Workshop: Strong reciprocity and welfare (scientific goal)

In economics, it has become customary to distinguish between self-regarding and other-regarding preferences. Individuals have self-regarding preferences if they derive utility from their own (monetary or non-monetary) payoffs; and they derive utility from their own monetary payoffs if they sell assets at a profit, are in good health, achieve professional goals, enjoy good food etc. Individuals have other-regarding preferences if they exhibit (advantageous or disadvantageous) inequity aversion (Fehr & Schmidt 1999) or strong reciprocity (Falk & Fischbacher 2005). They exhibit advantageous inequity aversion if they derive disutility from the difference between their own payoff and that received by the worst-off individual, disadvantageous inequity aversion if they derive disutility from the difference between the payoff received by the best-off individual and their own payoff, and strong reciprocity if they derive utility from increasing the payoff of individuals they experience as kind, and from decreasing the payoff of individuals they experience as unkind, potentially at a personal cost.

Prima facie the satisfaction of other-regarding preferences qualifies as welfare: individuals fare well when preferences in the shape of inequity aversion or strong reciprocity are satisfied. It has rarely been noted, however, that a standard welfare economic interpretation of the satisfaction of preferences in the shape of strong reciprocity is not straightforward. While preferences in the shape of inequity aversion satisfy the weak congruence axiom (the properties of completeness, transitivity, and reflexivity), and while the central concepts and tools of standard welfare economic (individual welfare optimum, Pareto optimum, compensating variation) are defined for these preferences, it is unclear whether preferences in the shape of strong reciprocity satisfy the weak congruence axiom, or whether the central concepts and tools of standard welfare economic are defined for these preferences.

This lack of clarity is unfortunate because empirical evidence deriving from laboratory experiments suggests that individuals do not only exhibit inequity aversion, but also strong reciprocity. There is also field evidence suggesting that individuals do not fare well when preferences in the shape of strong reciprocity remain unsatisfied: when behavior that is unkind in the sense of not conforming to "right" institutions (of behavior like tax evasion) or of conforming to "wrong" institutions remains unpunished. Case & Deaton (2020, pp. 11-12) refer to behavior of that sort when describing "the rent-seeking by pharma, by healthcare more generally, and by banks and many small- or medium-size business entrepreneurs, such as doctors, hedge fund managers, the owners of sports franchises, real estate businesspeople, and car dealers" as the behavior of people who "get rich from the 'oppressive monopolies' and special deals, tax breaks, and regulations that they have 'extorted from the legislature'", or who "are allowed to enrich themselves through unfair processes that hold down wages and raise prices".

The scientific goal of the workshop is to investigate the sense, in which the satisfaction of other-regarding preferences in the sense of strong reciprocity qualifies as welfare. Questions to be discussed include (but are not limited to) the following:

(1)   Is there a standard welfare economic interpretation of the satisfaction of preferences in the shape of strong reciprocity? If yes, it needs to be demonstrated that these preferences satisfy the weak congruence axiom, and that the central concepts and tools of standard welfare economics are defined for these preferences.

(2)   Can the satisfaction of preferences in the shape of strong reciprocity be interpreted in terms of the generalized framework that Bernheim & Rangel (2009) develop for behavioral economics? If yes, it needs to be demonstrated that these preferences are at least acyclic, and that the central concepts and tools of the generalized framework (generalized individual and Pareto optima, generalized compensating variation) are defined for these preferences.

(3)   Fehr & Schmidt (1999, p. 841) prove that one of the Nash equilibria (that of full cooperation) can be sustained as an equilibrium outcome if there is a group of $n \geq 1$ "conditionally cooperative enforcers". Hargreaves Heap and Ismail (forthcoming) demonstrate that any Nash equilibrium is a Pareto optimum in non-cooperative games under the "no-harm principle", which requires that the payoff of a player be not diminished by the actions of another. Can we specify a similar condition or principle that turns any Nash equilibrium into a Pareto optimum when players have preferences in the shape of strong reciprocity?

(4)   Bowles & Gintis (2011, chap. 10) suggest that individuals exhibiting strong reciprocity have "internalized" a norm or deontological constraint. Could this mean that individuals exhibiting strong reciprocity have "higher-order" (or moral) preferences, that the satisfaction of these preferences amounts to a kind of higher-order welfare, and that this higher-order welfare differs from "lower-order" welfare in that only the latter can be analyzed in terms of standard welfare economics or its behavioral generalization?

(5)   How is public policy to be conceived if individuals exhibiting strong reciprocity can be said to fare well (in any sense)? Is there a role for paternalistic policies that punish unkind behavior and promote kind behavior? Or will reciprocity work most effectively if public policy is decentralized, as Sugden (2018, chaps. 3, 7, 8) and Oliver (2019, chaps. 7-9) seem to suggest?

References

Bernheim, D. & Rangel, A. (2009). "Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics." *The Quarterly Journal of Economics* 124(1): 51-104.

Bowles, S. & Gintis, H. (2011). *A cooperative species: human reciprocity and its evolution*. Princeton: PUP.

Case, A. & Deaton, A. (2020). *Deaths of Despair and the Future of Capitalism*. Princeton: PUP.

Falk, A. & Fischbacher, U. (2005). "Modeling strong reciprocity." In H. Gintis, R. Boyd, S. Bowles and E. Fehr (eds.), *Moral sentiments and material interests*. Cambridge, MA: MIT Press, pp. 193-214.

Hargreaves Heap, S. P. & Ismail, M. (forthcoming). "No-harm principle, rationality, and Pareto optimality in games."

Oliver, A. (2019). *Reciprocity and the Art of Behavioral Public Policy*. Cambridge, MA: CUP.

Sugden, R. (2018). *The Community of Advantage. A Behavioral Economist's Defense of the Market*. Oxford: OUP.