Biased Preferences Equilibrium

Ariel Rubinstein
School of Economics, Tel Aviv University and
Department of Economics, New York University
and
Asher Wolinsky
Department of Economics, Northwestern University

October 24th, 2018

Abstract

We discuss economic environments in which individual choice sets are fixed and the level of a certain parameter that systematically biases the preferences of all agents is determined endogenously to achieve equilibrium. Our equilibrium concept, Biased Preferences Equilibrium, is reminiscent of competitive equilibrium: agents' choice sets and their preferences are independent of the behavior of other agents, the combined choices have to satisfy overall feasibility constraints and the endogenous adjustment of the equilibrating preference parameter is analogous to the equilibrating price adjustment. The concept is applied to a number of economic examples.

AEA Classification: D50
Keywords: Biased Preferences, Equilibrium, Economic environment

1. Introduction

The fundamental scarcity problem of economics– the tension between limited resources and insatiable wants – is resolved in conventional economic models by means of markets or other allocation mechanisms. In either case, preferences are viewed as exogenously fixed and the focus is on the role of instruments – prices or commands – that modify agents' choice sets.

This paper's approach is complementary to that framework. Agents' choice sets do not freely adjust and the resolution is obtained by a systematic bias of their preferences. Of course, if agents' preferences can be adjusted arbitrarily, then any feasible allocation can be trivially established as an "equilibrium". Hence, we focus on **systematic biases** that affect the preferences of **all** agents in the same way. Such a common bias can be interpreted as reflecting an adjustment of values or cultural norms. Consider, for example, a society whose survival depends on meeting a certain target level of a public good through voluntary contributions. If the original trade-offs between private consumption and the enjoyment from contributing are insufficient to achieve the goal, the survival pressure might trigger a social process that changes preferences so as to make contributing sufficiently attractive to achieve the target. We abstract from the details of how such a social process might work and let it just be reflected by the assumption of common adjustment of preferences.

We define an equilibrium concept - Biased Preferences Equilibrium. The environment is such that individual choice sets are fixed and the level of a certain parameter, which systematically biases the preferences of all agents, is determined endogenously to achieve equilibrium. This concept is reminiscent of competitive equilibrium: agents' choice sets and their preferences are independent of the behavior of other agents, and the combined choices must satisfy overall feasibility constraints. The endogenous adjustment of the equilibrating preference parameter is analogous to the equilibrating price adjustment in competitive equilibrium. The concept is then applied to several examples of economic environments in which limitations of various types preclude equilibrium in the absence of preference adjustment.

This approach is suggestive for situations where prices or commands might not be capable of equilibrating the system due to rigidities, restrictive social norms or free riding. However, we do not discuss how and why the preference bias mechanism might emerge instead of a price or command mechanism. Rather, we present the logic of the preference bias mechanism and demonstrate how it might work in some examples.

The closest papers to this one are Richter and Rubinstein (2015, 2018), which study interactive choice situations subject to social norms. In the first of these papers, the norm is a common ordering of all alternatives; in the second, it is a partition of the alternatives into permitted and forbidden actions. The other related literature consists of papers on endogenous preferences and norms. The main qualitative difference between this paper and the literature we are aware of is the role of preferences as endogenous equilibrating instruments in our model. In Section 5 below we will comment more specifically on the strands that we see as relatively more closely related to this paper and clarify the novelty of our approach.

2. The model and solution concept

An **economy** is a tuple $< N, Y, (X^i)_{i \in N}, F, P, \Lambda, T, (\succsim^i\ )_{i \in N} >$ where N is the set of agents, Y is the set of alternatives over which the agents' preferences are defined, $X^i \subseteq Y$ is the set of private alternatives available to agent i i.e. his "choice set", $F \subseteq Y^n$ is a subset of **feasible** profiles of agents' choices, P is a set of preference relations over Y and $(\succsim^i\ )_{i \in N}$ is a profile of preferences from P. From here on we use the notation (z^i) for $(z^i)_{i \in N}$.

For example, in an exchange economy environment one specification of these elements is as follows: Y is the set of all possible bundles, X^i is agent i's choice set, F is the set of all allocations satisfying the overall resource constraints.

