

Social Preferences and Reference Points

0500866

April 30, 2009

1 Introduction

The past decade has seen a tremendous growth in behavioural models which move beyond the self-centred nature of classical utility theory and allow the payoffs of other agents to affect an individual's utility. These can explain much behaviour seen in simple allocation games, however they omit the role of reference points and expectations in evaluating payoffs. Gains and losses relative to a reference point are often regarded as important for entirely selfish individuals (Tversky & Kahneman 1991) and there is no reason to expect that this is not so for social preferences.

As an example of reference-dependence, recent gains and losses experienced by individuals may affect the way in which payoffs are perceived. Suppose that individual i is endowed with money, and is considering how much to allocate to individual j : laboratory experiments show that many give a positive amount. Now consider a scenario with identical endowments but where j recently lost money due to some event outside their control. Introspection suggests that i may be more willing to give a greater sum to j to compensate for this loss, despite the two situations being identical in terms of asset position. For a real-life example of apparent reference-dependence in inter-personal preferences, consider the observation that donations to charity are typically towards those experiencing relative poverty (i.e. an income below some reference point for their society) and not just those with the lowest absolute incomes¹.

This paper seeks to create a parsimonious model to capture such behaviour and permit tests against more traditional models through simple experiments. In part II, we briefly review the relevant literature. This brings together two strands of research: models of *social preferences* in which individuals take other agents' payoffs into account, as well as the importance of *reference points*. It is shown how the social preferences model bears many similarities to a multidimensional reference-dependent model. Section III derives a specification which adds reference-dependence. In section IV, the main issues facing an experimental test of the model are discussed.

2 Literature Review

The difference-aversion model of Fehr & Schmidt (1999) adds dislike of inequality to the standard utility function. Players suffer from both disadvantageous inequality (being worse off than others), and also to a lesser extent advantageous inequality. With just two players, the model is written as follows:

$$U_i(x_i, x_j) = x_i - \alpha \max\{x_j - x_i, 0\} - \beta \max\{x_i - x_j, 0\}$$

U_i denotes the utility of player i , x_i refers to the monetary payoff of i , and α and β are the weights attached to disadvantageous and advantageous equality respectively. It is assumed that $\alpha_i \geq \beta_i$ (individuals suffer at least as much from disadvantageous inequality as advantageous inequality) and that $0 \leq \beta_i < 1$ (players do not prefer to be better off than others, however their aversion to advantageous inequality is dominated by their "selfish" utility from a monetary gain).

Charness & Rabin (2002) nest difference aversion within a more general model of social preferences (henceforth referred to as the "C&R model"). The two-player version is as follows:

¹This effect is of course confounded by the way in which individuals in wealthy countries may prefer to give to the poor in the same country rather than elsewhere. There has been little research done so far into the role of relative poverty and reference points in altruism, and experimental methods such as those outlined in this paper provide a possible way in which such effects can be isolated and tested.

Parameters	Resulting preferences
$\rho = \sigma = 0$	“Narrow self-interest”: player only cares about their payoff, as in the classical utility framework
$\sigma \leq \rho \leq 0$	“Competitive preferences”: player prefers payoffs to be as high as possible compared to other player
$\sigma < 0 < \rho < 1$	“Difference aversion”: as in Fehr & Schmidt (1999)
$0 < \sigma \leq \rho \leq 1$	“Social welfare preferences”: as in Andreoni & Miller (2002) players prefer to maximise all players’ payoffs, but are more concerned about their own payoffs if behind

$$U_i(x_i, x_j) = (\rho \cdot r + \sigma \cdot s + \theta \cdot q) \cdot x_j + (1 - \rho \cdot r - \sigma \cdot s - \theta \cdot q) \cdot x_i$$

r and s are indicator variables which are only equal to 1 if the monetary payoff of player i is ahead or behind player j respectively. The variable q is used to capture reciprocity and is set to -1 if the other player has “misbehaved” in the past. In what follows we ignore reciprocity, which can generally be controlled for in experiments. Utility is thus a weighted sum of each player’s payoffs. By setting constraints on the parameters σ and ρ , the following types of preferences can be modelled:

One notable limitation of these specifications is the linearity of the utility functions; other than the jump exhibited when i moves behind or ahead of j , the marginal utilities and marginal rate of substitution between their own wealth and that of the other player remains constant. This seems unrealistic; would an agent really gain the same utility when the other player’s wealth increases by a pound regardless of just how far ahead/behind they are? In the games studied in the literature, such issues are unlikely to be important: the utility function is likely to be almost linear when payoffs are a small fraction of total wealth (although the analysis applied generally assumes total wealth is *not* considered).

These models are applied to variants of the “Dictator Game” where a player picks one of two possible allocations for themselves and a second individual who has no input. Such games should remove the motivation for picking an allocation based on fear of retaliation (Hoffman, McCabe & Smith (1996) notes how individuals may still bring in repeated-game experience, although this effect can be minimised with careful experimental design). It is found that the social welfare preferences model consistently explains more decisions than the alternatives. ρ is found to be highly significant and equal to approximately 0.4, while σ is insignificant. This suggests a weak form of social welfare preferences or difference aversion where player i attaches some weight to j ’s payoff only when j is behind. Engelmann & Strobel (2004) carry out further experiments which generally support this interpretation of the C&R model over difference aversion.

Prospect Theory, introduced by Kahneman & Tversky (1979), suggests that individuals gain utility from changes in their asset position rather than absolute levels (prospect theory is generally applied to decisions under uncertainty, however we are only concerned about the reference-dependence it introduces). The *value function* yields the utility of deviations from a given “reference point” of assets, and generally exhibits concavity for gains, convexity for losses, and a steeper gradient for losses than gains (loss aversion). A typical value function is shown in figure 1.

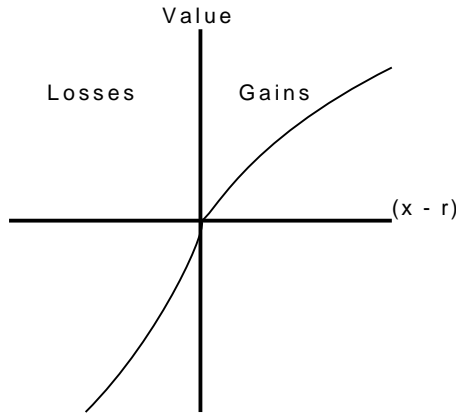


Figure 1: Prospect theory value function. X-axis measures deviations from reference point.

What determines a person’s reference point for their wealth? Possibilities include:

- **Current asset position** - the reference point is commonly assumed to coincide with the endowment of assets (Tversky & Kahneman 1991), which together with loss-aversion produces status-quo bias in riskless choice
- **Delays in integrating recent losses/gains** - Kahneman & Tversky (1979), in the context of choice under uncertainty, notes that the reference point may lag the current level of wealth. This is the relevant effect for the two-person example in the introduction
- **Expectations** - for example an expected pay rise, as in Kahneman (1992)
- **Social comparisons** - we argue below that the C&R model is a reference-dependent model where one agent’s payoff is the reference point for another
- **Foregone choices** - the role of regret is emphasised by Bell (1982) among others, and could affect the reference points in an experimental situation

While the C&R model does not explicitly take into account the reference-dependent nature of preferences, Fehr & Schmidt (1999) posit that, in the laboratory situation where all participants enter as equals, the egalitarian outcome forms a reference point. With just two agents, this implies that i ’s reference point for i ’s (j ’s) payoff is j ’s (i ’s) payoff. Figure 2 plots i ’s utility in the C&R model (assuming $0 < \sigma \leq \rho \leq 1$ and holding one player’s payoff at a constant positive level) where the horizontal axis is the difference between the two player’s payoffs and hence measures deviations from this reference point. The parallels to Prospect Theory are clear and although the linear nature of preferences rules out convexity/concavity, a form of loss aversion is present (an agent’s payoff has a higher marginal utility if it is below the equitable reference point). The key difference is the way in which the C&R “value function” moves vertically with changes in the asset position of the other agent. This is since, while the relative weight attached to each agents’ payoff is determined by if it is above or below the reference point (as in prospect theory), this is multiplied by the payoff itself rather than the difference between them. In this situation where the reference point of one asset is the quantity of the other, the usual value function would imply that increasing one of the two agents’ payoffs would result in disutility ².

