Teaching Social Preferences

Simon D. Halliday[1,*]

[a]*Smith College, Department of Economics, Wright Hall 234, Northapmton, MA, 01063*

Abstract

Over the past several decades, evidence from behavioral and experimental economics has shown that people have preferences that are not based entirely on self-interest, that is, that "economic man" (*Homo economicus*) does not represent the preferences of all people in all states. On the contrary, and in a variety of economically relevant contexts, people demonstrate preferences that are altruistic, inequality- or difference-averse, or reciprocal (so-called other-regarding or social preferences). Despite these advances in evidence and theory, few examples exist for teaching these preferences in undergraduate microeconomics courses. In this paper I address this gap in the literature by proposing four models in which such preferences can be taught in an undergraduate microeconomics course that use graphical analysis and algebra to teach constrained optimization, externalities, and exchange. I provide examples in which the results of the dictator game can be taught with the tools of indifference curves between goods with a budget constraint (and show how changing the "price" of altruism demonstrates a "law of demand" for altruistic behavior), a model of constrained optimization with a good (wealth) and a bad (inequality) where inequality-averse or self-interested preferences will result in different policy conclusions; a model of exchange in an Edgeworth box in which, instead of self-interested preferences, consumers are altruistic yet conflict can still exist among Pareto-efficient outcomes along the contract curve; and, lastly, a model of common pool resource extraction with negative externalities (based on a variant of the Cournot model) in which social preferences can result in Pareto-improvements over the Pareto-inefficient Cournot-Nash equilibrium. I provide suggestions as to how the models can be incorporated into existing courses as supplements to or variants of existing course content.

JEL Classification: A20; C7; D9

Keywords: social preferences; other-regarding preferences; economic experiments; intermediate microeconomics.

*Corresponding Author

Email address: shalliday@smith.edu (Simon D. Halliday)

Introduction

Humans and self-interested economic man respond differently to basic economic choices. Do people contribute to public goods and are they willing to punish those who infringe norms? Does a citizen choose to give to charity? Does a trader make a fair or an unequal offer in a bargaining situation? Does an employer respond to a worker with trust and provide them opportunities for voice and autonomy or does the employer cut their wages and reduce their work hours to maximize profits?

Not only has research shown that people behave differently to how standard economic theory would predict they might, but students also know this and have been campaigning internationally for their curricula to reflect these advances in economic research [citations]. Furthermore, recent work on the economics undergraduate curriculum shows that there are topics at the forefront of economics research that can be – and have begun to be – taught at both the introductory level and intermediate levels (Bowles and Carlin (2020), Bowles et al. (2019)).

I shall use the terms social preferences or other-regarding preferences to capture the sets of preferences in which altruism, difference-aversion, reciprocity, and so on, play a role in an economic agent's decision-making. Social preferences are salient in a variey of economically relevant situations. People give to charity on a regular basis, with charitable giving being approximately 3% of US GDP (citation). Many people involved in bargaining situations choose to make fair offers, even in take-it-or-leave-it contexts, rather than not trade at all. Employers choose to take temporary hits to profits and do not cut wages or employment as much as theory would predict worried about costs to morale and about costs to workers' families (Bewley (2009)). Workers respond to wage increase more than proportionately, contrary to the elasticities predicted by labor supply curves, suggesting that reciprocity and normative motivations play a role in their choices (Fehr and Goette (2007), Fehr, Goette, and Zehnder (2009)).

I provide an introduction to the ways in which other-regarding preferences and social preferences more broadly play a role in microecomic theory as supported be evidence from economic experiments in the lab and the field. I argue that these preferences can be modeled using standard tools in economics courses: consumers in a constrained optimization framework using indifference curves and a budget constraint for altruism and inequality aversion, agents with best-response functions to isolate a Nash equilibrium in a strategic interaction, and traders engaging in exchange in an Edgeworth box. Social preferences can be incorporated into each model in tractable ways and can be understood mathematically and graphically.

Yet core undergraduate economics courses often teach constrained optimization, trade with Edgeworth boxes, firm cost minimization, labor markets, and other economic models as if humans play no role and as if self-interested economic man rules all decisions. Such thinking has not been true in the leading journals for some time, so why does it remain true at the undergraduate economics level?[1] For many,

[1]See Bowles and Carlin (2020) for a discussion of the introductory economics class and Bowles

these topics are taught as if they are luxuries: add-ons if the instructor has the time. Even if they are covered in a textbook (for example, in a supplement on behavioral economics), the topics often do not use the tools that students are accustomed to, or the coverage is a minor morsel telling the student: self-interest is the norm, do not worry about minor "anomalies."

In this paper I confront a gap in the economics education literature and in the textbook market. Many instructors fail to address what have become core concerns at the forefront of economic theory and empirics: the role played by social preferences in the decision-making of economic agents. To that end, I argue, first, that social preferences and understanding the empirical support for other-regarding preferences are well grounded in the cutting edge of economics research, but not in many courses except as an exception to the standard of self-interested behavior. This contradicts many experimental results in which the purely self-interested choice is not the most common behavior. Second, the empirical results can be taught to undergraduates using standard tools that are taught in undergraduate introductory microeconomics, intermediate microeconomics, or public economics courses. Third, in addressing this gap in the literature the instructor will also satisfy growing concerns by students with the clash between their subjective experience (which often reflects their own social preferences) with the content of their economics courses, which often narrowly focuses on material self-interest.

People regularly behave in ways that contradict pure material self-interest, yet our models of rational actors acting within the framework of constrained optimization can still explain their behavior.[2] What this requires of us, then, is to consider how to implement changes to the undergraduate curriculum to reflect the advances in economic research. Among these changes, I argue that applications employing the tools student already learn provide an opportunity to engage students using alternative interpretations of models with which they are familiar or applying tools with which they are familiar to new contexts. This takes advantage of the knowledge they have lready accrued while asking them to transfer that knowledge to new, but related, domains.

