It is not implausible to argue that something would be missing in a value system that consisted only of first-order preferences. The imaginary agents with such simple preferences, which populate economic theory, have been variously labelled 'rational fools' (Sen 1976) or 'wantons' (Hirschman 2002 [1982]).

Process-Independence (*PI*). *R* is *process-indifferent* if, for all culmination outcomes *x* and choice processes *a,b*:

$$(x, a)R(x, b).$$

Process-Separability (*PS*). *R* is *process-separable* if, for all culmination outcomes *x,y* and choice processes *a,b*:

$$(x, a)R(y, a) \text{ if and only if } (x, b)R(y, b).$$

Strong Inter-Process Consistency (*SIPC*). For all culmination outcomes *x,y* and choice processes *a,b,c,d* (not necessarily distinct):

$$\text{if } (x, a)P(y, b), \text{ then } (x, c)P(y, d).$$

Weak Inter-Process Consistency (*WIPC*). For all culmination outcomes *x,y* and choice processes *a,b,c,d* (not necessarily distinct):

$$\text{if } (x, a)P(y, b), \text{ then not-}[(y, d)R(x, c)].$$

Weak Axiom of Revealed Preference (*WARP*). For all culmination outcomes *x,y*:

If an agent chooses *x* in some situation where both *x* and *y* are available, then in all situations where both are available and she chooses *y*, she also chooses *x*.

Optimization. An optimizer's choice behaviour selects those elements from a choice set that are weakly preferred to all the elements in that choice set.

Maximization. A maximiser's choice behaviour selects those elements from a choice set that are not strictly dispreferred to (dominated by) any other element in that choice set.

Practical and non-practical preferences.[27] For any two culmination outcomes *x,y*, and any two choice processes *a,b* (not necessarily distinct):

---

[27] *Note to definition of practical and non-practical preferences*. Strictly speaking, this definition of practical preferences is restricted to what we may call 'binarily practical' preferences. This is too narrow when what matters about the choice process is the choice set. Consider, for example, the preference between (*x* chosen from {*x,y,z*}) and (*y* chosen from {*x,y,z*}). This preference cannot be expressed in a choice between those two options *only*, since such a choice is logically impossible. Yet the preference could be practical; it depends on how the agent ranks the *third* option *z*. If she strictly prefers (*z* chosen from the set {*x,y,z*}) to the former two options, say, then the preference between the two is non-practical. If she strictly prefers both of them to the third, then it is practical. The idea is that a preference is practical if it can be expressed in choice. We might supplement the definition of 'binarily' practical preference as follows. A preference between (*x* chosen from {*x,y,z*}) and (*y* chosen from {*x,y,z*}), while not binarily practical, is practical if the agent chooses or rejects each of them (when choosing from {*x,y,z*}) exactly as she would choose or reject some arbitrary options *q* and *r* from the choice set {*q,r*} when the preference between (*q* chosen from

A preference between comprehensive outcomes (*x*,*a*) and (*y*,*b*) is *practical* if it is logically and causally possible to face a choice set between those two options only.

A preference is *non-practical* if it is not practical.

Notation

By *P* and *I* we refer to the asymmetric (strict preference) and symmetric (indifference) factor of the preference relation *R*, respectively. By $\text{WIPC}_{\text{practical}}$ and $\text{SIPC}_{\text{practical}}$ we understand the conditions WIPC and SIPC applied only to practical preferences.

Relationships between conditions on preferences:

1. PI entails PS

We assume that preferences are transitive. Let PI hold. For an arbitrary pair of comprehensive outcomes that contain the same choice process, suppose (*x*,*a*)*R*(*y*,*a*). Then for any feasible comprehensive outcomes (*x*,*b*) and (*y*,*b*):

$$(x, b)R(x, a) \text{ and } (y, a)R(y, b) \qquad \text{by PI,}$$

and therefore:

$$(x, b)R(y, b) \qquad\qquad\qquad \text{by T.}$$

So (*x*,*a*)*R*(*y*,*a*) ⇒ (*x*,*b*)*R*(*y*,*b*). A similar argument shows that (*x*,*b*)*R*(*y*,*b*) ⇒ (*x*,*a*)*R*(*y*,*a*). So (*x*,*a*)*R*(*y*,*a*) ⇔ (*x*,*b*)*R*(*y*,*b*), which is PS.