Unlike reference-dependent models which simply include one dimension of payoffs, this model includes two attributes (the wealth of each agent). There are two broad approaches which have been taken to extend reference-dependent preferences to multiple dimensions (Bleichrodt, Schmidt & Zank 2008): *Holistic evaluation* (the different attributes are somehow combined and weighed against a single reference point) and *attribute-specific evaluation* (gains and losses on each attribute are evaluated separately

²Consider the value function $\alpha(x_i - x_j) + \beta(x_j - x_i)$ where α and β are constants that vary with the ordering of payoffs. It can be seen that the coefficient on x_i is $\alpha - \beta$ while that on x_j is $\beta - \alpha$. Clearly, if one of these is positive then the other is negative.

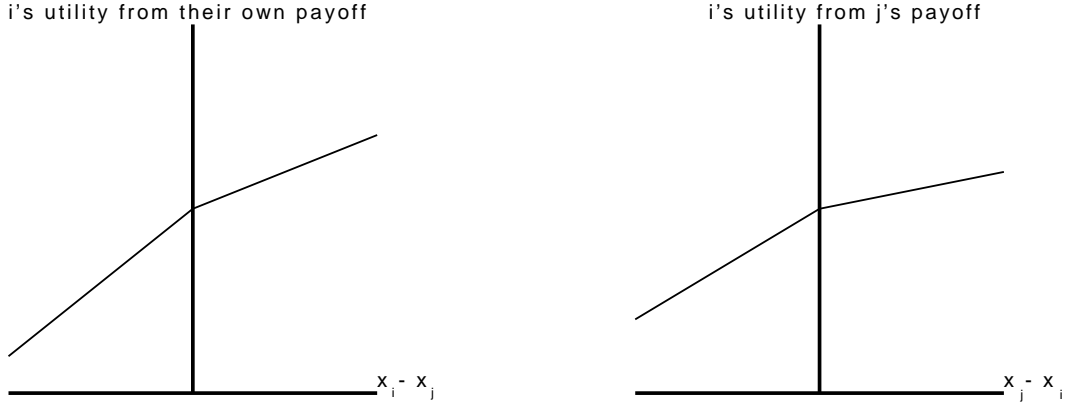


Figure 2: Utility from each agents' payoff in C&R model. X-axes measure deviation from other agent's payoff.

against distinct reference points). Much literature follows the latter route, including Tversky & Kahneman (1991) which includes a simple “additive constant loss aversion” specification that omits diminishing sensitivity:

$$U_r(x_1, x_2) = R_1(x_1) + R_2(x_2)$$

$$R_i(x_i) = \begin{cases} u_i(x_i) - u_i(r_i) & \text{if } x_i \geq r_i \\ (u_i(x_i) - u_i(r_i))/\lambda_i & \text{if } x_i < r_i \end{cases}$$

With a linear utility function u_i , this resembles the C&R model with the differences noted above.

We require two further reference points (one per agent). There are again two ways in which multiple reference points have been modelled (Ordez, Connolly & Coughlan 2000). The reference points may be *integrated*, where the payoff is evaluated against a weighted average of reference points or *separated*, where the payoff is compared against different reference points and an overall judgement then made. Ordez et al. (2000) examine multiple reference points for graduate salaries and find that subjects' responses are mainly indicative of separation. As the two reference points we are considering (for player i 's wealth they are j 's wealth and an arbitrary reference point that may reflect i 's expectations of their own wealth) are more differentiated in nature than salaries of different individuals, separation seems even more likely. Inman, Dyer & Jia (1997) utilise a separating model where deviations from the two reference points are multiplied by a constant (which can be different for losses and gains) and summed. This forms the basis for our model.