I progress as follows. I briefly cover the evidence and theory relating to social preferences and provide a discussion of the contexts in which the models have been taught. I proceed to discuss how the models can be incorporated in reasonable ways into four topics of undergraduate economics courses using, for the most part, tools that students will be accustomed to (indifference curves and constraints) and in one case arguing for the use of best-response curves that students would see in courses that teach Cournot Duopoly. I demonstrate the basics of each example, but also supplement the content with an online appendix and suggestions

---

et al. (2019) for an explanation of the economics research corpus and its relationship to the intermediate microeconomics curriculum.

[2]Recall that rationality merely requirest that an agent have preferences that are complete and transitive. Rationality implies consistent pursuit of ends, not self-interest.

for further reading.[3] I conclude with some recommendations for changes to the undergraduate curriculum that reflect ideas at the forefront of research.

Theoretical Background and Context

For much of the twentieth century in economics, the breadth of human motivations and preferences were ignored in favor of a model of preferences based on material self-interest, with scant attention paid to altruism, social norms, or other-regarding behavior more broadly. The centrality of self-interest was grounded in the marginal revolution in economics, going back to Marshall, Edgeworth, Menger, and others. For example, Edgeworth said, "The first principle of economics is that every agent is actuated only by self-interest" (Edgeworth (1881)). This tendency in economics research and thinking was reflected in undergraduate curricula, with altruism, reciprocity, and trust left to the other behavioral social sciences to explain. This abandonment of diverse other-regarding preferences, or what I shall call and broadly define as social preferences, was not consistent with the roots of economics and political eocnomy.

Adam Smith recognized the diversity of human motivations in *The Theory of Moral Sentiments* arguing for understanding fellow-feeling and sympathy in human relations and motivations (Smith (1982)). Researchers have thus called Smith, somewhat tongue-in-cheek, the first "Behavioral Economist" (Ashraf, Camerer, and Loewenstein (2005), see also Smith (1998) and Paganelli (2009)). Researchers in behavioral economics have buttressed Smith's arguments with evidence from a variety of decision-making and strategic experiments, such as the dictator game, ultimatum game, trust game, public goods game (or voluntary contribution game), among others.[4] In these contexts, though entirely self-interested behavior is common, in most cases the majority of people behave in ways that are not entirely self-interested and that reflect some measure of social preferences: concerns for reciprocity, conditional cooperation, inequality aversion, and intentions-based preferences. Furthermore, these have real-world correlates found in field experiments and lab-in-the field experiments paired with monitoring of field behavior (Fehr and Leibbrandt (2011)).

Since Kahneman, Knetsch, and Thaler (1986), dictator games have been used as an experiment to understand fairness in individual and corporate environments. A dictator is endowed with a monetary amount and asked how much they would choose to transfer to an anonymous partner. A recent meta-study shows that dictator-game giving is prevalent internationally with a comprehensive sample and show that, on average, dictators share 28.5 percent of their endowment (Engel 2011). In a small minority of samples, the dictator gives zero on averagem, which

[3] The content will be hosted at simondhalliday.com/teaching_social_preferences, though the page is currently not populated.

[4] For the sake of concision, I shall rely on the reader to check the references for the structure of each experiment, though I shall explain the dictator game in more detail.

tends to be more common when the dictator has the option to take, rather than just give, to their partner (List (2007),Bardsley (2008),Halliday (2011),Nikiforakis and Mitchell (2014)).

Though metastudies have not been done for all types of experiments, the evidence internationally shows that in anonymous interactions people are willing to give to others altruistically (dictator games), cooperate with others (public goods games, coordination games, common pool resource games), trust others with investments (trust games), reciprocate trust with remitted income (trust games), punish others who infringe perceived (injunctive) norms (public goods game with punishment; moonlighting game; third-party punishment game), among various others.

To explain the patterns in this behavior, a variety of theories have sprouted in the literature to identify what kinds of utility functions would be consistent with such behavior. Levine (1998) models concerns for altruism and spitefulness. Fehr and Schmidt (1999) concerned outcomes-based inequality aversion, where people dislike outcomes in which they are relatively more advantaged than others (advantageous inequality) and dislike outcomes in which they are worse off relative to others (disadvantageous inequality). Falk and Fischbacher (2006) examines more deeply concerns over reciprocity in a variety of experiments. More recently, a variety of models have focused on norms and norm enforcement (citation), differentiating intentions-based preferences from inequality-averse or differenece-averse outcomes-based preferences (see Falk, Fehr, and Fischbacher (2008), Cappelen et al. (2007),Cooper and Kagel (2016)). Moreover, a variety of work shows how these preferences extend into the workplace, with reciprocity and trust especially playing a role in principal-agent relationships (Falk and Kosfeld (2006),Ziegelmeyer, Schmelz, and Ploner (2012),Burdin, Halliday, and Landini (2018)). Other work has shown the role of such preferences in mechanism design and in understanding the crowding out of ethical behavior when policies presume that all citizens are self-interested (Bénabou and Tirole (2006),Bowles and Hwang (2008), Gneezy, Meier, and Rey-Biel (2011),Bowles and Polania-Reyes (2012))[5]

What, then, is the relevance of teaching these models at an undergraduate level? First, teaching economics in a particular way may attract or repel different students. In teaching these models, students have approached me to say that they had thought economics necessarily involved self-interest and that valuing self-interest was itself and end in economics. Seeing that economics could be used to understand preferences that were not self-interested meant that these students were interested in continuing with being an economics major. Second, those students who study these topics, realizing that the scope of economics is broader than they had originally anticipated, might themselves consider alternative careers to the careers to which many economics majors might normally go. For

[5]See Chetty (2015) and Bhargava and Loewenstein (2015) for the influence of behavioral economics broadly – beyond social preferences – on public policy.

example, instead of investment banking, I have had students who have said they would now consider government employment or working with an NGO. This is not to detract from the value of investment banking in value creation in a capitalist economy, but more to say that students see the application of economic ideas in the core curriculum to a wider array of economic problems. Third, there is a marketing or public relations aspect to teaching these topics in core courses.[6] When I can say to fellow faculty that I am teaching both self-interest and altruism in my classes, my fellow faculty members begin to see economics in another light. Students too enjoy surprising faculty in other disciplines with evidence and models that they learn here. For example, a student in my class corrected an anthropology professor's conceptualization of what economics was "about" when explaining models we had used in class and their varied applications to competition among firms and to models of extraction in common pool resources, or to choices about study time and leisure to choices about allocations between oneself and a charity.

