STATUS QUO MAINTENANCE RECONSIDERED: CHANGING OR INCOMPLETE PREFERENCES?*

Michael Mandler

After reviewing the evidence for status quo maintenance (SQM), I consider how to reconcile SQM with traditional consumer theory. Behavioural economists usually let agents' preferences change as a function of their endowments, treating the same person with different endowments as a set of distinct agents. Many properties of preferences then become immune to empirical test and it becomes impossible to judge whether an agent's decisions make the agent better or worse off. This impedes prediction of when decision rules are likely to change. SQM can alternatively be explained with unchanging preferences if preferences are incomplete. SQM is then consistent with self-interest and there is no reason why it should not persist.

The fundamental conservatism of economic agents, their reluctance to move away from what they deem to be the status quo, has been documented by several waves of academic study. Status quo bias, the endowment effect, loss aversion and the willingness to accept-willingness to pay disparity all describe agents who must be paid a premium to trade away from their endowments.

I review some empirically well-established cases of status quo maintenance, show how they violate traditional consumer theory and then consider how the phenomenon can be explained. Behavioural economists have usually let agents' preferences change as a function of their endowments, thus treating the same person with different endowments as a set of distinct agents. But then it becomes impossible to judge whether an agent's decisions leave the agent better or worse off and hence whether the agent has reason to adopt new ways of making decisions. One can alternatively explain status quo maintenance while preserving the identity of agents through time if preferences are allowed to be incomplete. Agents with incomplete preferences maintain the status quo when they follow the simple rule of refusing to trade their endowment for unranked bundles and waiting until offered an alternative that is ranked superior. Since this rule will respect agents' interests – agents will not be led to outcomes they judge to be inferior – there are no rationality grounds for thinking that status quo maintenance cannot persist indefinitely.

Part of the clash between behavioural theory and rationality theory hinges on the role played by theoretical models. In the behavioural view, models are designed to make accurate predictions, while under the rationality view, models are also supposed to identify the types of behaviour that violate self-interest and which therefore may disappear as agents learn or face selective pressure. One purpose of this essay is to argue that status quo maintenance passes the test of self-interest; perhaps these arguments will help persuade the doubters who insist that

the non-traditional behaviour revealed in experimental work must eventually disappear. A second purpose is more critical of the behavioural camp. Even when rationality models turn out to be empirically inaccurate they may still guide empirical research. This role for rationality theory has so far been blocked by an identification of self-interest with complete and transitive preferences, which leads to predictions that are irreparably inaccurate. But if we model self-interest more carefully, rationality may again prove to be a useful explanatory tool.

1. The WTA-WTP Disparity and other Evidence of Status Quo Maintenance

The willingness to accept-willingness to pay (WTA-WTP) disparity, whose discovery dates to Hammack and Brown (1974), provided the first case of status quo maintenance to receive careful scrutiny. The disparity was initially detailed in contingent valuation (CV) surveys that ask subjects to estimate the amount of money they would trade for various levels of environmental quality.[1] Agents usually announce a sizable gap between the value of increases in environmental quality and the value of decreases: the minimum amount of money agents will accept for a unit decrease in quality (their WTA) typically exceeds the maximum amount they would be willing to pay (their WTP) for a unit increase by a factor of three or four, and only rarely does this ratio fall below two. Agents thus insist on favourable terms of trade to move from the status quo: they must be offered a relatively large amount of money to diminish environmental quality and a large amount of environmental quality per dollar to part with their money.

Many have objected that as long as agents choose hypothetically between money and environment quality, their responses can veer from their true valuations, that is, how they would choose if the real environment and their pocketbooks were on the line (Seip and Strand, 1992; Diamond and Hausman, 1993; Duffield and Patterson, 1992). In answering questionnaires, agents may just use the platform to voice their outrage at the despoiling the environment. The same motive may explain the curious insensitivity of WTPs to the scope of environmental benefits, as when the WTP that one sample of respondents reports for a few clean lakes turns out to be similar to the WTP that a separate sample reports for hundreds of clean lakes (Kahneman, 1986; Kahneman and Knetsch, 1992; Desvousges *et al.* 1993). These objections have been countered by assurances that survey protocols have since been refined and that scope insensitivities have been exaggerated (Carson, 1997). Nevertheless suspicions linger.

If we accept the veracity of CV data, the question arises whether the WTA-WTP disparity is at odds with traditional consumer theory (Diamond and Hausman, 1993, 1994). We can depict contingent-valuation data by indifference curves defined on combinations of environmental quality and money (or, more precisely, the other goods that money buys). If, in Figure 1, an agent's initial endowment of environmental quality and money is given by x, then an agent's WTA is given by the

---

[1] Camerer (1995) and Rabin (1998) both survey the status quo maintenance literatures, including the WTA-WTP disparity, as part of broader evaluations of behavioural economics. Cummings *et al.* (1986) remains an excellent overview of contingent valuation. See Hausman (1993), Bjornstad and Kahn (1996), Kopp *et al.* (1997), Bateman and Willis (1999) for more recent contributions.
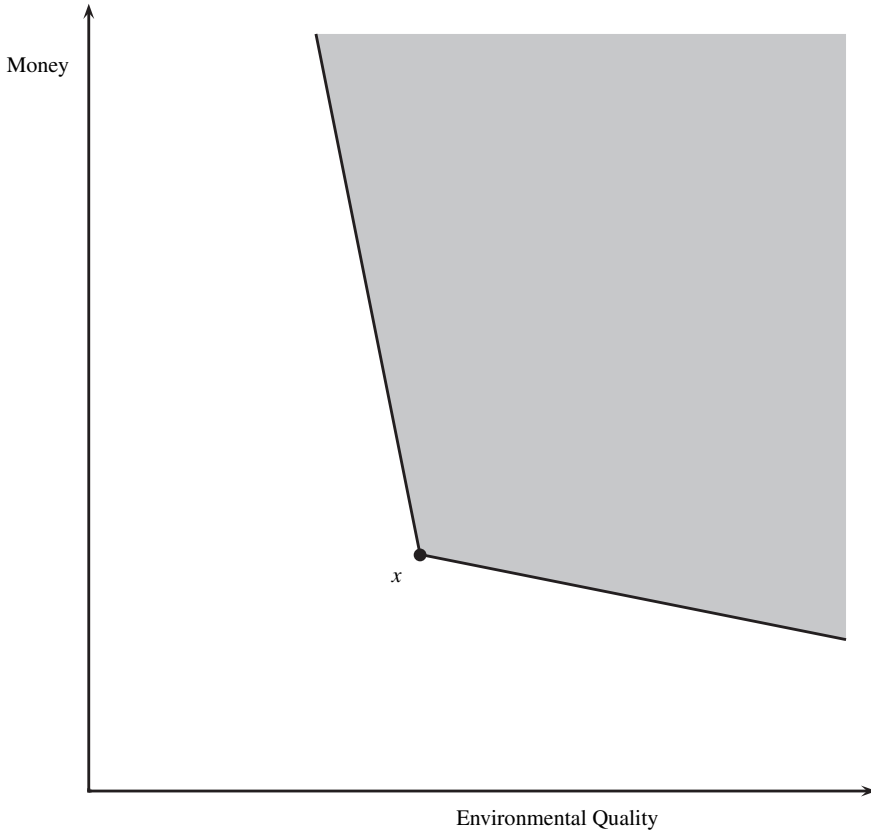
Fig. 1.   *The WTA-WTP Disparity*

(absolute value) of the slope of indifference curve segment to the upper-left of *x*, while the WTP is given by the (absolute value) of the slope of the segment to the lower-right of *x*; thus the agent will willingly move from *x* to any point in the shaded area.

