

SUCKERS, PUNTERS, PATHBREAKERS:  
WHEN *HOMO OECONOMICUS* IS SELFLESSLY  
SELFISH

*Anthony de Jasay*

Rational choice presupposes that people do what they like better than any available alternative. If, however, we mistrust what they declare to like or what psychology is supposed to tell us about it (a pardonable enough mistrust), we can only infer what they like from observing what they do. We must be content with revealed preference. The theory of choice is locked into the tautology of “they do what they like because they like what they do,” and requiring their preferences to be orderly and consistent is of little practical help. In its elegance, modern choice theory, as represented in neoclassical economics, is too smooth and slippery to be very useful.

The resulting frustration seems to me to have two consequences. One is a more or less unconscious backsliding into old-fashioned utility theory. We know more than revealed preference tells us; we know what people like, therefore we can predict their choices (more or less) before knowing what they chose. They like “utility,” the motive for choice. More formally, the things a person likes are arguments in his “utility function” that he seeks to maximize if he is rational. Further tempting detours on this road may lead to suppositions about a stable relation between “utility” and income (the “diminishing marginal utility of money”) and about the addition and subtraction of different people’s “utilities,” both suppositions permitting

---

*Cato Journal*, Vol. 31, No. 2 (Spring/Summer 2011). Copyright © Cato Institute.  
All rights reserved.

Anthony de Jasay is an independent scholar and author living in France.

irresistibly attractive conclusions about “maximizing aggregate social utility” and others of the same family.

The second consequence of the apparently barren elegance of modern choice theory is a repudiation of the backsliding involved in the first. The somewhat outdated utility theory of the latter points almost (though not altogether) inevitably to a *homo oeconomicus* who is opportunistic, self-interested, selfish. The lowbrow criticism of this image is that “neoliberal man is only interested in money,” while the highbrow one constructs ingenious laboratory experiments to confirm the evidence of everyday life—namely, that he often behaves as if he were not very interested in money. He will at certain junctures appear actually to be sacrificing his self-interest either for no discernible reason or in favor of some ideal, such as some notion of fairness. Of course, revealed preference remains unbeaten by all this. Man does what he does because that is what he likes better than doing anything else at that juncture. Opting for what he wishes is opportunistic and selfish, though it may be that he wishes the fulfillment of the wish of others and, acting accordingly, could only be described as selfless. If, as logic leads us to recognize, all choice is selfish, his conduct should perhaps be classified as “selflessly selfish.” This term is droll and impressionistic, but it does help to dispel the crude notion that rational choice means something like “maximizing money income.”

However, lest matters should start to look too easy, we must notice some very important junctures where behavior seems to be motivated by selfless selfishness, but where this refinement is in fact unnecessary, for typical choices can be shown to be rational in a simpler sense. Though there may well be others, the most significant of these choices are made by the sucker who contributes to a public good, the punter who makes the risky first move in an equilibrium selection for a “game,” and the pathbreaker who pioneers a new way rather than leave it to others to do it.

## The Sucker for Public Goods

Conventional wisdom, codified in Mancur Olson’s *The Logic of Collective Action*, has it that, as a general rule, rational agents do not voluntarily contribute to the cost of public goods (Olson 1965). In a large number case, each person’s contribution to the cost of the public good would be relatively small. Since the benefits of the pure

public good flow to everyone, and the cost of exclusion is prohibitive, each person would try to free ride—getting the benefits without contributing to the costs. If everyone acted in this fashion, there would be no voluntary provision of the public good, even though production would yield net social benefits.

The character of *homo oeconomicus* should forbid him to play the sucker and contribute voluntarily. If suckerdom is to be explained, it must be in behavioristic terms: The contribution was made out of solidarity with one's community, by a wish to look honest and not be despised as a free rider, or by decency proper. The sucker chooses the decent thing because he likes it better than available alternatives. However, a fairly plausible argument shows *homo oeconomicus* willingly acting the sucker under far from extravagant assumptions and without his having any care for solidarity, decency, or the semblance.