Contexts for Teaching the Models

The author has coordinated with a variety of instructors to teach these models in undergraduate intermediate microeconomics and related courses. The class sizes have ranged from 30 to 300 in institutions such as liberal arts colleges and public universities. The materials have been taught in the US, UK, India, Ireland, Germany, and South Africa, among others. In the majority of the cases where the content has been taught, they are used in courses where calculus has been presumed as a pre-requisite for taking the course, though in certain circumstances where calculus is not a pre-requisite (mainly in the US) the models have been taught mainly graphically for the purpose of understanding the dynamics involved in the underlying models and to articulate the core economic ideas: rational preferences that are other-regarding, tradeoffs among alternatives, opportunity costs from constraints, the role of externalities, conflict of interest over outcomes among people with different preferences, and so on.

Dictator Game Behavior and Altruism

There are two methods one can use to teach dictator games. I teach the simpler method and outline a more advanced method that can also be used (depending on the structure of your class).[7]

Consider two agents, Alex and Baharak who are playing a Dictator Game where Alex is the proposer and Baharak is the receiver. As the Receiver Baharak has no

[6] See, for example, Bauman (2011) and Stafford (2013) on public perceptions of economics making students selfish or conservative.

[7] A version of the dictator game model is also taught in *The Economy* (Bowles, Carlin, and Stevens (2017)), though the model in that context does not use calculus but simply uses the tangency conditions for an indifference curve to the budget constraint for constrained utility maximization.

actions, she simply receives what Alex gives her. Instead of the rather complicated theories offered in the previous section, we shall presume that two payoffs comprise Alex's utility function: his own payoff (π_A) and the payoff of Baharak (B), π_B. Alex has a utility function $u(\pi_A, \pi_B)$ where π_A and π_B indicate quantities of money in dollars for A and B respectively.

The two utilities enter Alex's value function in Cobb-Douglas fashion with a parameter λ giving the extent to which Alex is altruistic towards Baharak.

$$u_A(\pi_A, \pi_B) = (\pi_A)^{1-\lambda}(\pi_B)^{\lambda} \tag{1}$$

Where $0 \leq \lambda \leq 0.5$. When $\lambda = 0$, Alex is a *Homo economicus* and does not obtain any utility from making an allocation to Baharak. When $\lambda > 0$, Alex is altruistic and derives utility from transferring money to Baharak.

Alex's maximization problem is that he wants to maximize his utility given the allocations he can make between himself and Baharak. In a distribution choice like that offered by the dictator game, Alex faces a constraint, which is the amount of money with which he is endowed, y. Alex must choose a split of money (π_A, π_B) such that $\pi_A + \pi_B = y$, this is a constraint on Alex's choices similar to a budget constraint that students would be accustomed to seeing, that is $m = p_x x + p_y y$ for income, prices and quantities of each good. The equation for Alex's budget constraint (or the boundary of the feasible set of allocations between A and B) could therefore be given by a constraint like Equation 2:

$$\pi_A = y - \pi_B \tag{2}$$

Given this constraint and a utility function given by Equation 1, we can display a set of indifference curves in the (π_A, π_B) plane superimposed on the constraint given by Equation 2. In certain dictator games, however, the price of altruism is changed, such that a transfer to the other person (Baharak, B) results in either more or less money transferred. That is, $\pi_A = p * y - p \cdot \pi_B$. For example, if any amount given to the other person is tripled – in what is called a triple dictator game – then $p = 3$, so that maximum amount another person could obtain is $p \cdot y$, for example, \$60 if the endowment (y) is \$20 (Ashraf, Bohnet, and Piankov (2006)). Thus, when $\pi^A = 0$, $\pi^B = p \cdot y$. Otherwise, if $\pi_A > 0$, then $\pi_B = p \cdot y - p \cdot \pi_B$.

Alex's utility maximum would therefore be given by the point at which his highest indifference curve is tangent to his constraint (determined by the price and the amount of his endowment) and he would split the money according to a standard principle of constrained utility maximization, that is, where the marginal rate of substitution from his indifference curves equal his marginal rate of transformation from his constraint.