2. PS does not entail PI

It suffices to inspect the counterexamples given by the main text in the preference orders OIR and JIR (for complete preferences) or DIR (for incomplete preferences).

3. PI ⇒ SIPC ⇒ WIPC, and when preferences are complete, SIPC ⇔ WIPC ⇒ PS

3a. Suppose that PI holds and that for some arbitrary pair of comprehensive outcomes, (*x*,*a*)*P*(*y*,*b*). Then for any feasible comprehensive outcomes (*x*,*c*) and (*y*,*d*):

$$(x, c)R(x, a) \text{ and } (y, b)R(y, d) \quad \text{by PI,}$$

and therefore:

$$(x, c)R(y, d) \qquad\qquad\qquad \text{by T.}$$

{*q*,*r*}) and (*r* chosen from {*q*,*r*}) is (a) binarily practical and (b) the same as the preference between (*x* chosen from {*x*,*y*,*z*}) and (*x* chosen from {*x*,*y*,*z*}). Informally put, the preference is practical if the choice behavior mimics what it would be in a hypothetical binary choice between the options. With this definition, the proofs below would still be valid, and also cover preferences over the kind of comprehensive outcomes discussed in this footnote.

Furthermore, it cannot be that $(y,d)R(x,c)$ since, by a similar transitivity argument, this would imply $(y,b)R(x,a)$ which is a contradiction. So $(x,c)P(y,d)$. Therefore PI entails SIPC. Since $(x,c)P(y,d)$ implies not–$[(y,d)R(x,c)]$ by definition, SIPC entails WIPC. For complete preferences, the reverse implication holds.

3b. Assume that WIPC holds and that preferences are complete. Suppose $(x,a)R(y,a)$. This preference is either strict, so that not–$[(y,a)R(x,a)]$, or not.

1. If it is strict, then by WIPC, we must have not–$[(y,b)R(x,b)]$, so $(x,b)R(y,b)$.
2. If it is not strict, then $(x,a)I(y,a)$, and WIPC requires that the preference between (x,b) and (y,b) not be strict. We must therefore have $(x,b)I(y,b)$, so $(x,b)R(y,b)$.
   So $(x,a)R(y,a) \Rightarrow (x,b)R(y,b)$, and a similar argument proves the reverse implication. Therefore $(x,a)R(y,a) \Leftrightarrow (x,b)R(y,b)$, which is PS.

3c. The following incomplete preference pattern satisfies WIPC but violates PS:

$$(x, a)I(y, a), \text{ but} (x, b) \text{ and} (y, b) \text{ are not ranked.}$$

This shows that WIPC does not entail PS when preferences are incomplete.

3d. To see that PS does not imply WIPC, it suffices to note that unlike WIPC, PS imposes no constraints on preferences between comprehensive outcomes that do not share the same process component.

Relationships between conditions on preferences and axioms of choice invariance

4. WIPC$_{practical}$ and WARP are equivalent when preferences are complete
4a. Suppose WIPC holds for practical preferences. If the agent chooses (x,a) when (y,b) is available (a and b not necessarily distinct; we omit this rider in what follows), she must have the preference $(x,a)R(y,b)$. There are two possibilities:

1. If $(y,b)R(x,a)$, then $(x,a)I(y,b)$. WIPC$_{practical}$ requires that for any other choice processes c and d for which the preference between (x,c) and (y,d) is a practical one, it cannot be strict, so $(x,c)I(y,d)$. The agent therefore does not choose (y,d) without also choosing (x,c), thereby satisfying WARP.
2. If not–$[(y,b)R(x,a)]$, then $(x,a)P(y,b)$. WIPC$_{practical}$ requires that for any other choice processes c and d for which the preference between (x,c) and (y,d) is a practical one, not–$[(y,d)R(x,c)]$. The agent therefore rejects (y,d), thereby satisfying WARP.

Therefore, WIPC$_{practical}$ entails WARP when preferences are complete.