Up to now the model is fairly standard. The special elements of the model are Λ and T where Λ is a set of "social values" states and the function $T : P \times \Lambda \to P$ encodes the effect of a social values state on the preferences. In other words, $\lambda \in \Lambda$ is a parameter that systematically affects the preference relations of all agents and $T(\succsim, \lambda)$ is the preferences that an agent with fundamental preferences $\succsim$ ends up with when the social values state is λ. We use the shorthand notation $\succsim_\lambda$ for $T(\succsim, \lambda)$.

Note that although agent i has to choose from the set X^i, his preferences are defined over the larger set Y. This is because it is more natural to describe the systematic change of the preference relations over X as being induced by a systematic change of the preferences over the larger space Y.

An **equilibrium** is a pair $< (\widehat{x}^i), \widehat{\lambda} >$ such that $(\widehat{x}^i) \in F$ and $\widehat{x}^i \succsim^i_{\widehat{\lambda}} y$ for all i and all $y \in X^i$.

That is, an equilibrium is a profile of choices and a social value such that: (i) the profile $(\widehat{x}^i)$ is feasible; (ii) each agent's choice $\widehat{x}^i$ is individually optimal with respect to the endogenously determined preferences $\succsim^i_{\widehat{\lambda}}$.

The standard Pareto efficiency notion with respect to the fundamental preferences will be called pre-efficiency. That is, a profile (x^i) is **pre-efficient** if there is no other $(y^i) \in F$ such that $y^i \succsim^i x^i$ for all i with strict inequality for at least one agent. Obviously, any equilibrium $< (\widehat{x}^i), \widehat{\lambda} >$ is post-efficient (in the sense that there is no other $(y^i) \in F$ such that $y^i \succsim^i_{\widehat{\lambda}} \widehat{x}^i$ for all i with strict inequality for at least one of them) since $\widehat{x}^i$ is a maximizer over the entire set X^i of $\succsim^i_{\widehat{\lambda}}$.

Comment: A more general version of the model would allow the agents' choice sets to be endogenous. Then, the parameter λ would have a broader interpretation and agent i's choice set would be $X^i(\lambda)$. For example, in an exchange economy environment, λ might consist of a preference bias parameter, as above, and a price vector. The sets $X^i(\lambda)$ would then be the agents' budget sets determined by their respective endowments and the price vector. While such more general version might lead to more interesting applications, it is not required for the simple examples discussed below and we will therefore continue with the assumption of fixed exogenous X^i.

3. The Shapley-Shubik Assignment Model

Consider the Shapley and Shubik (1971) assignment model: N is a set of n individuals; H is a set of n houses; and Y is the set of all pairs (h,m) where $h \in H$ and $m \in R$ is a monetary transfer to the agent. We use the symbol h to also denote the pair $(h,0)$. Let $X^i = H$ for all i and let F be the set of profiles such that no house is allocated to two individuals. Let P be the set of all preferences over Y represented by a utility function of the form $v(h) + m$. Therefore, a preference relation in P is characterized by a vector of positive numbers $(v(h))_{h \in H}$ where $v(h)$ is the rate of substitution between the house h and money. Note that, although agents have to choose a house from H, their preferences are induced by their preferences over a larger domain that also involves money transfers. Denote by v^i agent i's utility function over Y.

Let $\Lambda = R_+^H$. A $\lambda \in \Lambda$ stands for a uniform systematic change of the rates of substitution between houses and money. The transformation $T(v,\lambda)$ multiplies each rate of substitution between h and money by $\lambda(h)$ in a uniform manner across all individuals. That is, $T(v,\lambda)$ is the preference relation in P characterized by the function $\lambda(h)v(h) + m$.

Claim 1. (i) An equilibrium exists. (ii) All equilibria are pre-efficient.

Proof: (i) The proof makes straightforward use of Shapley-Shubik's original theorem. Consider the Shapley-Shubik house allocation setting with preferences characterized by $u^i(h) = \ln v^i(h)$. By the Shapley-Shubik theorem, there is a vector of monetary transfers $(p_h)_{h \in H}$ and a feasible profile (h^i) such that, for each agent i, the assigned house h^i maximizes $u^i(h) + p_h$ over the set $B(p) = \{(h,p_h) \mid h \in H\}$. This implies that h^i also maximizes $e^{u^i(h)+p_h} = e^{p_h}v^i(h)$ over $B(p)$, for each i. Therefore, $< (h^i),(e^{p_h})_{h \in H} >$ is an equilibrium.