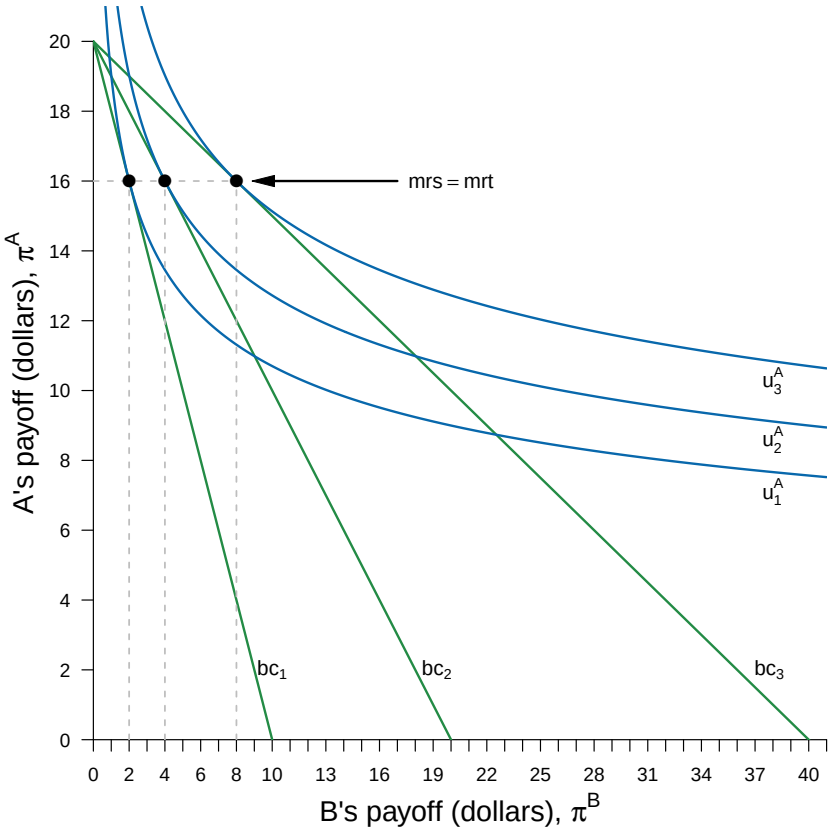

Figure 1: An altruistic person dividing an endowment

Figure 1 shows the case in which Alex has a high level of altruism (λ) such that $\lambda = 0.25$. As a consequence, Alex has convex indifference curves and his own payoffs and the payoffs to Baharak are goods to him (he has a positive marginal rate of substituion between his own payoff and Baharak's payoff). Alex's indifference curves become more horizontal as Alex becomes more self-interested (lower λ). At the extreme, if Alex has no altruism towards his counterpart, then $\lambda = 0$ and his indifference curves will be perfectly flat.

This framework for understanding altruism can also be used to demonstrate that altruistic preferences are also rational preferences (as in Andreoni and Miller (2002)). If an experimenter were to change the "price" of giving, for example by saying that for every \$1 dollar transferred the receiver gets \$0.50 or for every \$1 transferred, the receiver gets \$2 (both cases of which are shown with budget constraints bc_1 and bc_3) in Figure 1, then these can be used to derive a person's *demand curve for altruism*, which I depict in Figure 2. As is standard for someone

with rational preferences, the corresponding demand curve is downward-sloping: that is, the quantity demanded of altruism is greater the lower is the price, and the quantity demand is lesser the higher is the price.

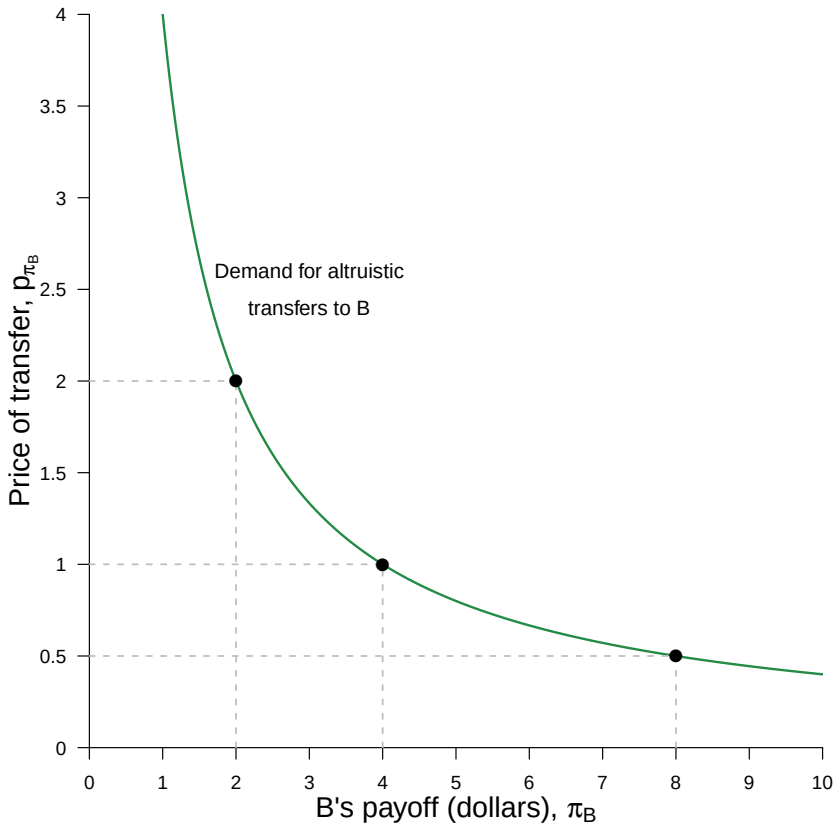

Figure 2: An altruistic person dividing an endowment

Difference aversion and constrained maximization

Inequality aversion or difference aversion involves people who dislike differences in outcomes between themselves and a counterpart (Fehr and Schmidt (1999),Fehr, Naef, and Schmidt (2006)). There are two tractable models that can be used to think about this idea for a model to teach undergraduate students and I show an application in which students use standard tools of indifference curves and a constraint to find a combination of average wealth and inequality that they would

judge to be acceptable.[8]

Consider two variables: inequality (Δ) and average wealth (ω) in society. There are two potential classes in society: Rich and Poor. Inequality (Δ) captures the difference between the wealth of the Rich and the wealth of the Poor. Figure 3 shows how average wealth first rises, reaches a maximum and then falls as inequality increases, as depicted by $g(\Delta)$. The shape of the curve reflects the following intuition: some inequality acts as an incentive to action and productivity; beyond a certain point, however, high inequality may result in lower average incomes overall (people may be disincentivized to work, for example, or there may be political instability as a consequence of the inequality, such as in pre-revolutionary France). The $g(\Delta)$ function is similar to a risk-return schedule: it provides the feasible combinations of average wealth and inequality. The negative of the slope of $g(\Delta)$ is the marginal rate of transformation of inequality into average wealth, or the opportunity cost of greater average wealth in terms of greater inequality. For low levels of inequality the opportunity cost of greater inequality is negative because greater inequality results in higher average wealth.