Let there be a small riverbank town that has had flood damage and seeks to protect itself from future floods. Consultants present the residents of the town with a probability distribution of floods of various severity over the foreseeable future, the corresponding damage, and the cost of the size of dam needed to protect against a flood of a particular severity. The consultants also calculate the ideal size of dam—namely, one whose marginal cost is just equal to the probability-weighted marginal damage that would be caused by the particular size of flood that the town would suffer if it chose to build a smaller and cheaper dam. Whoever is entitled to interpret the town's wishes decodes that an attempt should be made to build this ideal dam, which can be expected to yield a total benefit (avoided damage, virtual benefit) in excess of its total cost. It is this dam that can be expected to maximize its benefit as a public good.

There being no social contract obliging the townspeople to pay the taxes that a collective choice mechanism (e.g., a voting majority) imposes, the mayor calls for volunteers to pledge a contribution toward the dam's cost. If the sum is undersubscribed, the pledges are cancelled as if nothing had happened; if it is oversubscribed, the pledges are reduced pro rata. How the subscription is going is kept secret until it closes.

What does a rational individual do? If he does not subscribe but enough others do, the dam gets built and he, a free rider, benefits from its protection without bearing any of its cost. Conventional wisdom has it that free riding dominates suckerdom. On a closer look

and within the assumptions made in this section, however, this is not the case—in effect, there is no dominant strategy.

Ignoring what his fellow townspeople will choose to do (though having fragmentary bits of information about their dispositions), each individual must act as if he faced a probability distribution of the decisions of the others. The distribution ranges from one extreme where no one subscribes to the other extreme where all subscribe. Somewhere in between there is a probability that just enough others subscribe to make his own eventual contribution decisive for the success or failure of the attempt to find voluntary funding for the dam. If the proportion of subscribers is between zero and the decisive one, our rational individual would expect to do better to subscribe. Not subscribing and hoping that the dam will get built anyway would be to gamble against the odds. Subscribing, on the other hand, commits him to nothing in case the subscriptions of the others are insufficient, commits him to subscribe if his subscription is the decisive one, and commits him to a reduced subscription if more than the decisive proportion of the others subscribe. Only in the latter eventuality would he expect to be better off by taking a gamble on free riding, with the dam getting built without any contribution on his part.

Thus, while there is no dominant strategy and the thesis that free riding is dominant in collective action proves to be invalid under assumptions that are less than extravagant, presumptions can be formed about contingent strategies likely to be adopted by rational persons. Such a person will be the more likely to volunteer to be the sucker the lesser is the likely proportion of others doing the same. The converse is true for the likelihood of the rational person opting for the free-rider role.

This conclusion undermines to a significant extent the general belief that the imperative need of a society for public goods justifies submission to the coercive authority of the state, for only by coerced taxation can public goods be provided. The latter belief can be upheld only by conceiving of public goods as wholly or almost wholly divisible, so that each marginal contribution to the cost of such a good increases the benefit it provides to the public by a marginal amount, hence by near-nothing to any single member of the public, such as the contributor. If so, he would contribute only under coercion or some form of altruism or solidarity with others. However, the very concept of a public good—namely, that access to it is neither excluded nor rationed, and that every member of the public

consumes it freely at his discretion without depriving any other member of the public from doing the same—entails that the public good is an indivisible whole “tailored” to its purpose and the size of the public for whose unrestricted consumption it is destined. It is more like a dam than a school lunch.

In the language of neoclassical (and also of Austrian) economics, indivisibility dethrones the marginal and enthrones the decisive contribution. Marginal productivity is either zero or equal to total productivity. When total contributions are just sufficient to produce the public good, each contribution is decisive, for its withdrawal would entail failure of publicness of the good. Hence, the marginal product of each contribution is equal to the total product, the indivisible public good. This manner of putting the matter is to squeeze its logic to the point of abusing it. However, it is a useful abuse if it illuminates, albeit from an unusual angle, the strong force that makes people in important contexts act selflessly selfish.