3 Model

The C&R specification can be re-written in a form that resembles a prospect theory value function, including loss-aversion:

$$U_i(x_i, x_j) = \lambda(x_j) + (1 - \lambda)x_i$$

Where:

$$\lambda = \begin{cases} \rho & \text{if } x_i \geq x_j \\ \sigma & \text{if } x_i < x_j \end{cases}$$

We now incorporate deviations of both agents' payoffs from their respective reference points (r_i and r_j), using the linear specification outlined above:

$$U_i(x_i, x_j, r_i, r_j) = \lambda E_i(x_j) + (1 - \lambda)x_i + \alpha(E_i(x_j) - r_j) + \beta(x_i - r_i)$$

Where α and β are larger (smaller) if the relevant agent's payoff is below (above) the reference point.

Note that the payoffs and reference points include each individual's *entire* wealth. Given that in many situations (such as in an experiment) i will not know j 's entire asset position, we have explicitly indicated that the utility function should include i 's expectation of j 's wealth. As in the original model, the weight attached to agents' payoffs depends only on whether they are above/below the reference points but not the magnitude of such deviations.

Key limitations of this model include this linearity and the way in which there is no trivial generalisation to more than two individuals. By incorporating the C&R model, however, it can account for a wide variety of social preferences, and additional reference points could be added in a simple fashion.

The main implication is that i 's marginal utility from increasing an agent's (including themselves) payoff depends on whether they are above or below the reference point. In the dictator-game setting, we would expect a player to behave more generously if they are above their reference point, or the other player is below theirs. Experiments can enable us to determine whether this effect exists, its magnitude, and the determinants of the second reference point (the nature of which may itself affect the utility function).

4 Experimental considerations

As briefly described above, the social welfare model has been validated through variants of the "dictator game". Most such studies implicitly assume that individuals only consider payoffs within the laboratory environment (although Sobel (2005) notes the possibility that individuals may consider redistributing money *after* the experiment, the overall wealth of players, or even the experimenter's welfare). If players behave like this, it is reasonable to assume that their reference point for the payoff of all participants is equal to zero. Given that lab payoffs are never negative, the additional reference-dependent terms we postulate will simply act as additional weights on each players' payoff: such games can therefore neither prove nor disprove any reference-dependent component of preferences.

In order to test for the existence of reference-dependent effects, it is necessary for the experimenter to manipulate subjects' reference points independently of final outcomes, such that the payoffs can be above or below the reference points. Subjects could be provisionally allocated a set amount of money prior to playing, the level of which alters whether a given allocation is a "gain" or "loss". Alternatively, to test whether lagged endowments enter the reference point, agents' wealth could be raised or lowered at random before the dictator makes their decision. This does, however, rely on agents integrating the initial endowment (which may seem random from the agent's perspective) into their reference point, but not the subsequent random gain/loss. Individuals in the experimental setup may continue to consider any payoff a gain and there is no way for the experimenter to directly verify whether reference points have changed.

If we could manipulate reference points in this fashion, the most direct way to test for their effects would be to look for differences in choices made across decisions where allocations have the same final payoffs but are evaluated with respect to different reference points. An example is shown below where reference points are altered by modifying endowments. Note that in both cases the subject is choosing between final payoffs of (13, 10) and (15, 5), however the intended reference points are (10,10) in the first decision and (10,0) in the second (figures in brackets correspond to i and j respectively). By re-framing the 5 as a loss to j rather than a gain, we would expect players facing the second decision to gain relatively less utility from this option and potentially change their choice from (15, 5) to (13, 10).

Decision 1	Decision 2
i given 10	i given 10
j given 10	j given 0
i Chooses between:	
1) i gains 3, j stays on same payoff	1) i gains 3, j gains 10
2) i gains 5, j loses 5	2) i gains 5, j gains 5

Repeating questions in this way makes the equivalence of the two outcomes easily apparent and may affect reference points, however, which is likely to affect subjects' responses (certainly, the ordering of the decisions should be varied across individuals).