Because the classes are of equal size, average wealth is the mid-point between the wealth of the rich and the wealth of the poor. So the wealthy get average wealth plus one half of Δ and the poor get average wealth minus one half of Δ. Recall that the average wealth, ω, is determined by the level of inequality using the function $g(\Delta)$. Therefore:

$$\text{Wealth of The Poor} \qquad \omega^P = g(\Delta) - \frac{1}{2}\Delta \tag{3}$$

$$\text{Wealth of The Rich} \qquad \omega^R = g(\Delta) + \frac{1}{2}\Delta \tag{4}$$

[8] The functional form used by Fehr and Schmidt (1999) is not appropriate in an intermediate microeconomics course, but an alternative specification that I have found successful with advanced undergraduates involves the following utility function: $u_i(\pi_i, \pi_j) = \pi_i - \delta_i \max[\pi_j - \pi_i, 0] - \alpha_i \max[\pi_i - \pi_j, 0]$, where π_i and π_j are the payoffs to two players in, for example, an ultimatum game, and δ_i and α_i are positive constants reflecting person i's preferences for advantageous and disadvantageous inequality respectively. See also Bowles (2009).

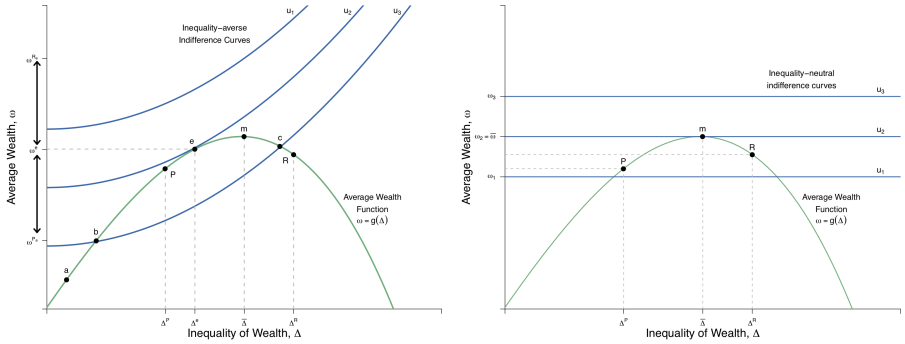

Figure 3: Inequality-Wealth Schedule and indifference curves

I focus on the use of a quadratic, quasi-linear utility function where wealth is a good and inequality is a bad, an example of which is shown in Equation 5 below:

$$u_A(\omega, \Delta) = \omega - \frac{\Delta^2}{2} \tag{5}$$

The indifference curves corresponding to this utility function are upward-sloping as a result of wealth being a good, but inequality (the difference between rich and poor) is a bad. A utility-maximizing policy-maker will choose the point where their marginal rate of substitution of inequality for wealth equals the marginal rate of transformation of inequality into average wealth, or where the tradeoffs they are willing to make between the outcomes are equal to the opportunity cost of inequality for wealth. As is standard, this occurs where the policy-maker's indifference curves are tangent to the constraint – in this case the inequality-wealth schedule.

An inequality-averse policy-maker would choose point **e** in the left-hand panel of Figure 3 with a level of inequality Δ^e with corresponding level of average wealth ω^e. On the contrary, an inequality-neutral policy maker would choose point **m**, maximizing average wealth $\bar{\omega}$ with a higher corresponding level of inequality Δ^m. These differences make clear to students a variety of issues: first, that people can differ in the preferences that they have over outcomes such as inequality and that these differences will manifest in contrasting indifference curves; and, second, given the same feasible set of outcomes, one policy-maker who is inequality-averse and another who is not will make differing policy choices that have material consequences for inequality and average wealth. These models help students to see the ways in which a standard economic model of constrained maximization can be applied to a policy-relevant topic with people who have different types of preferences as supported by data from economic experiments – some people are concerned with inequality and others are not and they will make different utility-maximizing choices as a result.

Coordination Failures and Social Preferences

A variety of problems in intermediate microeconomics can be explained using models in which players impose costs on each as a consequence of negative externalities (these externalities can be otherwise thought of as manifestations of contractual incompleteness). Consider a model of a common pool resource, a good that is rival and non-excludable. Each player wishes to extract resources from the common pool resource, such as fish in a lake, lumber from a forest, or grazing a herd on common land. I call the players Angela (A) and Bruno (B). Each player can control their own hours of work, h_A and h_B respectively.

Each player has a maximum amount of hourly productivity for resource extraction for their first hour of activity, $\bar{y}$ that is common across the players. Each player has decreasing marginal productivity and each imposes a negative external effect on the other, such that one player's hours of work reduces the productivity of their counterpart. Therefore, each player has the following production function:[9]

$$y_i(h_i, h_j) = (\bar{y} - \beta(h_i + h_j))h_i$$

For each player i against each other player j, β captures both the negative exterality and the decreasing productivity of a player's own hours of time spent in the extractive activity. Each player incorporates their production into their utility function, such that $u_i(h_i, h_j)$ is given by the following:

$$u_i(y_i, h_i) = y_i - \frac{1}{2}h_i^2$$

Therefore, a player has increasing disutility of work hours in extraction, but they benefit from output, y_i. They dynamics of this interaction can be captured with the indifference curves and best-response functions shown in Figure 4.

[9] In using the construction below, we follow a standard Cournot model with quadratic disutility in effort or hours worked rather than a constant returns to scale technology as would be more common in an example with firms. The results should be unsurprising as, from a firm's point of view, the problem of a Cournot market is that the are 'over-harvesting' customers when both firms would do better by restricting output and acting like a monopoly. The situation here is analogous for consumers, but without the welfare concerns of high prices for consumers in the Cournot market with a monopolist or cartel.