### The Punter in Equilibrium Selection

In noncooperative games with multiple equilibria, it cannot be predicted which of the potential ones will turn out to be the solution of the game. One equilibrium being Pareto-optimal, or at least greatly superior to another, is no reason for expecting it to be selected. The actual choice of one equilibrium can, of course, always be explained by imputing to the players behavioral motives that would appear consistent with, and adequate to provoke, the actual choice. More ambitiously, an *ex post* explanation in the rational-choice rather than the behavioral mode can also be constructed by imputing to each player appropriate conjectures about the other player. Admittedly, they cannot generate valid predictions. What they can do, though, is to improve our understanding of the problematic nature of equilibrium selection. They may, in particular, help to identify the circumstances under which the selection of mutually more advantageous, Pareto-superior solutions becomes more rather than less likely. Selection of a particular equilibrium out of the potentially available ones may occur *ex nihilo* or by way of changing over from another, pre-existing equilibrium.

In the latter case, the player who seeks to initiate the change must take a gamble whose odds are hidden in the recesses of the other player’s mind. The first player, by abandoning the existing

equilibrium, loses its protection. He suffers losses, and the other player reaps gains, that continue until either the second player follows the lead of the first and they create a new equilibrium, or the first player gives up his initiative and reverts to the protection of the old equilibrium. Prima facie, he is a hero, like the soldier who volunteers to be the first over the parapet and advances without assurance of being backed up. He does it because, as we may say analytically, he would rather do it than not do it, or he selfishly pleases himself; however, as his course of action is also an attempt to serve his side's interest it may claim to be selflessly selfish. It can be argued, though, and will be argued below, that sheltering behind the parapet of the old equilibrium is not a dominant strategy; that being a punter in the equilibrium selection attempt is perfectly consistent with being a *homo oeconomicus*; and that selflessness may (but by no means need) mean self-sacrifice.

The odds the punter accepts when betting on his success to initiate a new, improved equilibrium are, of course, subjective matters of his own conjectures about how the other player plans to respond to any move on his part, and that plan, in turn, depends on what the other player conjectures the first player's plan to be. Common knowledge cannot be assumed. Instead, the players are in a situation of the reflecting mirrors type: "I think that he thinks that I think, etc." Such infinite regress will naturally be stopped quite short if it is to serve any practical purpose and avoid cumulative error. In any case, nothing prejudices its result to be typically dissuasive, and nothing prevents the ideal punter who neither loves nor fears risk, from judging each such situation on what he takes it to be its merits and finding the odds that it seems to offer perfectly acceptable in some cases, though probably not in all.

Let First Mover be a maker of brooches and Second Mover a maker of necklaces. Every day each steals a bauble from the other. First Mover would rather sell his brooches to his legitimate customers than have stolen necklaces he can only sell to a fence, and Second Mover would likewise prefer to have necklaces to sell than stolen brooches. However, they stick to their daily routine of mutual theft. On a Sunday, First Mover refrains from stealing Second Mover's necklace. The latter, however, continues the daily routine and steals the former's brooch. In this disequilibrium, a loss for one and a gain for the other are created. The following Sunday, the same moves are repeated and produce the same result. For Second

Mover, the best outcome would be the indefinite iteration of disequilibrium every Sunday; next best would be that he, too, refrains from stealing on Sundays; the worst that First Mover gives up the expensive attempt to lure him into a mutual truce on Sundays, reverts to seven-day stealing and the old, Pareto-inferior equilibrium is restored. Second Mover must decide how far to push his luck, since persistent seeking of his seemingly best alternative must at some point lose him not only it, but also the second-best alternative of the new equilibrium with the Sunday truce. The First Mover must decide on how many Sundays he will unilaterally refrain from stealing and wait to see whether Second Mover will reciprocate. If neither player miscalculates the odds that attach to alternative moves between rational players in these circumstances, a new, improved equilibrium will have been successfully selected. Further improvements to a full weekend no-theft truce and eventually even to a Pareto-optimal seven-days-a-week respect for property might then become progressively less difficult to attain.