4b. Next suppose that the agent's choices satisfy WARP. If her preference is $(x,a)P(y,b)$, and this preference is a practical one, she chooses (x,a) and rejects (y,b) from the choice set $\{(x,a),(y,b)\}$. Satisfying WARP means she will always reject y when x is also available. Consider a choice set $\{(x,c),(y,d)\}$. The requirement not to choose y entails $(x,c)P(y,d)$, which means that SIPC$_{practical}$ is satisfied. Therefore, WARP entails SIPC$_{practical}$ and *a fortiori* WIPC$_{practical}$ when preferences are complete.

5. WIPC$_{practical}$ and WARP are equivalent when preferences are incomplete if choice behaviour is optimizing
5a. Suppose WIPC holds for practical preferences. If an optimizing agent chooses (x,a) when (y,b) is available, her preference must be $(x,a)R(y,b)$. As in 4a, there are two possibilities:

1. If $(x,a)I(y,b)$, then WIPC$_{practical}$ requires that for any other choice processes c and d for which the preference between (x,c) and (y,d) is a practical one, it cannot be strict. So either (x,c) and (y,d) are unranked, or $(x,c)I(y,d)$ holds. If the former, an optimizer cannot choose (y,d), so WARP is satisfied. If the latter, the argument in 4a.1 applies.
2. If $(x,a)P(y,b)$, the argument in 4a.2 applies.

Therefore, WIPC$_{practical}$ entails WARP for optimizers with incomplete preferences.

5b. Suppose that the agent's choices satisfy WARP. If her preference is $(x,a)P(y,b)$, she chooses (x,a) and rejects (y,b) from the choice set $\{(x,a),(y,b)\}$. Satisfying WARP means she will always reject y when x is also available. Consider a choice set $\{(x,c),(y,d)\}$. Rejecting y requires not–$[(y,d)R(x,c)]$, which means that WIPC$_{practical}$ is satisfied. Therefore, WARP entails WIPC$_{practical}$ for optimizers with incomplete preferences.

6. SIPC$_{practical}$ is a necessary condition for WARP to be satisfied by maximisers with incomplete preferences, but no condition on pairwise comparisons is sufficient
6a. Suppose the maximizing agent's choices satisfy WARP. If her preference is $(x,a)P(y,b)$, she chooses (x,a) and rejects (y,b) from the choice set $\{(x,a),(y,b)\}$. Satisfying WARP means she will always reject y when x is also available. Consider a choice set $\{(x,c),(y,d)\}$. Rejecting y requires $(x,c)P(y,d)$, which means that SIPC$_{practical}$ is satisfied. Therefore, WARP entails SIPC$_{practical}$ for maximisers with incomplete preferences.

6b. There is no sufficient condition on *pairwise* preferences that ensures that a maximiser with incomplete preferences will not violate WARP, except in the trivial case where she happens to have such preferences as to behave identically with an optimizer. To see the general case, consider a maximiser's behaviour when she faces the choice set $\{(x,a),(y,b)\}$. Assume she does not have a strict preference between the two options, so she chooses both. WARP requires that, from a different choice set

$\{(x,c),(y,d),(z,e)\}$, if she chooses either (x,c) or (y,d), she must choose the other as well. But this cannot be ensured by any binary preference relation between (x,c) and (y,d). It clearly is not ensured (indeed it is violated) if one is strictly preferred to the other. If the agent is indifferent or does not have a preference between them, then the two options' respective ranking against the *third* option (z,e) determines which will be chosen. Specifically, if (z,e) is strictly preferred to one and unranked against the other, the former will be rejected and the latter chosen, in violation of WARP. Thus to ensure that WARP is satisfied, we would need to impose invariance conditions not just on pairwise preference rankings across choice processes, but on whole preference *patterns*.

Summary table

	Optimizers	Maximisers	Proof
Complete and incomplete preferences:	PI $\Rightarrow$ PS		1
	not-[PS $\Rightarrow$ PI]		2
	PI $\Rightarrow$ SIPC $\Rightarrow$ WIPC		3a
	not-[PS $\Rightarrow$ WIPC]		3d
Complete preferences:	PI $\Rightarrow$ SIPC $\Leftrightarrow$ WIPC $\Rightarrow$ PS		3ab
	SIPC$_{practical}$ $\Leftrightarrow$ WIPC$_{practical}$ $\Leftrightarrow$ WARP		3a, 4
Incomplete preferences:	not-[WIPC $\Rightarrow$ PS]		3c
	WIPC$_{practical}$ $\Leftrightarrow$ WARP	WARP $\Rightarrow$ SIPC$_{practical}$	5,6

REFERENCES

Andreoni, J., P. M. Brown, and L. Vesterlund 2002. What makes an allocation fair? Some experimental evidence. *Games and Economic Behavior* 40:1–24.