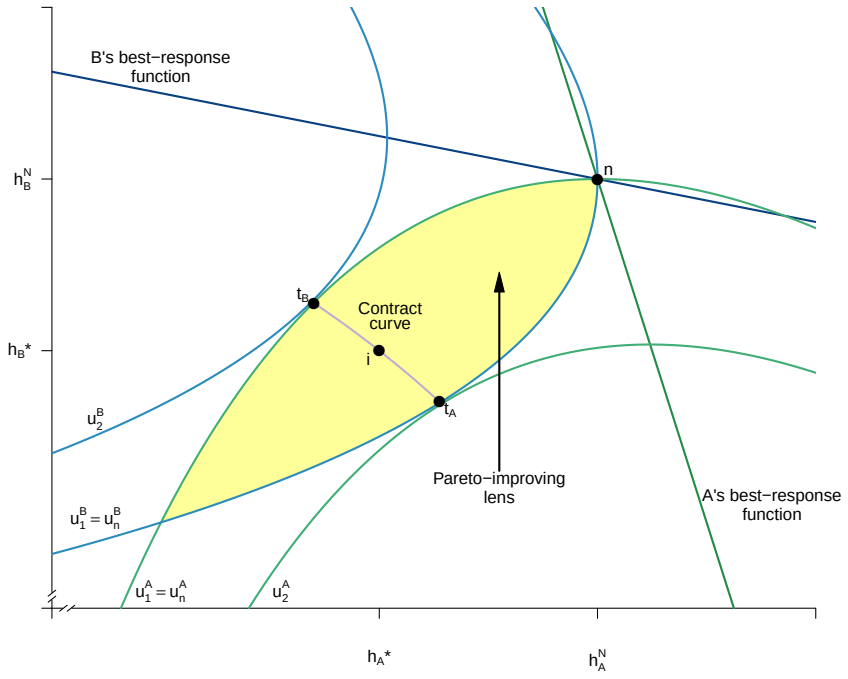

Figure 4: Common pool resource interaction

The demonstration of the best-response function and Nash equilibrium are provided in the appendix. The Nash equilibrium of the interaciton is given by a combination of hours of work (h_A, h_B), such that $h_A = \frac{\bar{y}}{1+3\beta} = h_B$. The Nash equilibrium is Pareto-inefficient and it would benefit both players if they could agree on a mutually beneficial decrease in their hours of work.

The Pareto-inefficiency can be clearly seen by the indifference curves being perpendicular to each other at their intersection; that is, the marginal rates of substitution are not equal, therefore there are unexploited gains from trade that can be obtained along the contract curve or Pareto-efficient curve which indicates the locus of all points where the players' marginal rates of substitution are equal. How then can players get to a Pareto-efficient point or at least a Pareto-superior point that is in the Pareto-improving lens?

Channeling the work of Elinor Ostrom and others, one means by which the players can achieve a Pareto-superior – if not Pareto-efficient – outocome is through interactions within the community that cultivate social preferences, fellow feeling, or other-regarding behavior (Ostrom, Walker, and Gardner (1992), Ostrom, Walker, and Gardner (1994), Ostrom (n.d.), Falk, Fehr, and Fischbacher (2002), Càrdenas and Ostrom (2004)). Social preferences operate in this regard by encouraing the agent (partially) to offset the cost that they inflict on others as a consequence of the externality. Consider a simplified idea of social preferences in which a player

has a coefficient of altruism λ towards another player in which $0 \leq \lambda \leq 1$. Each player therefore has a new utility function

$$u_i(h_i, h_j) = (\bar{y} - \beta(h_i + h_j))h_i - \frac{1}{2}h_i^2 + \lambda\left((\bar{y} - \beta(h_i + h_j))h_j - \frac{1}{2}h_j^2\right) \quad (6)$$

In Equation 6, the term multiplied by λ changes the best-response functions and the resulting Nash equilibrium for each player resulting in a reduction in work hours commensurate with a player's level of altruism. If player's are perfectly altruistic – that is they treat their counterpart as well as they themselves would like to be treated and $\lambda = 1$ – then they would choose allocation (h_A^*, h_B^*) at point **i** in Figure 4. When they are imperfectly altruistic – that is they treat their counterparts well, but not as well as they themselves would like to be treated and $0 < \lambda < 1$ – then they would choose a point in the Pareto-improving lens.[10] Furthermore, if agents are not purely altruistic, as is most often the case, the instructor can also discuss how social punishment can act like a per unit tax. Social punishment typically involves peers punishing each other or the group punishing an agent who breaks a social norm. In this case, altruism can be supplemented by punishmen to move outcomes closer to the contract curve or closer to point **i**.

Exchange and Social Preferences in Edgeworth Boxes

Many instructors teach exchange in intermediate microeconomics using Edgeworth boxes with self-interested Cobb-Douglas preferences. We can use a variation on the the earlier model with an agent, Alex (A), who has a value function the depends on the his personal utility and the utility of his counterpart in exchage, Babalwa (B). A and B each hold endowments of goods x and y respectively, where the total amount of x and y available for exchange in the Edgeworth box are given by $\bar{x} = x_A + x_B$ and $\bar{y} = y_A + y_B$.

Alex's value function is therefore given by the following Equation:

$$v_A(x_A, y_A, x_B, y_B) = (u_A)^{1-\lambda}(u_B)^{\lambda}$$

Where, once again, λ captures the extent of an agent's altruism towards their counterpart. Because u_A and u_B are both Cobb-Douglas in their arguments, and x_B and y_B can be re-written as $\bar{x} - x_A$ and $\bar{y} - y_A$ respectively, the value function simplifies to the following:

$$v_A(x_A, y_A) = \left(x_A^{\alpha} y_A^{1-\alpha}\right)^{1-\lambda}\left((\bar{x} - x_A)^{\beta}(\bar{y} - y_A)^{1-\beta}\right)^{\lambda}$$

[10] We do not make a claim here as to which specific point they would choose as the players would have a conflict of interest over points they would prefer, so prevailing institutions of exchange would likely determine which allocations would be more likely to occur, as Ostrom and others argue.