### The Pathbreaker to Fairness

When a distribution of good or bad things among a designated set of individuals is fair has never been defined with even tolerable clarity. The best practice in the matter is probably the somewhat cavalier one of saying that fair is what most ordinary people in ordinary speech call fair—though such a solution is both erratic and does too much honor to ordinary speech. It is indeed remarkable how the almost total absence of fairness criteria fails to trouble those who employ the word so frequently with such confidence. The confidence may be justified in some very limited contexts. One of these arises in two- or more-person interactions that provoke approval or condemnation by applying to them such yardsticks as decency and not taking advantage. The present section treats fairness in this narrow but not insignificant sense.

Justice is rooted in rules, ownership, and reciprocal agreement. In its pure form, its rules are spontaneous conventions that are all voluntarily adopted rather than agreed by some and imposed on others by virtue of some rule-making rule. In contrast to justice, fairness is rooted in ethical intuitions that need not be unanimous. Some may profess the intuitions that prevail in average opinion without being guided by them in their actions. It is widely held that

fair dealing and the maximization of personal interests in the narrow *homo oeconomicus* mode tend to be in conflict, and that there is no built-in machinery for the enforcement of fair conduct in the same way as the machinery that enforces the rules of justice by the self-defense built into behavioral conventions that punish deviation.

Consider now the hoary parable of slicing the cake fairly. The cake has been baked in heaven and is meant to be divided among a defined number of related people, none of whom has any prior claim to it. The person wielding the knife must have the decency not to favor anyone with a bigger slice than the others, and must not take advantage of his having the knife. This platitudinous form of a basic fairness problem is transformed into one of acute interests in the Ultimatum Game, invented by Werner Gueth and tested by him and others in many cunningly devised experiments. In this game, the Proposer offers to divide an unowned cake fallen from heaven between himself and the Respondent in two slices. If the Respondent accepts the share offered to her, they both get their agreed share, while if she refuses, neither gets anything and the cake is snatched away from them. Ostensibly, the Respondent is better off if she accepts any size of slice down to a paper-thin one than if she refuses it and gets nothing.

If the Proposer's offer is indecently low, the Respondent may indulge herself by punishing him and refuse his offer. It costs her but little to do so. The Proposer can reduce the risk of such punishment by pitching his offer higher and making refusal of it more costly to Respondent. The higher the offer, the safer it is for the Proposer because the more expensive it is for Respondent to refuse it. Except if the Proposer plays "maximin" and would rather take a paper-thin, minimal slice than run the least risk of getting nothing, he will at some point find that any further reduction of the risk of getting nothing is not worth any further reduction of his share of the cake. The Proposer can only guess where this point might lie and in experimental play is content to go to what Thomas Schelling (1960: 57) called the "focal point." He offers to share half-and-half, and the Respondent seems content with this. Fairness triumphs.

How this fairly uniform and stable-looking result is achieved is a matter of conjecture. The least interesting of these is to suppose that Proposers are fair-minded, like to play fair, and will play fair even at the cost of getting no more than the "fair" share of the cake. This, of course, is a tautology rather than a real conjecture.

In a more interesting one, fairness would figure as an instrument helping to maximize the probability-weighted slice of cake the Proposer can expect to earn by issuing his ultimatum.

The real question then arises: Why does the Proposer expect the Respondent to kick over the tray and send the whole cake to hell if she thinks the Proposer is taking advantage of her? The tautological answer is that unfairness to herself and others makes her indignant and wishful to punish it, therefore she will do what she would rather do and punish it albeit at a cost. The non-tautological conjecture is that punishing unfairness is instrumental to maximizing the share of the cake she expects to get. It does so by teaching the Proposer that the greedier his ultimatum, the greater the probability of its being rejected and earning a zero payoff.