Blount, S. 1995. When social outcomes aren't fair: the effect of causal attribution on preferences. *Organizational Behavior and Human Decision Processes* 63:131–44.

Brandts, J. and G. Charness. 1999. Retribution in a cheap-talk experiment. Unpublished manuscript. Available on http://www.econ.upf.edu/docs/papers/downloads/437.pdf.

Brandts, J. and Sola, C. 2001. Reference points and negative reciprocity in simple sequential games. *Games and Economic Behavior* 36:138–57

Broome, J. 1991. Utility. *Economics and Philosophy* 7:1–12. Reprinted in J. Broome. 1999. *Ethics out of Economics*. Cambridge University Press.

Broome, J. 1993. Can a Humean be moderate? In *Value, Welfare and Morality*, ed. R. G. Frey and Christopher Morris. Cambridge University Press. Reprinted in J. Broome. 1999. *Ethics out of economics*. Cambridge University Press.

Elster, Jon. 1983. *Sour Grapes*. Cambridge University Press.

Falk, A., E. Fehr and U. Fischbacher. 2003. On the nature of fair behavior. *Economic Inquiry* 41:20–6.

Güth, W., S. Huck and W. Müller. 2001. The relevance of equal splits in ultimatum games. *Games and Economic Behavior* 37:161–69.

Hirschman, A. O. 2002 [1982]. *Shifting involvements* (20th Anniversary edition). Princeton University Press.

Lind, E. A. and T. R. Tyler. 1988. *The Social Psychology of Procedural Justice*. Plenum Press.

Mas-Colell, A., M. Whinston, and J. Green. 1995. *Microeconomic Analysis*. Oxford University Press.

Nozick, R. 1993. *The Nature of Rationality*. Princeton University Press.

Prasnikar, V. and A. E. Roth. 1992. Considerations of fairness and strategy: experimental data from sequential games. *The Quarterly Journal of Economics* 107:865–88.

Raz, Joseph. 1986. *The Morality of Freedom*. Clarendon Press.

Samuelson, Paul A. 1938. A note on the pure theory of consumer behavior. *Economica* 5:61–71.

Sandbu, Martin E. 2003. Explorations in process-dependent preference theory. Ph. D. dissertation. Harvard University.

Sandbu, Martin E. (Forthcoming). Fairness and the roads not taken: an experimental test of non-reciprocal set-dependence in distributive preferences. *Games and Economic Behavior*.

Scanlon, T. M. 1986. The significance of choice. *The Tanner Lectures on Human Values*. University of Utah.

Schuessler, A. A. 2000. *A Logic of Expressive Action*. Princeton University Press.

Sen, A. 1971. Choice functions and revealed preference. *The Review of Economic Studies* 38:307–17.

Sen, A. 1974. Choice, orderings and morality. In *Practical Reason*, ed. S. Körner, 54–67. Blackwell. Reprinted in A. Sen. 1982. *Choice, Welfare and Measurement*. Harvard University Press.

Sen, A. 1976. Rational fools: a critique of the behavioural foundations of economic theory. *Philosophy and Public Affairs* 6:318–44.

Sen, A. 1993. Internal consistency of choice. *Econometrica* 61:495–521.

Sen, A. 1997. Maximization and the act of choice. *Econometrica* 65:745–79.

Sen, A. 1999. *Development as Freedom*. Anchor Books.

Smart, J. J. C. 1956. Extreme and restricted utilitarianism. *Philosophical Quarterly* 6:344–54.

Tversky, A. and Kahneman, D. 1991. Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics* 106:1039–61.

Williams, B. 1973. A critique of utilitarianism. In *Utilitarianism: For and Against*, ed. J. C. C. Smart and B. Williams, 75–155. Cambridge University Press.