   Figure 1 makes clear that one observation of a high WTA-WTP ratio is entirely consistent with standard consumer theory. Controversy persists on this point, however, with many claiming that even a single large WTA-WTP disparity is pathological (Arrow, 1993; Milgrom, 1993; Diamond, 1996). Evidently, an agent's indifference curve must display a rapid change in slope near the agent's endowment point for the WTA-WTP gap to be large; so the elasticity of substitution between environmental quality and other goods must be small. The extreme case occurs if indifference curves are kinked (as in Figure 1) but, since the measurement of a WTA-WTP disparity necessarily involves comparing discrete points, even a differentiable indifference curve can produce a large disparity if its slope changes rapidly. The asserted pathology is that large WTA-WTP disparities would entail unreasonably large income effects for environmental goods if hypothetically those goods were purchasable. This claim is not correct for the WTAs and WTPs that

typical CV surveys measure. Suppose for instance that higher and lower sets of preferred bundles are vertical translations of the set in Figure 1 and that we gave the consumer the chance to buy environmental quality along some price line tangent at x to the preferred set in Figure 1. Then the consumer would hypothetically purchase x. And if we increased or decreased the consumer's income (which would shift the price line vertically but not change its slope) the consumer's purchases of environmental quality would remain unchanged: a zero income effect.[2]

The conflict with traditional consumer theory would come if agents were to report WTPs and WTAs similar to those above from several *different* endowments, say, x and y in Figure 2. Then an agent's sets of preferred points would overlap, an undeniable violation of received consumer theory – although we will see that there is ambiguity about which axiom of the theory is contravened. Multiple reports of the same WTAs and WTPs would not imply an overlap of preferred sets unless the agent's endowment of environment quality were changing, and such changes lie beyond the scope of survey questionnaires or traditional laboratory experiments. Preferences over environmental goods thus do not provide an ideal terrain in which to nail down an indisputable case where status quo maintenance violates standard preference theory, and it is fair to conclude that only since experimental work has reported similar findings have CV studies been accepted as part of a broader indictment of traditional consumer theory.

The experimental literature has dealt with more tractable goods and provided clinching evidence. In Knetsch (1989), for example, half of a sample of n agents are given candy and the other half are given a mug. If no agent is indifferent or undecided, some $c \leq n$ agents will prefer the candy to the mug and the remaining $n - c$ agents will prefer the mug to the candy. So $c/2$ agents from the half given a mug and $(n - c)/2$ from the half given candy, half the sample in all, will want to switch their gift for the other item. In actual fact, only 10% of Knetsch's sample chose to switch – as opposed to a separate sample with no initial gift that divided evenly between preferring candy and mugs.

It is now well-established experimentally that status quo maintenance or a WTA-WTP disparity arises when agents have an entitlement to an endowment (Thaler, 1980; Knetsch and Sinden, 1984, 1987; Samuelson and Zeckhauser, 1988; Knetsch, 1989; Kahneman *et al.*, 1990). Although seemingly conclusive, these findings too have drawn criticism. One objection complains that subjects sometimes have to choose over novel items or are not disciplined by the practice of repeated market experience. The influential study by Coursey *et al.* (1987) examined changes in the

---

[2] The claim, put formally by Hanemann (1991) but asserted widely, that a large income effect *ceteris paribus* accompanies a large WTA-WTP disparity requires a distinct definition of WTA. If q is the agent's status quo level of environmental quality, q^+ is the one-unit higher level, and q^- is the one-unit lower level, then the typical CV measure of WTP is the maximum income loss an agent would pay for a move from q to q^+ and the typical CV measure of WTA is the minimum income compensation an agent would accept for a move from q to q^-. Hanemann and others use this definition of WTP but define WTA as the minimum compensation an agent would accept for a move from q^+ to q. To elicit this WTA, an agent would have to be asked, 'suppose the level of environmental quality were higher – how much would you have to be compensated if we were to return it to its original level?'. There are CV studies that ask this question, but they are the exception. Hanemann, we should add, does not conclude that WTA-WTP disparities are pathological.
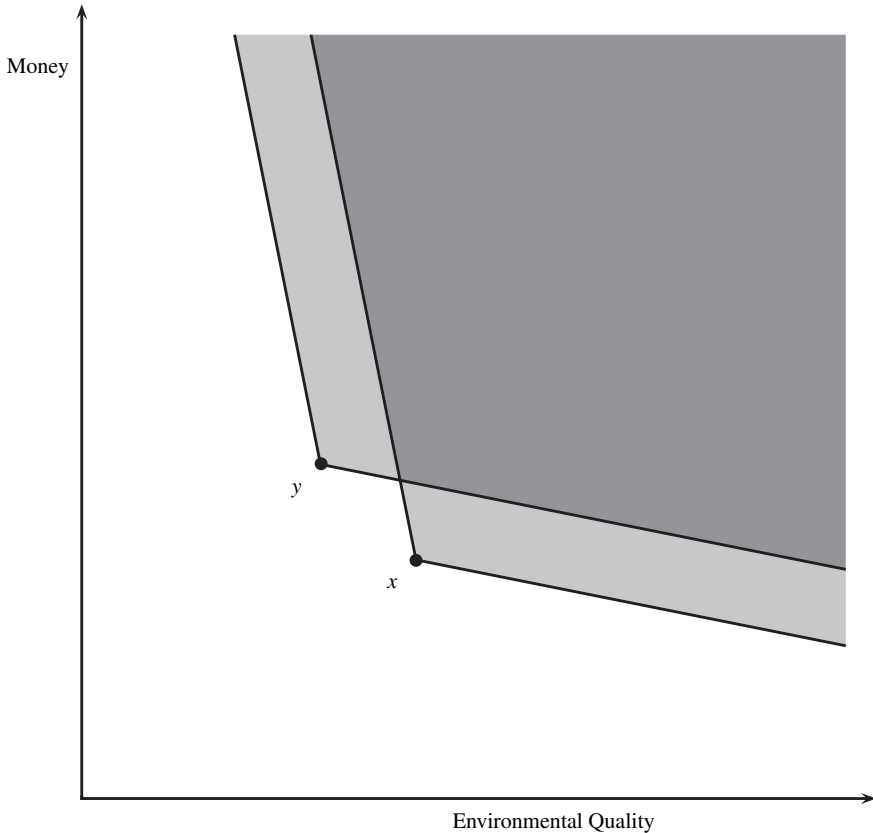
Fig. 2. *Two Simultaneous WTA-WTP Disparities*

WTA-WTP disparity where the good in question is truly exotic: the right *not* to swallow a bitter but medically harmless liquid. Not surprisingly, the disparity diminishes a little when, after sampling a bit of the liquid, agents are asked repeatedly if they will lower their WTAs or raise their WTPs, giving agents the time to mull over their valuations. The disparity can diminish more substantially when agents buy and sell the good in a Vickrey auction. Still, a WTA-WTP disparity persists. Although there is disagreement about how big the residual disparity is in the Coursey *et al.* data (Knetsch and Sinden, 1987), other work shows sizable disparities even after market or auction practice (Brookshire and Coursey, 1987; Kahneman *et al.*, 1990). But regardless of the disparity's size, it is unclear why experience in an institution like an auction should matter. Decisions whether to consume or not, or to sell or not, are all economic life usually asks of agents, while the distinctive settings of auctions, where questions of winning and losing inevitably arise, are rare. Buying and selling familiar goods requires no laboratory practice; and the students who populate economics experiments are certainly familiar with the mugs and candy and the like that show a WTA-WTP disparity.