Admittedly, for the Respondent's tacit threat to be wholly credible and not a bluff, the game cannot be a pure one-off neither played again with the same Proposer, nor played by other Proposers with her or with other Respondents linked to one another by shared information. However it must always be the case to some extent in real-life versions of Ultimatum Games. In feudal times, some "unequal bargains" driven by some lord provoked the exodus of serfs (at some immediate cost to themselves) and must have provided some inducement to other lords to treat their serfs with some degree of humanity. One might impute to the same kind of influence the fact that wages in industrial and mining locations far removed from competitive labor markets were not simply driven down to near-subsistence level.

The long and short of the argument of this section is that in order to explain the prevalence of "fair rather than exploitative" solutions in situations where the dice seem loaded in favor of one of the players, it is not necessary to have recourse to an assumption of love of fairness. This is not to claim that such love may not exist, but simply that we might expect some prominent distributive interactions to have the same result when no such love exists as when it does.

In an imaginary all-encompassing Ultimatum Game, the Respondent who is the very first of all Respondents to refuse an offer that by rights she ought to accept because it is better than nothing, is making a sacrifice in much the same way as all pathbreakers do who make the way easier for all who come after them. At least metaphorically, there has to be a first, and the pathbreaker who assumes the role would presumably prefer to be second if she could count on

someone else to go first and break a path. Yet, if it were always preferable to wait for someone else to go first, nobody would go first, nobody could go second, and no path would be made to fairness. The sacrifice of teaching a lesson or breaking a path, like the acceptance of the role of sucker in public goods provision and of first mover in equilibrium selection, is a wager whose odds may not attract everybody but should at any rate attract the number needed to produce gains for themselves and everybody else.

## Conclusion

If all we can say about people's choices is that they do what they would rather do than anything else available, it is a pointless truism to call all choices selfish. The truism obviously continues to hold when people make manifest sacrifices that benefit other people. They would do these benign things rather than anything else. Perhaps we should say, a little wryly, that they are being selflessly selfish. The behavioral approach to the moral sciences finds it sufficient to attribute this to people liking to advance the interests of other people, wishing to conform to recognized behavioral norms, or seeking approval by others. Pursuing diverse notions of fairness also figures as an end explaining behavior. It is reasonable enough to concede that all these putative ends do influence human conduct, though it is disconcerting to learn that people seek approval because they like to be approved of, or that a sense of solidarity induces them to support a group objective they personally do not much care for. The trouble with such a list of disparate final ends is that, being incommensurate, their combination cannot be maximized. Any combination that is chosen is the best; it cannot be improved or criticized.

This article has taken a close look at three well-known interactions that between them cover an important part of a society's functions: the provision of public goods, the selection and eventually the improvement of coordination equilibria, and the pathbreaking action by which its potential victim discourages the "unfair" division of the proverbial cake. In each, the probability of expected actions by others plays a key part. Such interaction postulates some apparent sacrifice of a player's *prima facie* interest or dominant strategy; the player appears to be selfless for the purposes of the interaction. It turns out, however, that the selflessness, though the sacrifice it requires is real enough, does not serve the good of others as a final end. The good of

others emerges as a by-product, a sort of positive externality. The final end is the good of the player himself. Moreover, his good (“welfare”) can at least conceptually be maximized in the neoclassical tradition where *homo oeconomicus* can use what A. C. Pigou (1912: 3) called “the measuring rod of money” to add up his goods. Let at least that part of purposive human action be illuminated by the criteria of rationality. The rest, important as it may be, will have to be left, unfalsifiably, to look after itself.

## References

- Olson, M. (1965) *The Logic of Collective Action*. Cambridge, Mass.: Harvard University Press.
- Pigou, A. C. (1912) *Wealth and Welfare*. London: Macmillan.
- Shelling, T. (1960) *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