While subjects in Charness & Rabin (2002) make multiple decision, they use two forms of analysis which could both be applied to an experiment where participants only make one decision. Firstly, some allocation decisions can unambiguously rule out or confirm certain forms of preferences (e.g. if an individual chooses to give more money to another player who is already ahead, they are not difference-averse). It is not possible to determine anything about the reference-dependent nature of preferences from a single choice however³. An extension to more than two agents allows reference points to differ across potential beneficiaries (e.g. does the dictator choose to give to someone who has recently lost money?) but is beyond the scope of the simple model utilised here.

The second type of analysis undertaken assumes that subjects have identical preferences and maximise them with error. Maximum likelihood estimation can then be applied to the dataset to determine the underlying parameters. This appears to be the most fruitful avenue for research regarding reference-dependent preferences.

5 Conclusions

It has been demonstrated how models of interpersonal preferences can be interpreted as incorporating both reference-dependence and loss-aversion. Our model further allows the weight that individuals place on payoffs to depend upon whether payoffs are above or below arbitrary reference points.

There are certainly many difficulties in designing an experiment to quantitatively estimate the parameters of such a model. Fundamentally, there is no way of directly measuring any reference points which a person may have, and it is surely impossible for an experimenter to accurately manipulate them without a complete theory of how they are formed. Nevertheless, conventional models of inter-personal preferences predict that individuals facing multiple choices which are identical in terms of final outcomes should make the same decision. This assertion can be tested relatively easily, via modifications of the Dictator Game as suggested above.

³To test agents' preferences across social payoffs, we just need to ensure that the two choices differ sufficiently in the allocations to each player. To isolate for reference-dependent effects, we would need choices in which the allocations themselves are identical but different in relationship to the reference point. This requires setting a different reference point for each allocation choice, but reference points are by their nature the same across different choices available at the same time.

References

- Andreoni, J. & Miller, J. (2002), 'Giving according to garp: An experimental test of the consistency of preferences for altruism', *Econometrica* **70**(2), 737–753.
- Bell, D. E. (1982), 'Regret in decision making under uncertainty', *Operations Research* **30**(5), 961–981.
- Bleichrodt, H., Schmidt, U. & Zank, H. (2008), Additive utility in prospect theory, The School of Economics Discussion Paper Series 0811, Economics, The University of Manchester.
- Charness, G. & Rabin, M. (2002), 'Understanding social preferences with simple tests', *The Quarterly Journal of Economics* **117**(3), 817–869.
- Engelmann, D. & Strobel, M. (2004), 'Inequality aversion, efficiency, and maximin preferences in simple distribution experiments', *The American Economic Review* **94**(4), 857–869.
- Fehr, E. & Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics* **114**(3), 817–868.
- Hoffman, E., McCabe, K. & Smith, V. L. (1996), 'Social distance and other-regarding behavior in dictator games', *The American Economic Review* **86**(3), 653–660.
- Inman, J. J., Dyer, J. S. & Jia, J. (1997), 'A generalized utility model of disappointment and regret effects on post-choice valuation', *Marketing Science* **16**(2), 97–111.
- Kahneman, D. (1992), 'Reference points, anchors, norms, and mixed feelings', *Organizational Behavior and Human Decision Processes* **51**(2), 296 – 312. Decision Processes in Negotiation.
- Kahneman, D. & Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**(2), 263–291.
- Ordez, L. D., Connolly, T. & Coughlan, R. (2000), 'Multiple reference points in satisfaction and fairness assessment', *Journal of Behavioural Decision Making* **13**, 329–344.
- Sobel, J. (2005), 'Interdependent preferences and reciprocity', *Journal of Economic Literature* **43**(2), 392–436.
- Tversky, A. & Kahneman, D. (1991), 'Loss aversion in riskless choice: A reference-dependent model', *The Quarterly Journal of Economics* **106**(4), 1039–1061.