So far, we have considered choice over riskless items, which provide the most stark illustrations of the endowment effect and status quo maintenance.

Preferences over risky prospects, such as lotteries over money or income levels, provide yet more evidence; indeed the impact of endowments on choice behaviour accompanied the first applications of von Neumann-Morgenstern expected utility theory. Friedman and Savage (1948) sought to explain why agents simultaneously display risk aversion in their insurance purchases and risk loving in their decisions to gamble. Their famous answer was to posit an expected utility function that is concave for decreases in income and convex for small increases (and a further concave region at the highest income levels, to explain why lotteries are designed with multiple large prizes rather than one huge prize). Friedman and Savage did not themselves point out that for this explanation to apply to agents whose income varies through time the inflection point of their expected utility functions must shift along with income but Markowitz (1952) did. Markowitz therefore recast utility theory to apply to changes in income not levels. The same 'reference dependence' later became a centerpiece of the prospect theory of Kahneman and Tversky (1979). In prospect theory, an agent evaluates lotteries in terms of how they deviate from the agent's pre-existing reference level of asset holdings – in practice their endowment of assets – as in Markowitz. The reference-dependent utility functions both of Markowitz and of Kahneman and Tversky display convexity for decreases in income and concavity for increases (curiously the opposite of what Friedman and Savage supposed). But Kahneman-Tversky agents distinctively do not weight the utility of a lottery's outcomes by the true probabilities of the lottery. Instead they transform those probabilities nonlinearly into new decision weights, and so their decisions can differ from what Markowitz's theory predicts. Although this feature of prospect theory has drawn the lion's share of attention – it can, for example, allow for Allais paradox behaviour – we concentrate instead on how Kahneman and Tversky further developed Markowitz's analysis of reference dependence.

To understand Kahneman and Tversky's most striking demonstration of reference dependence, let $(x, p; y, 1 - p)$ denote a lottery that delivers x dollars with probability p and y dollars with probability $1 - p$, and let $(x, 1)$ denote the degenerate lottery that delivers x with certainty. Unless stated otherwise, the dollar prizes are differences relative to initial income. Suppose one pool of subjects chooses between a lottery that gives \$1,000 with probability 0.5 and 0 dollars with probability 0.5 and a lottery that delivers 500 dollars with certainty:

$$(1000, 0.5; 0, 0.5) \text{ vs. } (500, 1).$$

A standard concave expected utility function, a Kahneman-Tversky utility and a Markowitz utility function (since the latter two are concave for income increases) all predict that subjects will select (500, 1). A second pool of subjects faces the following choice:

$$(-1000, 0.5; 0, 0.5) \text{ vs. } (-500, 1).$$

Here the risk-loving implied by the convexity of income decreases in Kahneman-Tversky or Markowitz utilities predicts that agents will choose (−1000, 0.5; 0, 0.5). But suppose that prior to choosing, everyone in the first pool is given \$1,000 dollars and everyone in the second pool is given \$2,000. Even with these prior gifts,

subjects in the first pool say they would choose (500, 1) and subjects in the second pool say they would choose (−1000, 0.5; 0, 0.5). If we suppose that the gifts lead to a change in subjects' reference levels, these data align with Kahneman and Tversky's theory. But both pairs of choices involve decisions between the same changes of income relative to agents' pre-gift holdings: for a subject with initial income I the decision in both cases is between the lotteries $(I + 2000, 0.5; I + 1000, 0.5)$ and $(I + 1500, 1)$. Strikingly, the mere announcement to agents that they are temporarily receiving gifts of \$1,000 and \$2,000, even though they have no right to walk away with those gifts and will have to choose between risky prospects, is enough to change their reference points.

These predictions about reference points are consistent with Markowitz's theory as well. The distinctive Kahneman and Tversky assumption that provides a theory of status quo maintenance is that the utility gains of income increases are small relative to the utility losses of similarly sized income decreases. This ensures that agents will exhibit risk aversion in choices where one option is the null lottery $(0, 1)$ that leads to no change in income (and is made consistent with the concavity and convexity mentioned earlier by letting utility be linked at the reference point). For example, Kahneman-Tversky agents, and actual experimental agents, will choose the null lottery over fair gambles that involve a 0.5 probability gain of x dollars and a 0.5 probability loss of x dollars. Reference-dependence implies just that choice behaviour changes as a function of agents' endowments; it is the further assumption of risk aversion relative to the null lottery that adds the status quo maintenance twist.

2. Explaining Status Quo Maintenance: Changing or Incomplete Preferences?

We return to the simpler case of choice under certainty and the example of status quo maintenance in Figure 2. The figure is evidently at odds with traditional rationality theory, but it is less clear which feature of the theory is implicated. Putting aside supplemental assumptions that arise in particular settings, a rational preference relation $\geqslant$ satisfies three properties:

(1) completeness ($x \geqslant y$ or $y \geqslant x$ for any pair of bundles x and y)
(2) transitivity (if $x \geqslant y$ and $y \geqslant z$ then $x \geqslant z$), and
(3) constancy through time.

While (1) and (2) are routinely identified with rationality, (3) may provoke doubts since economic theory commonly distinguishes goods by date, e.g., watermelons-next-winter versus watermelons-next-summer. Consumption bundles must therefore specify the dates (and possibly states) at which goods appear. For example, $x \geqslant y$ might say that a bundle $x = (x_w, x_s)$, denoting x_w watermelons-next-summer and x_s watermelons-next-winter, is weakly preferred to $y = (y_w, y_s)$. The constancy of preferences through time then means only that if $x \geqslant y$ today then $x \geqslant y$ at any subsequent date prior to receiving any substantive information about the bundles (e.g., about the weather next summer or winter); it does not say that one quantity of watermelons is preferred to another regardless of date and climate.

Dropping each of the above properties leads to a different way of explaining status quo maintenance. The most common nowadays is to drop (3) and suppose that preferences change through time as a function of an agent's endowment or reference point. Tversky and Kahneman (1991), (see also Kahneman *et al.* (1990)), offer an explicit model of reference dependence in the choice over certain prospects; a great many experiments and theories of the last twenty years follow the same programme. If we allow preferences to change through time, then we can retain properties (1) and (2). In the case of Figure 2, we could evidently posit two different complete and transitive preference relations, one preference relation, say $\succcurlyeq_x$, when the agent's endowment is x and a second, $\succcurlyeq_y$, when the agent's endowment is y.

