Journal of Health Economics 32 (2013) 1250–1257

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Testing the rate of preference reversal in personal and social decision-making

Adam Oliver*

LSE Health and Social Care, and the Department of Social Policy, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

ARTICLE INFO

Article history:
Received 3 July 2013
Accepted 4 October 2013
Available online 17 October 2013

JEL classification:
C91
D81
I10

Keywords:
Preference reversals
Choice
Valuation
Personal preference
Social preference

ABSTRACT

Classic preference reversal, where choice and valuation procedures generate inconsistent preference orderings, has rarely been tested in hypothetical health care treatment scenarios. Two studies – the first non-incentivised and the second incentivised – are reported in this article. In both studies, respondents are asked to make decisions that affect themselves (a personal decision making frame) and those for whom they are responsible (a social decision making frame). The results show non-negligible and systematic rates of preference reversal in both frames, although these rates are slightly, but non-significantly, lower in the incentivised condition. Moreover, in both studies, the rate of predicted preference reversal was somewhat higher in the social than in the personal decision making frame, a finding that is explained by greater risk aversion when choosing treatment options for others than when choosing treatments for oneself.

1. Introduction

1.1. Background

Evidence of preference reversal, whereby a respondent chooses one good over another but places a higher monetary value on the non-chosen good, have been reported in the economics and psychology literatures for more than forty years (for a review, see Seidl, 2002). These preference reversals appear to violate procedural invariance and, in contrast to the assumptions of standard economic theory, demonstrate that people may not always hold fixed and stable preferences, but might rather construct their preferences according to how a decision problem is framed, or to the elicitation procedure that they face.

Classically, the preference reversal phenomenon involves two bets, commonly termed the $-*bet* and the *P-bet*. The $-*bet* offers a modest probability of winning a relatively large amount, and the *P-bet* offers a high probability of winning a modest amount; the two bets have similar expected values. In a direct choice between the two bets, many respondents tend to choose the *P-bet*, but then when asked to value the bets, they place a higher monetary value on the $-*bet*. Early work on this phenomenon was reported by the psychologists Lichtenstein and Slovic (1971, 1973), who, to illustrate, presented respondents with the following bets:

$-*bet:* ($16, 11/36; −$1.50, 25/36)
P-bet: ($4, 35/36; −$1, 1/36)

Here, the $-*bet* offers an 11/36 chance of winning $16 and a 25/36 chance of losing $1.50. The *P-bet* can be similarly read. Lichtenstein and Slovic (1971) undertook three tests and reported that between 51% and 83% of their respondents reversed their preferences in the *predicted* direction of choosing the *P-bet* but placing a higher value on the $-*bet*. The opposite – *unpredicted* – preference reversal pattern, such that respondents chose the $-*bet* but placed a higher value on the *P-bet*, occurred at a rate of only 6–27%. Such a systematic (i.e. predominantly unidirectional) pattern of reversed preferences cannot be attributed to random error.

Initially, economists tended to be sceptical of the preference reversal phenomenon, and somewhat unfairly argued that the psychologists had insufficiently incentivised their respondents. However, a carefully conducted study by Grether and Plott (1979) revealed that, if anything, the predicted preference reversal rate increased after the inclusion of what they deemed to be appropriate incentives. The preference reversal phenomenon has since proven to be a robust finding, observed in a multitude of studies

* Tel.: +44 020 7955 6471.
E-mail address: a.j.oliver@lse.ac.uk

0167-6296/$ – see front matter
http://dx.doi.org/10.1016/j.jhealeco.2013.10.003

(see Seidl, 2002), with that undertaken by Tversky et al. (1990) being reasonably representative with predicted and unpredicted preference reversal rates of 45% and 4%, respectively. Moreover, preference reversals have been observed across outcome domains that transcend monetary lotteries, including, for example, time dimensions (see the second study reported in Tversky et al., 1990), environmental goods (Irwin et al., 1993), and income distributions (Camacho-Cuena et al., 2005).

There is also evidence that respondents sometimes reverse their preferences across different elicitation procedures in the context of health (e.g. Bleichrodt and Pinto Prades, 2009; Gyldmark and Morrison, 2001; Olsen, 1997; Olsen and Donaldson, 1997; Robinson et al., 1997; Shackley and Donaldson, 2002) – for example, inconsistencies are often found between respondents' willingness to pay (WTP) for health care programmes and their ordinal ranking of those same programmes. Shackley and Donaldson (2002) speculate that this inconsistency may be caused by respondents basing their WTP answers on their estimated cost of the interventions, causing them to give high values for costly, relatively undesired options. Gyldmark and Morrison (2001) attribute at least some of the observed inconsistency to the fact that respondents often fail to distinguish between interventions in terms of the WTP amounts they offer. For instance, respondents are regularly found to not want to pay anything for the interventions, a form of protest response towards options that they believe ought to be free at the point of use, or due to an intention to free ride, both also noted by Irwin et al. (1993) in relation to valuing environmental goods. Other respondents assign identical non-zero WTP values to different health interventions in order to gain the moral satisfaction of offering a fixed amount towards something they perceive as 'good' irrespective of their relative preferences for different goods; in essence, the 'warm glow' effect. Along these lines, Ajzen and Peterson (1988) have criticised WTP techniques for capturing attitudinal intentions rather than behavioural preference, a problem that may be avoided with ordinal ranking.

