P000152

preference reversal

Preference reversal is a widely observed behavioural tendency for the preference ordering of a pair of alternatives to depend on the process used to elicit it. The phenomenon appears to be both a robust and a systematic departure from conventional preference theory. Competing theoretical explanations variously interpret it as a violation of procedure invariance (the presumption that preferences should be independent of the method of eliciting them); a failure of transitivity; or a consequence of loss-averse (and reference-dependent) preferences. This article discusses these interpretations, the related evidence, and reflects on some of the broader implications of the phenomenon.

Preference reversal (PR) is a widely observed behavioural tendency for the preference ordering of a pair of alternatives to depend, in a predictable way, on the process used to elicit it.

The existence of preference reversal sets an empirical challenge to fundamental assumptions of conventional economic theory: PR is an apparent failure of procedure invariance (that is, the traditional presumption that preferences should be independent of the method of eliciting them). Some see it as a challenge to the very idea that human decisions are governed by preferences.

Much of the empirical PR literature has examined decisions relating to pairs of simple gambles. One of the gambles (typically called the ‘P-bet’) will offer a relatively good chance of winning a modest prize, otherwise nothing (or sometimes a small loss); the other bet (the ‘$-bet’), offers a relatively small chance of winning a larger prize. In classic PR experiments, subjects are required to make straight choices between such pairs of bets and to provide separate (usually monetary) valuations for each bet. For any individual and gamble pair, conventional economic theory implies that the chosen gamble would also be the more highly valued of the pair. But while many individuals are so consistent, a significant proportion, typically, are not. The existence of some such inconsistency, by itself, is not especially surprising. People might, for instance, make a mistake in one or more task, leading to some level of inconsistency in comparisons of rankings. Interest in PR, however, stems largely from the fact that observed inconsistencies tend to be patterned in a highly predictable way: the typical finding is that considerable numbers of subjects choose the P-bet and value the $-bet more highly (let us call this the standard reversal), while very few commit the opposite reversal ($-bet chosen and P-bet valued more highly). It is this *asymmetric* pattern of inconsistencies between rankings based on choice and valuation that constitutes the intriguing PR phenomenon.

Evidence

PR was first predicted and then observed by psychologists (Lichtenstein and Slovic, 1971; Lindman, 1971). It was later brought to the attention of economists by Grether and Plott (1979) who described its potential significance for economics in the following passage:

> Taken at face value the data are simply inconsistent with preference theory and have broad implications for research priorities within economics. The inconsistency is deeper than mere lack of transitivity or

> even stochastic transitivity. It suggests that no optimisation principles of any sort lie behind even the simplest of human choices. (Grether and Plott, 1979, p. 623)

Like many economists who have followed in their footsteps, Grether and Plott did not immediately accept this face-value interpretation and, instead, looked for ways of explaining PR while retaining the assumption that individuals do have a unique preference ordering over gambles. A substantial body of research in this spirit has examined whether PR might be an experimental artefact arising from imperfectly designed experiments. Early research of this genre – including Grether and Plott (1979); Reilly (1982); Pommerehne, Schneider and Zweifel (1982) – investigated issues such as whether PR might be a consequence of subjects failing to understand the tasks confronting them, or of having insufficient motivation to take those tasks seriously. But a large body of evidence now shows that PR is a highly replicable phenomenon, robust to many variations in experimental procedures. Seidl (2002) provides a review.

A more subtle critique of PR experiments and evidence emerged in the late 1980s with the publication of a series of theoretical papers (Holt, 1986; Karni and Safra, 1987; Segal, 1988) arguing that PR might be a spurious artefact of experimental design after all. These papers shared a common strategy, pointing to a potential weakness of two experimental procedures which had been commonly used to incentivize decision tasks in PR experiments: the Becker–DeGroot–Marschak (1964) mechanism and the random lottery incentive system. The thrust of these papers is to show that, if individuals have non-expected utility preferences (violating either the independence axiom of expected utility theory, or the reduction of compound lotteries principle, or both), these standard incentive mechanisms could be biased and might generate the spurious appearance of PR. On this interpretation, PR would not be evidence against procedure invariance: instead it would be evidence of consistent, but non-expected utility, preferences interacting with specific features of experimental design. This interpretation has, however, been largely discounted in the light of subsequent research (including Tversky, Slovic and Kahneman, 1990; Cubitt, Munro and Starmer, 2004) which reproduces the PR phenomenon in experiments using incentive mechanisms immune to this critique of earlier studies.

Theory

There remains considerable interest in trying to find a satisfactory explanation of PR. In what follows, we discuss three types of theory that may contribute to that objective: *regret theory*, *reference-dependent theory*, and *constructed preference theory*.