But there are drawbacks to letting preferences be so flexible. First, preference relations become largely unverifiable. For $\succcurlyeq_x$, only choices between x and other bundles can be observed, but $\succcurlyeq_x$ also postulates rankings between every other pair of bundles, e.g., between y and other bundles. The ordinalist principle that assumptions on preferences be confirmable using choice data therefore must fall by the wayside. Transitivity, for instance, can be neither affirmed nor falsified: verification may require for a triple of bundles (a, b, c) that the three rankings a vs b, b vs c, and c vs a be checked but $\succcurlyeq_x$ allows only rankings *vis-à-vis* x to be observed. Thus, although dropping (3), preference constancy, allows us to hold on to (1) and (2), it does so at the cost of immunising (1) and (2) from empirical test.

Second, if preference judgements at different dates are uncoupled, welfare evaluation of an agent's actions is severely circumscribed. If an agent endowed with x willingly moves to a bundle w and then to v (so $w \succcurlyeq_x x$ and $v \succcurlyeq_w w$), should $\succcurlyeq_x$ or $\succcurlyeq_w$ be used to judge if the agent is better off with v than with x? If $\succcurlyeq_x$ and $\succcurlyeq_w$ issue different judgements, no clear conclusion emerges. This inability to evaluate an agent's welfare is not just a normative problem, it hobbles empirical analysis. When agents make choices that diminish their welfare, there is good reason to investigate whether they subsequently adopt new ways of making decisions: agents who do not start off as rational can learn from their mistakes and make future choices differently. But if we drop (3), we cannot even say whether status quo maintenance itself hurts those who practise it and hence whether the behaviour is likely to continue.

To retain the constancy of preferences, we can instead drop the completeness assumption.[3] An agent with incomplete preferences does not make preference judgements between certain pairs of bundles. Particularly with intangible goods, an agent may not form a definitive view of how much money an incremental unit of a good is worth (Diamond and Hausman, 1994). But even for standard consumption goods there is no reason why agents should always be able to judge which bundles leave them better off. When agents cannot make welfare judgements, they must decide by other means. One simple way to proceed is to maintain

[3] Incomplete preferences have recently been the subject of renewed research: see, e.g., Bewley (1986), which has proved highly influential, Danan (2001), Dubra *et al.* (2004), Eliaz and Ok (2002), Ok (2002), and Mandler (2004).

the status quo – to hold on to one's current endowment until offered a bundle that one can rank as superior.

We have begun to give preference a psychological interpretation where $x \succcurlyeq y$ means that the agent thinks he or she is at least as well off with x as with y, or, put differently, thinks he or she will experience more welfare with x than with y.[4] Incomplete preference between x and y then holds when an agent cannot judge if x or y is superior. Although we will need this psychological or welfare interpretation to judge the rationality of an agent's decisions, we can still test behaviourally when an agent believes he or she is better off by concluding that $x \succcurlyeq y$ holds if and only if the agent, when endowed with y, agrees to switch to x. Thus an agent has to reject an endowed (or 'status quo') y actively in favour of x for an observer to infer than the agent prefers x over y. We therefore infer that incomplete preference between x and y holds if an agent when endowed with y refuses to switch to x and when endowed with x refuses to switch to y. In contrast, an agent who is indifferent between x and y will agree both to switch from an endowed x to y and to switch from an endowed y to x (or agree to these exchanges when offered arbitrarily small bonuses). Incompleteness is therefore behaviourally distinct from indifference and is an entirely ordinal concept.

To pursue an incompleteness explanation of Figure 2, we need to disentangle completeness and transitivity. It is so common to assume that preferences satisfy or violate both assumptions at once that it is easy to overlook that they can be represented graphically as distinct conditions.

Given the preferences $\succcurlyeq$ and the consumption bundle x, we label the bundles weakly preferred to x (or better than x) as $B(x)$ and the bundles that are weakly preferred to x (or worse than x) as $W(x)$. That is, $B(x) = \{w: w \succcurlyeq x\}$ and $W(x) = \{w: x \succcurlyeq w\}$. In behavioural terms, $B(x)$ is the set of bundles that the agent would switch to when endowed with x, and $W(x)$ is the set of endowed bundles from which the agent would switch to x. We label the bundles strictly preferred to x as $SB(x)$ and the bundles x is strictly preferred to as $SW(x)$: $SB(x) = \{w: w \succ x\}$ and $SW(x) = \{w: x \succ w\}$.[5] When w is in $SW(x)$, for example, the agent is willing to switch from an endowed w to x but not to switch back to w if x were the endowment.

If $\succcurlyeq$ is transitive, then $z \succcurlyeq y$ and $y \succcurlyeq x$ imply $z \succcurlyeq x$. Since we can express $z \succcurlyeq y$ and $y \succcurlyeq x$ equivalently as saying that $B(x)$ and $W(z)$ have the common element y, $\succcurlyeq$ is transitive if and only if whenever $B(x)$ and $W(z)$ intersect, then z is in $B(x)$. See Figure 3, where the intersection of $B(x)$ and $W(z)$ is shaded, for a characteristic violation of transitivity.

Now return to Figure 2. Suppose that the bundles the agent willingly accepts in exchange for x or for y all happen to be strictly preferred, that is, suppose the sets $SB(w)$ and $B(w)$ coincide for $w = x, y$ except that w is in $B(w)$ but not $SB(w)$. Then the overlap of the B sets in Figure 2 does *not* imply a violation of transitivity. Figure 2 identifies only the bundles preferred to x or preferred to y, but to either

---

[4] See Sen (1973 1982, 1997) and Levi (1986) on the distinction between psychological and choice interpretations of preference.

[5] Of course, $\succ$ is defined in terms of $\succcurlyeq$: $a \succ b$ obtains if and only if $a \succcurlyeq b$ and not $b \succcurlyeq a$.
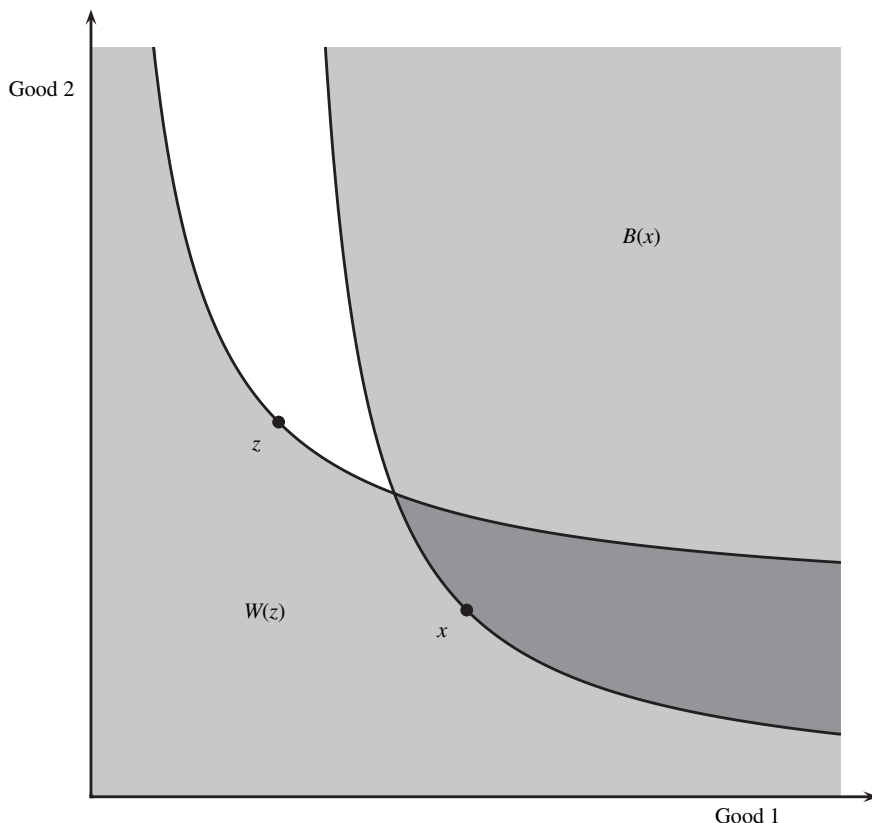
Fig. 3.   *A Violation of Transitivity*