Where α and β are the coefficients of preference intensity in each agent's Cobb-Douglas utility function. The indifference curves of such a value function, however, are unconventially shaped in (x_A, y_A) space. In Figure 5 we present two cases with these indifference curves for $0 < \lambda \leq 0.5$.

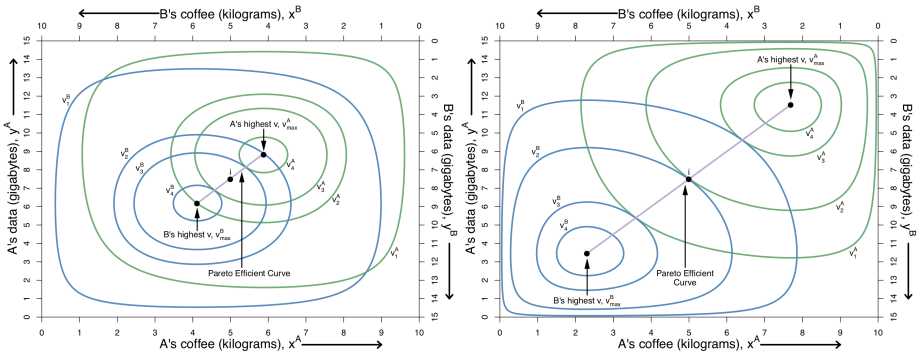

Figure 5: Edgeworth Boxes with altruistic preferences

In the first case (left-hand panel), the agent has a high level of altruism ($\lambda = 0.4$) whereas in the second case (right-hand panel), the agent has a high low of altruism ($\lambda = 0.2$). Because we constrain $\lambda \leq 0.5$, these cases therefore correspond to someone being approximately 40% and 80% altruistic (as a proportion of treating the other as well as the treat themselves, which would be the case for $\lambda = 0.5$ where the other would be as important as themselves).

Why are the indifference curves helpful for teaching social preferences in exchange? A student needs to master their understanding of marginal utilities and the marginal rate of substitution. A trader who is altruistic has a marginal rate of substitution (the negative of the slope of the indifference curve) that changes signs at different points along their indifference curves. Why is this true? Because x_B decreases as x_A increases or y_B decreases as y_A increases, along an indifference curve an altruistic agent is continuously balancing the marginal utilities of their own consumption of x and y with their counterpart's consumption of the those same goods. Holding constant the consumption of y, as A's consumption of x, x_A increases, this implies that the marginal utility to A of x_A is decreasing. At the same time, the marginal utility to A of B's consumption of x, x_B, is increasing. Depending on an agent's level of altruism (λ), to maintain their utility constant along an indifference curve, at some point A must increase B's consumption, x_B, rather than increase their own consumption x_A. At the margin, an increase in B's consumption of x with commensurate change in the marginal utlity to A of that consumption will more than offset the marginal utility to A of increase their own consumption of x. Consequently, at different points along the indifference curves the quality of the the commodities x_A and y_A change from being goods to bads. For example, at lower levels of consumption of both x and y, both are goods to A

for his own consumption. At higher levels of consumption of both goods, however, both are bads, it would improve A's utility if they consumed less of both goods and increased the consumption of their counterpart. We can identify this by seeing that the slope of the indifference curve changes and therefore the marginal rate of substition changes correspondingly.

Consequently, when x and y are constrained in their total amounts (as they are in an Edgeworth box), A will maximize his utility at a point internal to the Edgweworth box, such as $\bar{v}_A$. The location of $\bar{v}_A$ depends on A's level of altruism: the higher the degree of A's altruism, the close is $\bar{v}_A$ to the center of the Edgeworth box (proportional to the total endowments of x and y). Consequently, A's level of altruism also acts as a constraint on voluntary exchange. A would rather choose a point like $\bar{v}_A$ rather than an allocation where they have a higher post-exchange allocations, such as at point **j**. The result therefore shows how altruism results in each person wanting their counterpart to have more rather than less of the goods, as demonstrated by the contract curve (Pareto-efficient curve) contracting towards the mid-point of the Edgeworth box as the level of altruism increases. Unless they are purely altruistic – treating the other as well as they treat themselves – traders continute to have a conflict of interest over where they would prefer to be along the contract curve (as is standard in the self-interest case), but they unequivolcall prefer not be at the extreme corners, which are the preferred points when traders are self-interested.

Conclusion

Social preferences play a significant role in helping us to understand a variety of behaviors, from adherence to social norms, the role of institutions in selecting focal points among multiple Nash equilibria in strategic interactions, promoting Pareto-superior outcomes over Pareto-inefficient outcomes, among many other outcomes. Teaching social preferences to undergraduates – especially in intermediate microeconomics courses – can be implemented through a set of examples that involve applications of tools with which students are familiar from these courses: constrained optimization, the construction of best-response functions, and the use of Edgeworth boxes to represent exchange. Though there are tradeoffs for any instructor in deciding what to teach, showing students that the methods of economics encompass preferences and explain behaviors that expand beyond shopping or hard-nosed negotiation to generosity and charitable giving, trade with trust and regard for one's trading partner, or altruism and community when dealing with externalities and market failures, rewards both the student and instructor alike.

References

Andreoni, James, and John Miller. 2002. "Giving According to Garp: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (2). Wiley Online Library: 737–53.

Ashraf, Nava, Iris Bohnet, and Nikita Piankov. 2006. "Decomposing Trust and Trusworthiness." *Experimental Economics* 9: 193–208.

Ashraf, Nava, Colin Camerer, and George Loewenstein. 2005. "Adam Smith, Behavioral Economist." *Journal of Economic Perspectives* 19 (3): 131–45.

Bardsley, Nicholas. 2008. "Dictator Game Giving: Altruism or Artefact?" *Experimental Economics* 11: 122–33.