Regret theory (Loomes and Sugden, 1982; 1983) explains PR as a form of intransitivity. In this theory preferences are defined over pairs of acts which map from states of the world to consequences (as in Savage, 1954). Suppose A_i and A_j are two potential acts that result in, respectively, outcomes x_{is} and x_{js}, in state of the world s. If A_i is chosen, the resulting utility in each state is given by a 'modified utility function' $M(x_{is},x_{js})$. Notice that this function allows the consequences of the chosen act to depend upon those that *might have been* experienced under the forgone act A_j. In particular, the utility from having x_{is} may be suppressed by 'regret' when x_{is} is worse than x_{js}. Regret theory assumes that individuals attempt to maximize the expectation of modified utility $\Sigma_s\, p_s.M(x_{is},x_{js})$ where p_s is the probability of state s. Regret theory reduces to expected utility theory in the special case where $M(x_{is}, x_{js}) = u(x_{is})$ and $u(.)$ is a von Neumann–Morgenstern utility function.

Loomes and Sugden (1982) show that, if preferences in this theory satisfy particular restrictions, then regret theory provides a possible explanation of several well-known violations of expected utility theory including some cases of the famous Allais paradox. The most important of these restrictions is a property (subsequently) called regret aversion and, in a follow-up paper, Loomes and Sugden (1983) show that regret aversion may also explain PR. The argument works roughly as follows. Consider the following three acts labelled $, P and M with monetary consequences $x > y > m > 0$ defined over three states.

	State 1	State 2	State 3
$	x	0	0
P	y	y	0
M	m	m	m

The acts labelled $ and P have the structure of typical $- and P-bets: they are binary gambles where $ has the higher prize, and P the higher probability of 'winning'; the third act gives payoff m for sure. Regret theory allows choices over acts with this structure to be non-transitive and, if preferences are regret averse, if a cycle occurs it will be in a specific direction: P chosen over $; M over P; and $ over M. Now recall that, in a typical PR experiment, the standard reversal occurred when a subject chose P over $ but valued $ more highly than P. So, if we interpret choices from {$, M} and {P, M} as analogues of valuation tasks asking 'is $ (or P) worth more or less than m?', then the cycle predicted by regret theory can be interpreted as a form of PR.

This explanation for PR has been tested via experiments designed to look for the pure choice analogue of PR by confronting subjects with pairwise choices among triples of bets with the structure of $, P and M above. The outcome of this strand of research has produced good and bad news for regret theory. The good news is that the non-transitive choice cycles predicted by it have been observed and replicated (Loomes, Starmer and Sugden, 1991). Since these choice cycles occur in studies that involve no valuation tasks at all, this is evidence for the intransitivity interpretation of PR. The bad news is that subsequent research (Starmer and Sugden, 1998) has cast considerable doubt on regret theory's account of these choice cycles. The current state of play appears to be that regret theory has led to the discovery of a surprising new choice phenomenon, but it turns out not to be the right explanation for it! It remains possible that these intransitive choice cycles are manifestations of regret-type influences at work but that formal models of regret must be refined to properly account for them. Another possibility is that they have nothing to do with 'regret' and that their discovery, as a consequence of testing regret theory, was just accidental.

A new account of PR has emerged in the form of reference-dependent subjective expected utility theory (Sugden, 2003). In this model, preferences are again defined over acts. The key structural departure from Savage's (1954) subjective expected utility theory is that consequences in each state are modelled as gains and losses relative to a *reference act* (the status quo). The resulting theory is a formulation of expected utility (that is, a model that is linear in probabilities) that can accommodate loss aversion (that is, losses of a given size being weighted more highly than corresponding magnitude gains). Sugden demonstrates that, when preferences are loss averse, this model predicts standard PR in experiments where values are elicited as selling prices (which they usually are). This prediction depends on the assumption that, in selling tasks, an agent's reference act is the lottery being sold: given this, seemingly reasonable, assumption, $ valuations become particularly 'inflated' by consideration of the large $ prize which becomes a (probabilistic) loss if the $-bet is given up for a certain amount of cash. Hence, on this account, PR is the consequence of loss aversion operating through selling tasks. As yet, there have been no direct tests of this explanation, though the evidence of loss aversion operating in other contexts (see Starmer, 2000, for some discussion) perhaps gives it some initial credibility.

Thus far we have discussed various preference-theoretic accounts of PR. The final type of explanation we discuss is the oldest and belongs to a class of theory that has evolved in the psychology literature. From the outset, most psychologists accepted PR as evidence against the very thing that economists have invested their efforts in defending: the presumption that behaviour can be adequately explained in terms of unique underlying preferences. Psychologists have, instead, focused on accounts of PR which attribute it to aspects of human *decision processes*. Viewed from this perspective, there is nothing fundamentally surprising about the fact that rankings delivered via choice and valuation tasks differ; those working within this paradigm will, typically, attempt to read such inconsistencies as clues to the, potentially distinct, mental heuristics invoked in those different tasks.

Numerous theories in this spirit have been proposed as putative accounts of PR, and one of the best known examples is the scale-compatibility hypothesis due to Tversky, Sattath and Slovic (1988). The general hypothesis assumes that the way in which an individual is required to respond to a task ('the response mode') can affect the weights that he or she places on particular dimensions of alternatives being evaluated. In application to PR, the hypothesis implies that, because valuation tasks require a money amount as output, individuals place particularly high (low) weight on the money (probability) dimension, leading to relatively 'inflated' values for $ bets. Some recent support for this particular hypothesis is reported in Cubitt, Munro and Starmer (2004). There is, however, a vast theoretical and empirical literature connecting PR with the constructed preference approach and, for those interested in pursuing it, an excellent source is Lichtenstein and Slovic (2006).