confirm or contradict transitivity we must have three bundles *a*, *b*, *c* such that
$a \succcurlyeq b \succcurlyeq c$, or, equivalently, bundles *a* and *c* such that $B(c)$ and $W(a)$ intersect.[6]
The overlap of two *B* sets in Figure 2 would imply the type of overlap of a *B* set and
a *W* set pictured in Figure 3 only if we required preferences to be complete (since
then the complement of a *B* set would be a *W* set).

   If the behaviour in Figure 2 is in accord with transitivity (and we do not allow
preferences to change through time) it must then be completeness that is violated.
Incomplete preferences allow unchosen bundles, which are typically classified as
dispreferred, to be labelled as unranked instead: when some *y* is not in $B(x)$,
it is now permissible for *x* not to be in $B(y)$. It is this innovation that allows
the evidence of status quo maintenance to be generated by a single transitive
preference violation.

---

[6] We would have such a triple if either $y \in B(x)$ or $x \in B(y)$. If say $y \in B(x)$ then *y* would be in the
intersection of $B(x)$ and any $W(z)$ such that $z \in B(y)$. And then an overlap of $B(x)$ and $B(y)$ would imply
a violation of transitivity: an overlap means there would be a *z* in $B(y)$ but not in $B(x)$, which under
transitivity is inconsistent with *y* being in both $B(x)$ and $W(z)$. We can now also explain why we have
assumed that the *SB* and *B* sets coincide. If one of the sets in Figure 2 were a *B* set, say $B(x)$, and its
boundary were an indifference curve, then for some point *w* both on *x*'s indifference curve and in the
darkly shaded area, we would have $x \sim w$ and $w \succcurlyeq y$. So, if $\succcurlyeq$ were transitive, $x \succcurlyeq y$, that is, $x \in B(y)$,
which evidently does not obtain in Figure 2.

We can depict incomplete preferences graphically by designating a set of unranked bundles. Given an arbitrary *x*, three types of bundles must be distinguished: preferred, dispreferred and unranked. In behavioural terms, the preferred bundles are those the agent actively selects when endowed with *x*, the dispreferred are endowments from which the agent agrees to switch to *x* and the unranked are those bundles that relative to *x* the agent neither switches to nor switches from. The three categories are depicted in Figure 4, where the unranked bundles are shaded.

The preferred set in Figure 4 coincides with the preferred set in Figure 1, and thus the agent behind Figure 4 displays the same active choice behaviour and hence the same WTA-WTP disparity as the agent behind Figure 1. And since incompleteness permits an overlap of preferred sets, as in Figure 2, agents with incomplete preferences can display similar WTA-WTP disparities from multiple status quos, as well as the other characteristic signs of status quo maintenance.

Unlike the case where $\geqslant$ changes through time, we can now confirm empirically which properties $\geqslant$ satisfies. In the case of transitivity, we can check, when *x* is in $B(y)$ and *y* is in $B(z)$, whether *x* is in $B(z)$. To confirm incompleteness, we can see if
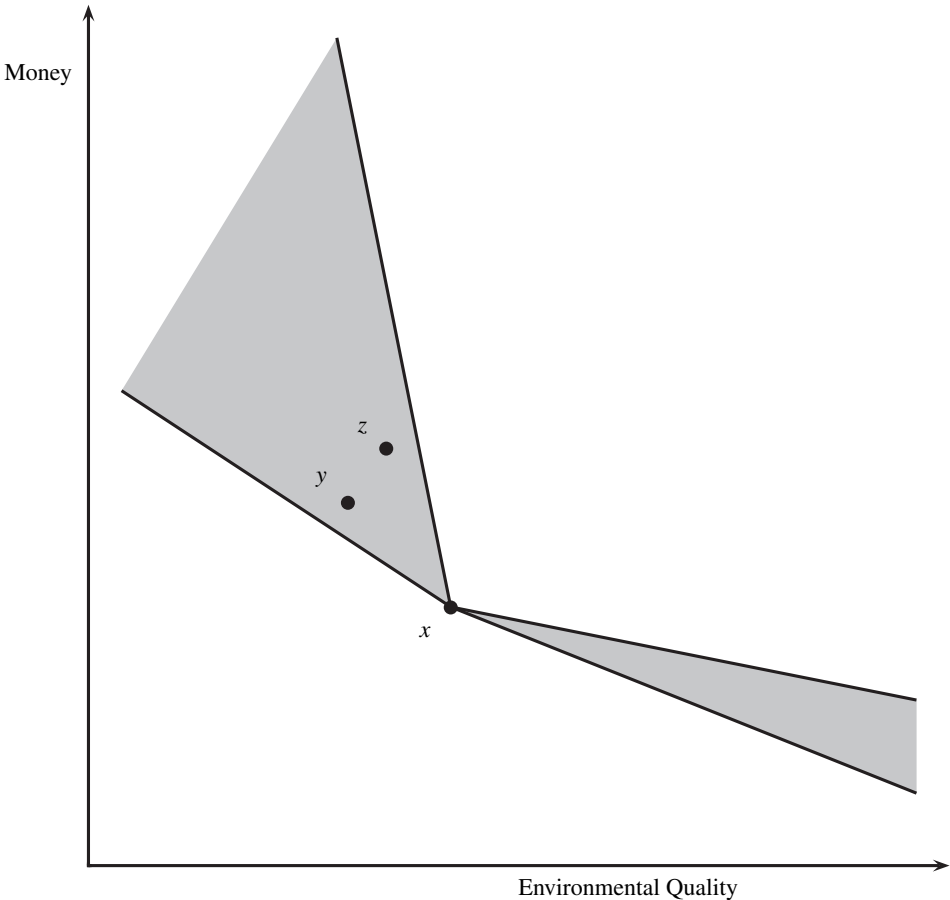

Fig. 4. *Incomplete Preferences*

there is some pair (x, y) such that x is not in $B(y)$ and y is not in $B(x)$. We can also examine whether behaviour becomes more rational through time, e.g., whether trading patterns that expose agents to manipulation disappear when agents have a chance to practice trading. Incomplete but constant preferences thus open up a programme of answerable empirical questions.