Health care treatment options can also be framed as lotteries, and therefore this domain is ripe for testing preference reversals within the structure of the classic design outlined earlier. Perhaps the closest replication of the classic test in health was undertaken by Oliver (2006), although rather than focussing on health care interventions, he asked respondents to choose between two countries with different life expectancy distributions, and to give their certainty equivalents in terms of life expectancy as a means of placing a value on the options. The life expectancy distributions in the two countries were as follows:

$-*bet:* (64 yrs, 70%; 84 yrs, 30%)
P-bet: (65 yrs, 3%; 70 yrs, 97%)

Here, the $-*bet* is a hypothetical country with a modest percentage of people enjoying a long life expectancy, with 30% of the population expecting to live to 84 years of age and 70% expecting to live for 64 years. The *P-bet* offers a large percentage of the population a modest life expectancy, and can be read similarly. Oliver (2006) observed that 36% of his respondents demonstrated a strict preference reversal, with all of those offering the predicted pattern of preferring the *P-bet* over the $-*bet*, but then valuing the $-*bet* higher. More recently, Oliver (2013) attempted to test for preference reversals over options defined as health care treatment options, but, probably due to the complicated presentation of the options, non-systematic, although still quite substantial (measuring 35–40%), preference reversal rates were observed, with respondents often seemingly constructing their answers according to the design of the tasks that they faced.

1.2. Explanations for classic preference reversals

In order to elicit respondents' monetary valuations, it is common to ask them for their minimum selling prices of the bets that they face. To do this, economists have traditionally insisted that an incentive compatible elicitation method be used; to these ends, economists have often preferred to use the Becker–DeGroot–Marschak (BDM) procedure (Becker et al., 1964). This method requires the respondent to state her minimum selling price for a bet (which can be thought of as a lottery ticket). Through a random device, an offer price is then generated for the bet. The respondent receives the offer price if it exceeds her minimum selling price but plays the bet otherwise. In theory, the BDM procedure incentivises the respondent to state her 'true' minimum selling price: if she states a selling price higher than her true minimum, the random device might generate an offer price that falls between her true minimum and her stated minimum, and she will end up playing the bet when she would have preferred to have accepted the offer.

Theoretically, the BDM procedure requires non-violation of the axioms of expected utility theory (EUT), and, in particular, the independence axiom, in respondents' decision making. We have known for a long time, however, that independence does not always hold (e.g. see Allais, 1990). This has led some to contend that preference reversals may be attributable to use of the BDM method (Holt, 1986; Karni and Safra, 1987), but, as noted by Tversky et al. (1990), observations of substantial and systematic predicted preference reversals are as prevalent in studies that explicitly do not use a value elicitation method that requires independence (e.g. Tversky et al.'s own study; Cubitt et al., 2004; Lichtenstein and Slovic, 1971) as they are in those that do (e.g. Grether and Plott, 1979). The contamination effect on the rate of preference reversals caused by possible violations of independence when using the BDM procedure therefore appears weak at most (Camacho-Cuena et al., 2005).

A second possible cause of preference reversals is intransitivity in respondents' answers. That is, let *X* be a monetary amount that falls above the valuation of the *P-bet* but below the valuation of the $-*bet* for a respondent demonstrating predicted preference reversal. It might then be hypothesised that *P-bet* > $-*bet* > *X* > *P-bet*, where > is her preference relation, 'is preferred to', in a series of direct choice tasks involving that respondent. However, several studies have found that intransitivity in respondent choice can explain only 10–20% of observed preference reversals (Cox and Grether, 1996; Loomes et al., 1989; Tversky et al., 1990).

A third, and most broadly accepted, explanation for predicted preference reversals is that procedural invariance is violated; in other words, different decision procedures produce inconsistent preference orderings. Tversky et al. (1988) posit two hypotheses as the drivers behind procedural invariance violations. First, and probably most importantly with respect to preference reversals, there is the scale compatibility hypothesis, which, according to Thaler and Tversky (1992), itself has a twofold rationale: (i) if the attributes of an option (e.g. outcomes, probability) and the response mode (e.g. choice, valuation) do not match, additional mental operations are required, which increases effort and errors; (ii) the response mode tends to focus attention on (i.e. causes anchoring on) the attributes of an option that are compatible with that mode (e.g. monetary valuation tends to focus respondents' attentions on money outcomes), which can in turn cause overpricing of the $-*bet*. A probable example of the latter in the context of health can be observed in Stalmeier et al. (2001), who found that many respondents (45% of their sample) prefer to live for ten rather than twenty years if each week involves four and a half days of migraine, and yet place a greater healthy year equivalent (i.e. a certainty equivalent measured in life