Bauman, Yoram. 2011. "The Dismal Education." The New York Times (Dec 18, 2011). https://www.nytimes.com/2011/12/18/opinion/sunday/economists-are-grinches.html?hp. https://www.nytimes.com/2011/12/18/opinion/sunday/economists-are-grinches.html?hp.

Bewley, Truman F. 2009. *Why Wages Don't Fall During a Recession*. Harvard university press.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78. https://doi.org/10.1257/aer.96.5.1652.

Bhargava, Saurabh, and George Loewenstein. 2015. "Behavioral Economics and Public Policy 102: Beyond Nudging." *American Economic Review* 105 (5): 396–401.

Bowles, Samuel. 2009. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton University Press.

Bowles, Samuel, and Wendy Carlin. 2020. "What Students Learn in Economics 101: Time for a Change." *Journal of Economic Literature* Forthcoming.

Bowles, Samuel, Wendy Carlin, Simon Halliday, and Sahana Subramanyam. 2019. "What Do We Think an Economist Should Know? A Machine Learning Investigation of Research and Intermediate-Level Texts." *Santa Fe Institute Working Papers*.

Bowles, Samuel, Wendy Carlin, and Margaret Stevens. 2017. *The Economy*. Oxford University Press.

Bowles, Samuel, and Sung-Ha Hwang. 2008. "Social Preferences and Public Economics: Mechanism Design When Preferences Depend on Incentives." *Journal of Public Economics* 92: 1811–20.

Bowles, Samuel, and Sandra Polania-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature* 50 (2): 368–425.

Burdin, Gabriel, Simon Halliday, and Fabio Landini. 2018. "The Hidden Benefits of Abstaining from Control." *Journal of Economic Behavior & Organization* 147. Elsevier: 1–12.

Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3): 818–27.

Càrdenas, Juan-Camilo, and Elizabeth Ostrom. 2004. "What Do People Bring into the Game? Experiments in the Field About Cooperation in the Commons." *Agricultural Systems* 82: 307–26.

Chetty, Raj. 2015. "Behavioral Economics and Public Policy: A Pragmatic Perspective." *American Economic Review* 105 (5): 1–33.

Cooper, David J, and John Kagel. 2016. "Other-Regarding Preferences." *The Handbook of Experimental Economics* 2. Princeton University Press Princeton, NJ: 217.

Edgeworth, Francis Ysidro. 1881. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. Vol. 10. Kegan Paul.

Engel, Christoph. 2011. "Dictator Games: A Meta Study." *Experimental Economics* 14 (4). Springer US: 583–610. https://doi.org/10.1007/s10683-011-9283-7.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2002. "Appropriating the Commons - a Theoretical Explanation." In *The Drama of the Commons*, edited by Thomas Dietz, Nives Dolsak, Elinor Ostrom, Paul Stern, Susan Stonich, and Elke Weber, 157–92. National Academy Press, Washington DC.

———. 2008. "Testing Theories of Fairness - Intentions Matter." *Games and Economic Behavior* 62: 287–303.

Falk, Armin, and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior* 54 (2): 293–315.

Falk, Armin, and Michael Kosfeld. 2006. "The Hidden Costs of Control." *American Economic Review* 96 (5): 1611–30. https://doi.org/10.1257/aer.96.5.1611.

Fehr, Ernst, and Lorenz Goette. 2007. "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review* 97 (1): 298–317.

Fehr, Ernst, Lorenz Goette, and Christian Zehnder. 2009. "A Behavioral Account of the Labor Market: The Role of Fairness Concerns." *Annu. Rev. Econ.* 1 (1). Annual Reviews: 355–84.

Fehr, Ernst, and Andreas Leibbrandt. 2011. "A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons." *Journal of Public Economics* 95 (9-10). Elsevier: 1144–55.

Fehr, Ernst, Michael Naef, and Klaus M Schmidt. 2006. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment." *American Economic Review* 96 (5): 1912–7.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation." *The Quarterly Journal of Economics* 114 (3): 817–68.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25 (4): 191–210.

Halliday, Simon D. 2011. "Rarer Actions: Giving and Taking in Third-Party Punishment Games." Working Paper 211. Economic Research Southern Africa.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review* 76: 728–41.

Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiment." *Review of Economic Dynamics* 1: 593–622.

List, John. 2007. "On the Interpretation of Giving in Dictator Games." *The Journal of Political Economy* 115 (3): 482–93.

Nikiforakis, Nikos, and Helen Mitchell. 2014. "Mixing the Carrots with the Sticks: Third Party Punishment and Reward." *Experimental Economics* 17 (1). Springer US: 1–23. https://doi.org/10.1007/s10683-013-9354-z.

Ostrom, Elinor. n.d. *Governing the Commons*. Cambridge university press.

Ostrom, Elizabeth, James Walker, and Roy Gardner. 1992. "Covenants with and Without a Sword: Self-Governance Is Possible." *American Political Science Review* 86: 404–17.

———. 1994. *Rules, Games and Common-Pool Resources*. University of Michigan Press.

Paganelli, Maria Pia. 2009. "Smithian Answers to Some Puzzling Results in the Experimental Literature." In *The Elgar Companion to Adam Smith*, edited by Jeffrey T. Young. Edward Elgar.

Smith, Adam. 1982. *The Theory of Moral Sentiments*. Glasgow Edition. Liberty Fund Inc.

Smith, Vernon L. 1998. "The Two Faces of Adam Smith." *The Southern Economic Journal* 65 (1): 1–19.

Stafford, Tom. 2013. "Does Economics Make You More Selfish?" British Broadcasting Corporation, BCC Future. http://www.bbc.com/future/story/20131022-are-economists-more-selfish. http://www.bbc.com/future/story/20131022-are-economists-more-selfish.

Ziegelmeyer, Anthony, Katrin Schmelz, and Matteo Ploner. 2012. "Hidden costs of control: four repetitions and an extension." *Experimental Economics* 15 (2): 323–40. https://ideas.repec.org/a/kap/expeco/v15y2012i2p323-340.html.