Status quo maintainers with incomplete preferences can still satisfy transitivity; this is, moreover, indispensable to their rationality since intransitivity is the classical case of a behaviour that invites manipulation. To evaluate whether an agent is rational, we must here distinguish between an agent's choice behaviour and his or her psychological judgements of well-being (though as we pointed out we may infer these judgements from choice behaviour). Otherwise we would have no way to decide if a sequence of choices leaves an agent better or worse off. If $\geqslant$ denotes these psychological welfare judgements and $\geqslant$ is transitive and the agent is always allowed to hold on to his current bundle, status quo maintenance can never lead the agent from an initial bundle x to a terminal bundle z that is worse than x, i.e., where $x \succ z$. In contrast, with intransitive psychological preferences, where there is a triple a, b, c, such that $a \geqslant b$, $b \geqslant c$, and $c \succ a$, there are always sequences of choices that lead an agent from a superior to an inferior bundle: beginning with c, the agent would agree to switch to b, and then subsequently to a, and a is inferior to c. Sometimes such sequences are said to be a *money pump*, particularly if the agent ends up with a bundle that has less of each of the goods than his initial bundle. For these reasons, transitivity is rightly considered to be a fundamental characteristic of rational psychological preferences. Since status quo maintainers with transitive but incomplete preferences are always immune to money pumps, again assuming they are always allowed to retain their endowed bundle, completeness does not share a comparable rationality status. In fact, even when agents are not allowed to retain their current bundle, agents with incomplete but transitive preferences can always violate the predictions of traditional consumer theory without exposing themselves to money pumps or other hazards (Mandler, 2004).

Although status quo maintenance is not the only rational way for an agent with incomplete preferences to choose, an unrestricted willingness to switch between unranked bundles can readily expose an incomplete-preferences agent to a money pump, even when the agent's preferences are transitive. In Figure 4, for example, the agent could be led from z to x and then to y. Status quo maintenance blocks such sequences of trades and it does so with an easily executable rule.

A third way to account for status quo maintenance is to retain the constancy of preferences through time and completeness and instead to drop transitivity.[7] Many believe that it is impossible for agents to display an incompleteness of preference or to verify such an occurrence empirically: agents can always be compelled to choose between any pair of options and thus reveal a preference one way or the other. If an agent holds to a status quo bundle x when offered the chance to switch to y and also refuses to switch to x when y is the status quo, then the agent should be labelled as indifferent between x and y. So, instead of assuming that $\geqslant$ is

---

[7] A literature on intransitive indifference flourished in the 1950s and 1960s (see Fishburn, 1970) for a survey) and these models may in retrospect be seen as formalisations of status quo maintenance.

incomplete but still possibly transitive, we can ascribe enough cases of indifference to force $\geqq$ to be complete.

Of course if $\geqq$ is complete then in cases such as Figure 2 $\geqq$ must be intransitive. Completeness means that if it is *not* the case that $y \succ x$, then $x \geqq y$: graphically, the complement of $SB(x)$ must equal $W(x)$ for any x. Hence, if $\geqq$ is complete, then Figure 2 would depict an intransitivity: the complement of the points in $SB(x)$ would be $W(x)$, and so Figure 2 would have the same form as Figure 3. Indeed, for complete and transitive preferences, B sets cannot 'overlap': it is easy to confirm that $B(x) \subset B(y)$ or $B(y) \subset B(x)$ or both. And similarly pairs of W or SB or SW sets cannot overlap when $\geqq$ is complete and transitive.

Using complete but intransitive preferences to represent the agent behind Figure 2 has some awkward features. First, even though x and y are designated as indifferent, the agent always rebuffs offers to exchange x for y when endowed with x. Second, in contrast to standard cases of indifference, a small improvement to one of the bundles need not lead it to be preferred: a point z just to the northeast of y (see Figure 4) can, like y itself, also be designated as indifferent to x. Incompleteness can explain this behaviour; for complete and intransitive preferences they are puzzles. Finally, as we argued earlier, incompleteness *can* be distinguished observationally from indifference: when an agent is genuinely psychologically indifferent between x and y, he or she will be willing to switch from x to y and back again (or induced to do so with arbitrarily small bonuses) but when no preference between x and y obtains, the agent may refuse to switch either from x to y or from y to x.

In the end, 'indifference' and 'incompleteness' are labels; what matters is agents' behaviour. So, we can certainly call bundles such as x and y in Figure 4 indifferent. But we would then need another set of classifications to keep track of the fact that agents are willing to switch among some cases of indifferent bundles but not among others, and that indifference can be converted to preference through small improvements in some cases but not in others. Verbal gymnastics aside, we would have returned to the incomplete preference theory.

The incompleteness view distinguishes between two types of choice behaviour: the active selection of a bundle over the status quo versus a passive 'revealed preference' in favour of the status quo over alternative bundles. Although both are types of preference, they have different origins and consequences. Active choice usually requires some minimum of conscious thought and deliberation whereas retaining the status quo is a default that can operate on automatic pilot; thus different mental operations come into play; see Kahneman (2003) on the distinction, though not directed to this conclusion. It is a sign of the power of the received view of preference as homogeneous that even behavioural economists do not normally distinguish between the two types of choice. Active choice is not only mentally distinctive; it also allows us to extract at least some welfare conclusions from choice behaviour. When an agent both rejects offers to switch to w when x is the status quo and deserts w in favour of x when w is the status quo, it is reasonable to infer that the agent believes that he or she is better off with x than with w. We cannot deduce as many welfare rankings as we can from the agents of traditional preference theory but we do not need to adopt the wholesale welfare agnosticism of behavioural economics.

3. Conclusion: Incomplete Preferences vs Ordering Principles

The prevalence of status quo maintenance shows that the orthodox theory of economic rationality fails as a predictive model. The response of behavioural economics has been to dissociate the choices agents make at different dates. If instead we allow preferences to be incomplete, or, put differently, distinguish between actively chosen bundles and passively retained bundles, then we can continue to suppose that the various selves that make decisions at different times are all manifestations of the same preferences. More empirical claims – e.g., that preferences are transitive – can then be tested and more normative questions – which decision rules lead agents to better outcomes? – can be asked and answered. Incompleteness can preserve the longstanding connection in economics between rationality and choice behaviour, although we must understand rationality, not as completeness and transitivity, but as the more basic property of making choices that do not lower welfare. On this view, status quo maintenance appears as a simple and effective way to make self-interested decisions.

When are incomplete preferences and status quo maintenance most likely to arise? Simple familiarity with goods is not enough to ensure that agents have complete orderings. But when agents can turn to *ordering principles* that rank their underlying choices or objectives, then completeness is more likely to obtain or to develop through time (Mandler, 2001). For instance, if an agent feels that some set of choices can and should be ordered according to the amount of physical pleasure the alternatives deliver, then a small consumption sample, as at a deli counter, may be able to decide the matter; incompleteness is then likely to be nonexistent or temporary. Or if choices vary only in their efficiency in delivering some service – say various drilling sites by how much oil they contain – they can be ordered by their expected efficiency; even with incomplete information, agents may be able to estimate efficiency and thereby build a complete ordering. In these examples, there is a criterion by which agents can gauge their decisions, and so information or even just time to think can resolve an initial inability to rank.