Developing themes

One developing theme in empirical PR research examines the persistence of PR in environments where individuals receive feedback on the consequences of their decisions. A famous experiment by Chu and Chu (1990) exposed preference reversers to 'money pumps': subjects who committed PR had their stated preferences implemented across a series of trades which ultimately resulted in monetary losses. Individuals quickly learned to avoid PR in this environment. While this is an interesting finding, since Chu and Chu use such an explicit method for disciplining inconsistent preferences, it would be a mistake to view this as persuasive evidence that PR would be eroded in any naturally occurring market. There is some limited evidence to suggest that PR may decay in some specific experimental markets (Cox and Grether, 1996) but the findings here are both tentative and mixed, and further investigation is warranted before any firm conclusions can be drawn.

Another theme of current research explores the implications of preference anomalies (including PR) for the formulation of economic policy. A discussion of this topic is contained in Braga and Starmer (2005).

Chris Starmer

See also

<xref=xyyyyyy> adaptive heuristics;
<xref=xyyyyyy> Allais paradox;
<xref=E000178> expected utility hypothesis;
<xref=xyyyyyy> paradoxes and anomalies;
<xref=xyyyyyy> preferences;
<xref=xyyyyyy> prospect theory;
<xref=xyyyyyy> rational behaviour;
<xref=xyyyyyy> rationality;
<xref=xyyyyyy> bounded rationality;
<xref=xyyyyyy> Savage's subjective expected utility model;
<xref=xyyyyyy> transitivity.

Bibliography

Becker, G.M., DeGroot, M.H. and Marschak, J. 1964. Measuring utility by a single-response sequential method. *Behavioral Science* 9, 226–32.

Braga, J. and Starmer, C. 2005. Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics* 32, 55–89.

Chu, Y.P. and Chu, R.L. 1990. The subsidence of preference reversals in simplified and marketlike experimental settings: a note. *American Economic Review* 80, 902–11.

Cox, J.C. and Grether, D.M. 1996. The preference reversal phenomenon: response mode, markets and incentives. *Economic Theory* 7, 381–405.

Cubitt, R.P., Munro, A. and Starmer, C. 2004. Testing explanations of preference reversal. *Economic Journal* 114, 709–26.

Grether, D. and Plott, C.R. 1979. Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69, 623–38.

Holt, C.A. 1986. Preference reversals and the independence axiom. *American Economic Review* 76, 508–15.

Karni, E. and Safra, Z. 1987. 'Preference reversal' and the observability of preferences by experimental methods. *Econometrica* 55, 675–85.

Lichtenstein, S. and Slovic, P. 1971. Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89, 46–55.

Lichtenstein, S. and Slovic, P. 2006. *The Construction of Preference*. New York: Cambridge University Press.

Lindman, H.R. 1971. Inconsistent preferences among gambles. *Journal of Experimental Psychology* 89, 390–97.

Loomes, G., Starmer, C. and Sugden, R. 1991. Observing violations of transitivity by experimental methods. *Econometrica* 59, 425–39.

Loomes, G.C. and Sugden, R. 1982. Regret theory: an alternative theory of rational choice under uncertainty. *Economic Journal* 92, 805–24.

Loomes, G.C. and Sugden, R. 1983. A rationale for preference reversal. *American Economic Review* 73, 428–32.

Pommerehne, W.W., Schneider, F. and Zweifel, P. 1982. Economic theory of choice and the preference reversal phenomenon: a re-examination. *American Economic Review* 73, 569–74.

Reilly, R.J. 1982. Preference reversal: further evidence and some suggested modifications in experimental design. *American Economic Review* 73, 576–84.

Segal, U. 1988. Does the preference reversal phenomenon necessarily contradict the independence axiom? *American Economic Review* 78, 233–36.

Seidl, C. 2002. Preference reversal. *Journal of Economic Surveys* 6, 621–55.

Savage, L. 1954. *The Foundations of Statistics*. New York: Wiley.

Starmer, C.V. 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* 38, 332–82.

Starmer, C. and Sugden, R. 1998. Testing alternative explanations of cyclical choices. *Economica* 65, 259–347.

Sugden, R. 2003. Reference-dependent subjective expected utility. *Journal of Economic Theory* 111, 172–91.

Tversky, A., Sattath, S. and Slovic, P. 1988. Contingent weighting in judgement and choice. *Psychological Review* 95, 371–84.

Tversky, A., Slovic, P. and Kahneman, D. 1990. The causes of preference reversal. *American Economic Review* 80, 204–17.

Index terms

Allais paradox
decision processes
expected utility hypothesis
intransitivity
loss aversion
preference reversal
preferences
procedure invariance
regret
Savage's subjective expected hypothesis

Index terms not found:

expected utility hypothesis
Savage's subjective expected hypothesis