When agents cannot find an ordering principle that reduces disparate alternatives – the proverbial apples and oranges – to a common benchmark, there is no reason for preferences to be complete or to become complete through time, whether or not choices involve familiar goods. It is common, even in behavioural economics (Kahneman and Tversky, 1979), to argue that violations of the standard model of complete and transitive preferences will normally be self-correcting when agents have a chance to learn from their mistakes. This position is often defended by reasoning that preferences must be complete since an agent can be compelled to choose between any pair of alternatives and that these compelled choices must be transitive since otherwise the agent could be money pumped. But this argument supposes that an agent will always choose between a pair of bundles in the same way. Status quo maintainers, in contrast, hold to their endowment *whatever it is*, unless offered a superior bundle. It may then appear that a status quo maintainer has overlapping indifference curves but, as we have seen, this conclusion is not warranted if we do not interpret retention of the status quo as preference over nonselected bundles. If the agent's psychological welfare judgements (or active

choices) are transitive, status quo maintainers will be immune to money pumps and will not end up with inferior outcomes. In terms of the bundles agents ultimately consume, status quo maintenance is therefore fully rational and there is no reason for evolution or self-interest to weed it out.[8]

When ordering principles are present and preferences are complete, choice behaviour may of course still violate some of the axioms of traditional rationality theory. But those axioms nevertheless have an important role to play in empirical analysis. A couple of examples will illustrate. I mentioned choice over oil drilling sites as a characteristic case where an ordering principle is present. Despite this fact, there is good evidence that oil companies bid irrationally for drilling rights and fall victim to the winner's curse (Capen *et al.*, 1971; Hendricks *et al.*, 1987). Companies do not adequately take into account the fact that if they base their bids solely on their information about the oil content of sites then they are more likely to win an auction when their private information misleadingly indicates high oil content. Optimal bids should be adjusted downwards to take account of this correlation between overoptimistic information and winning an auction. But this is a sophisticated principle that amateurs and experimental subjects rarely grasp without repeated practice; see Kagel (1995) for a guide to a large literature. An outsider studying a mining industry therefore cannot know in advance if industry participants appreciate the problem. Moreover, if it turns out that bidders for mining sites indeed fall prey to the winner's curse, the question whether bidders will eventually discover their errors and adjust their future bids is necessarily a leading topic for investigation. If it turns out that mining companies continue decade after decade in overbidding, a fully satisfactory analysis must explain this persistence. Thus, even when the predictions of rationality theory are not directly accurate, they guide empirical research. In this case, this conclusion hinges on the presence of an ordering principle – profit maximisation – that gives agents a way to order their alternatives and on this ordering principle remaining stable through time. If agents have no systematic way to rank their choices, they can learn nothing that would lead them to reconsider their decisions, and rationality theory would not then predict any long-run corrective pressure on behaviour.

As a second example, consider the theory of rational herding and informational cascades; see Banerjee (1992) and Bikhchandani *et al.* (1992), and Bikhchandani *et al.* (1998) and Gale (1996) for surveys. In its simplest form, this theory considers a queue of agents facing a binary choice, say, motorists deciding between a high road or a low road as the quickest way to get from *A* to *B*. Each motorist in the queue receives an independent signal (or has an independent opinion) about which road is speedier and can observe the turns made by at least two previous motorists further along the road. Let the probability p that any given signal is correct be greater than 0.5. Herding theory argues that if the motorists are rational Bayesians, it will not be unusual to see every car in the queue taking the slower route, say the low road. Suppose the first two drivers take the low road, driver 1

---

[8] For more detail on the defects of arguments that link the traditional rationality model with self-interest, see Mandler (2004).

because he receives a low-road signal and driver 2 because either he too receives a low-road signal or he receives a high-road signal, and, realising that driver 1 must have received the opposite signal, chooses on a coin toss that happens to come up low. Each subsequent driver then realises that his information is swamped by the previous motorists' information. Even if driver 3, say, were to receive a high-road signal, as a rational Bayesian he would realise that though his and driver 1's information 'cancel,' it is more likely that driver 2 received a low-road signal, since in some fraction of the cases that 2 receives a high-road signal he takes the high road by coin toss and 3 has seen 2 take the low road. Thus, regardless of 3's signal, 3 also takes the low road. Drivers 4 and beyond can therefore infer nothing about any signal received by any driver after the second, and so they also take the low road regardless of their signal. Thus, two unlucky signals and the entire queue takes the wrong turn – the drivers act like a herd.

It would be a rare motorist who pursues or follows the details of this reasoning. Perhaps most make no calculations of self-interest at all and act out of conformism and regret aversion: drivers contemplating a high-road deviation might want to avoid at all costs the feeling that they foolishly broke from the pack only to hit a two-hour delay. Although real motorists may be motivated primarily by regret aversion, herding theory still has an important message: even when motorists are exempt from social psychological pressure, they may still act like a herd. The force of rationality thus does not overturn agents' instincts to conform and this should increase confidence in a prediction of persistent herding. Social scientists, as out-siders, rarely know agents' exact motives or how rational they are or, in this case, how adept they are at Bayesian updating. In the face of this ignorance, predictions that hold regardless of the extent of agents' rationality are more likely to be correct.

Curiously, it is the descriptive *in*accuracy of herding theory that makes it useful. If we enriched the model by endowing motorists with the full range of motives they might have, the logic of why it can be rational to mimic others would be muddied. Shearing away the complexity of human motivation spotlights the isolated effect of rational information transmission.

Analogously to the previous example, the support that herding theory gives to the prediction that agents act as a crowd depends on the ordering principle that the drivers desire only to get from A to B quickly. If motorists have additional goals – e.g., how picturesque the drive is – and they believe that each of the routes serves a different goal more effectively, then a herding theory has no bite; it would not make any firm prediction about how the accuracy of agents' updating affects their choice of route.

I have chosen these examples so that in one case (herding) typical behaviour follows the dictates of self-interest but in the other (the winner's curse) typical behaviour appears not to. Regardless of whether actual behaviour is rational, therefore, the categorisation of behaviour as rational or irrational is germane for empirical work.

Returning to status quo maintenance and consumer theory, we may conclude that the mere empirical inaccuracy of the traditional rationality theory of complete and transitive preferences does not by itself render the theory irrelevant. Rather the problem is that the traditional theory is often employed in settings where no

ordering principle is applicable and where therefore the theory does not model agents' self-interest accurately. The traditional theory is not pertinent in these cases even in the long-run and there is no need to explain why agents do not obey its dictates (as there was in the winner's curse model).

The behavioural literature on status quo maintenance has never been clear about the rationality standing of the behaviour it documents, leading traditionalists to suspect that status quo maintenance must ultimately be fleeting. But if incomplete psychological preferences lie behind status quo maintenance, the behaviour need not leave agents worse off. And if the objects of choice are not subject to an ordering principle, there is no reason for status quo maintenance to disappear with time. Once the compatibility between incompleteness and rationality is recognised, the longstanding link in economic theory between behaviour and self-interest can be re-established.

Royal Holloway College, University of London

References

Arrow, K. (1993). 'Contingent valuation of nonuse values: observations and questions', in (J. Hausman, ed.) *Contingent Valuation: a Critical Assessment*, Amsterdam: Elsevier Science.

Banerjee, A. (1992). 'A simple model of herd behaviour', *Quarterly Journal of Economics*, vol. 107, pp. 797–818.

Bateman, I. and Willis, K. (1999). *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford: Oxford University Press.

Bewley, T. (1986). 'Knightian decision theory: Part I', Cowles Foundation Discussion Paper 807.

Bikhchandani, S., Hirshleifer, D. and Welch, I. (1992). 'A theory of fads, fashion, custom, and cultural change as informational cascades', *Journal of Political Economy*, vol. 100, pp. 992–1026.

Bikhchandani, S., Hirshleifer, D. and Welch, I. (1998). 'Learning from the behaviour of others: conformity, fads, and informational cascades', *Journal of Economic Perspectives*, vol. 12, pp. 151–70.

Bjornstad, D. and Kahn, J. (eds) (1996). *The Contingent Valuation of Environmental Resources*, Cheltenham: Edward Elgar.

Brookshire, D. and Coursey, D. (1987). 'Measuring the value of a public good: an empirical comparison of elicitation procedures', *American Economic Review*, vol. 77, pp. 554–66.

Camerer, C. (1995). 'Individual decision making', in (J. Kagel and A. Roth, eds.) *The Handbook of Experimental Economics*, Princeton: Princeton University Press.

Capen, E., Clapp, R. and Campbell, W. (1971). 'Competitive bidding in high-risk situations', *Journal of Petroleum Technology*, vol. 23, pp. 641–53.

Carson, R. (1997). 'Contingent valuation and tests of insensitivity to scope', in (R. Kopp, W. Pommerhene and N. Schwartz, eds.) *Determining the Value of Non-marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, Boston: Kluwer.

Coursey, D., Hovis, J. and Schulze, W. (1987). 'The disparity between willingness to accept and willingness to pay measures of value', *Quarterly Journal of Economics*, vol. 102, pp. 679–90.

Cummings, R., Brookshire, D. and Schulze, W. (eds.) (1986). *Valuing Environmental Goods*, Totowa, NJ: Rowman and Allanheld.

Danan, E. (2001). 'Behavioural foundations of incomplete preferences', mimeo, Université de Paris 1.

Desvousges, W., Johnson, F., Dunford, R., Boyle, K., Hudson, S. and Wilson, K. (1993). 'Measuring natural resource damages with contingent valuation: tests of validity and reliability', in Hausman (1993).

Diamond, P. (1996). 'Discussion of the conceptual underpinnings of the contingent valuation method by A. C. Fisher', in Bjornstad and Kahn (1996).

Diamond, P. and Hausman, J. (1993). 'On contingent valuation measurement of nonuse values', in Hausman (1993).

Diamond, P. and Hausman, J. (1994). 'Contingent valuation: is some number better than no number?', *Journal of Economic Perspectives*, vol. 8, pp. 45–64.

Dubra, J., Maccheroni, F. and Ok, E. (2004). 'Expected utility theory without the completeness axiom', *Journal of Economic Theory*, vol. 115, pp. 118–33.

Duffield, J. and Patterson, D. (1992). 'Field testing existence values: an instream flow trust fund for Montana rivers', mimeo, Department of Economics, University of Montana.

Eliaz, K. and Ok, E. (2002). 'Indifference or indecisiveness: choice theoretic foundations of incomplete preferences', mimeo, New York University.

Fishburn, P. (1970). 'Intransitive indifference in preference theory: a survey', *Operations Research*, vol. 18, pp. 207–28.

Friedman, M. and Savage, L. (1948). 'The utility analysis of choices involving risks', *Journal of Political Economy*, vol. 56, pp. 279–304.

Gale, D. (1996). 'What have we learned from social learning?', *European Economic Review*, vol. 40, pp. 617–28.

Hammack, J. and Brown, G. (1974). *Waterfowl and Wetlands: Toward Bioeconomic Analysis*, Baltimore: Johns Hopkins University Press.

Hanemann, W. (1991). 'Willingness to pay and willingness to accept: how much can they differ?', *American Economic Review*, vol. 81, pp. 635–47.

Hausman, J. (ed.) (1993). *Contingent Valuation: A Critical Assessment*, Amsterdam: Elsevier Science.

Hendricks, K., Porter, R. and Boudreau, B. (1987). 'Information, returns, and bidding behaviour in OCS auctions: 1954–1969', *Journal of Industrial Economics*, vol. 35, pp. 517–42.

Kagel, J. (1995). 'Auctions: a survey of experimental research', in (J. Kagel and A. Roth, eds.) *The Handbook of Experimental Economics*, Princeton: Princeton University Press.

Kahneman, D. (1986). 'Comments', in Cummings *et al.* (1986).

Kahneman, D. (2003). 'Maps of bounded rationality: psychology for behavioral economics', *American Economic Review*, vol. 93, pp. 1449–75.

Kahneman, D. and Knetsch, J. (1992). 'Valuing public goods: the purchase of moral satisfaction', *Journal of Environmental Economics and Management*, vol. 22, pp. 57–70.

Kahneman, D. and Tversky, A. (1979). 'Prospect theory: an analysis of decision under risk', *Econometrica*, vol. 47, pp. 263–91.

Kahneman, D., Knetsch, J. and Thaler, R. (1990). 'Experimental tests of the endowment effect and the Coase theorem', *Journal of Political Economy*, vol. 98, pp. 1325–48.

Knetsch, J. (1989). 'The endowment effect and evidence of nonreversible indifference curves', *American Economic Review*, vol. 79, pp. 1277–84.

Knetsch, J. and Sinden, J. (1984). 'Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of values', *Quarterly Journal of Economics*, vol. 99, pp. 507–21.

Knetsch, J. and Sinden, J. (1987). 'The persistence of evaluation disparities', *Quarterly Journal of Economics*, vol. 102, pp. 691–5.

Kopp, R., Pommerhene, W. and Schwartz, N. (eds.) (1997). *Determining the Value of Non-marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, Boston: Kluwer.

Levi, I. (1986). *Hard Choices: Decision Making Under Unresolved Conflict*, Cambridge: Cambridge University Press.

Mandler, M. (1998). 'The economics of incomplete preferences', mimeo, Harvard University.

Mandler, M. (2001). 'A difficult choice in preference theory: rationality entails completeness or transitivity but not both', in (E. Millgram, ed.) *Varieties of Practical Reasoning*, Cambridge: MIT.

Mandler, M. (2004). 'Incomplete preferences and rational intransitivity of choice', *Games and Economic Behavior*, (forthcoming).

Markowitz, H. (1952). 'The utility of wealth', *Journal of Political Economy*, vol. 60, pp. 151–8.

Milgrom, P. (1993). 'Is sympathy an economic value? Philosophy, economics, and the contingent valuation method', in Hausman (1993).

Ok, E. (2002). 'Utility representation of an incomplete preference relation', *Journal of Economic Theory*, vol. 104, pp. 429-49.

Rabin, M. (1998). 'Psychology and economics', *Journal of Economic Literature*, vol. 36, pp. 11–46.

Samuelson, W. and Zeckhauser, R. (1988). 'Status quo bias in decision making', *Journal of Risk and Uncertainty*, vol. 1, pp. 7–59.

Seip, K. and Strand, J. (1992). 'Willingness to pay for environmental goods in Norway: a contingent valuation study with real payment', *Environmental and Resource Economics*, vol. 2, pp. 91–106.

Sen, A. (1973). 'Behaviour and the concept of preference', *Economica*, vol. 40, pp. 241–59.

Sen, A. (1982). 'Introduction', in: *Choice, Welfare and Measurement*, Oxford: Blackwell.

Sen, A. (1997). 'Maximization and the act of choice', *Econometrica*, vol. 65, pp. 745–80.

Thaler, R. (1980). 'Toward a positive theory of consumer choice', *Journal of Economic Behaviour and Organization*, vol. 1, pp. 39–60.

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