(ii) Note that if $< (h^i),(\lambda(h))_{h \in H} >$ is an equilibrium with the utilities (w^i), then (h^i) is Pareto efficient for any vector of biased utilities (w^i_λ). Otherwise, for some λ, there is a feasible allocation (x^i) such that
$w^i_\lambda(x^i) = \lambda(x^i)w^i(x^i) \geq w^i_\lambda(h^i) = \lambda(h^i)w^i(h^i)$ for all i with at least one inequality. But then $\Pi_{i \in N}\lambda(x^i)w^i(x^i) > \Pi_{i \in N}\lambda(h^i)w^i(h^i)$ and since both (h^i) and (x^i) are allocations, the two products must be equal. Thus, the equilibrium is pre-efficient. ■

Comment: If λ transforms the preferences in a more elaborate way, then there can be multiple equilibria, some of them may not be pre-efficient. For example, suppose that λ is a vector $\lambda(h) = (\alpha(h),\beta(h))$ such that $T(v,(\alpha,\beta))(h) = \alpha(h) + \beta(h)v(h)$. Order the houses arbitrarily $h_1,..,h_n$. First fix $\beta(h_1)$ such that the difference between the highest value of $\beta(h_1)v_i(h_1)$ and the second highest value is between 10^{2n} and 10^{2n-1}. Continue with h_2 so that the corresponding difference is between $10^{2(n-1)}$ and $10^{2(n-1)-1}$ and so on. Then, apply the Shapley-Shubik result to calculate the equilibrium monetary transfers $(\alpha(h))_{h \in H}$. Since the equilibrium assignment maximizes the sum of values, it must be describable by a sequential allocation procedure whereby at the m-th stage the individual with the highest value from among those who were not assigned earlier to $h_1,\ldots,h_{m-1}$ is assigned to h_m. This allocation may not be pre-efficient. To see this, consider the case of two individuals 1 and 2 and $H = \{a,b\}$ with $v^1(a) = 1$, $v^1(b) = 0$, $v^2(a) = 2$ and $v^2(b) = 3$. Applying the above procedure with the order $(h_1,h_2) = (a,b)$ results in an equilibrium allocation (b,a), which is Pareto-dominated by (a,b) according to the fundamental preferences.

4. An Exchange Economy with Fixed Prices

Consider an exchange economy with K goods. Each agent i has an initial endowment w^i and let denote the total endowment as w. Trade can take place only at the fixed prices $p = (p_k)$. In this case $Y = R^K_+$ (the set of all bundles), $X^i = \{x \mid px = pw^i\}$ for all i, and $F = \{(x^1,\ldots,x^n) \mid \sum_i x^i = w\}$. Let P be the set of all preference relations represented by a separable, differentiable, strictly increasing (in each argument) and concave function. Let U denote the set of all such functions and let $u^i \in U$ denote the utility function representing agent i's preferences. The set Λ consists of all vectors $\lambda = (\lambda_1,\lambda_2,..,\lambda_K) \in R^K_{++}$. Given $u \in U$ such that $u(x) = \sum_{j=1,..,K} v_j(x)$, let $T(u,\lambda)$ be the preferences represented by $u_\lambda(x) = \sum_{j=1,..,K} \lambda_j v_j(x_j)$.

Notice the similarity between the preference bias in this case and that of the previous section. Here, the subjective marginal rates of substitution of all individuals between any good k and good 1 are biased by the same multiplier. There, too, the subjective rate of substitution for all individuals between any particular house h and money are biased by the same multiplier.

Claim 2. (i) The model has an equilibrium. (ii) All equilibria are pre-efficient.

Proof. For simplicity, we prove the proposition for the case in which every bundle in every efficient allocation is either strictly positive or all zero. To see the intuition behind this, consider first the $n = K = 2$ case depicted by the Edgeworth box in Figure 1:

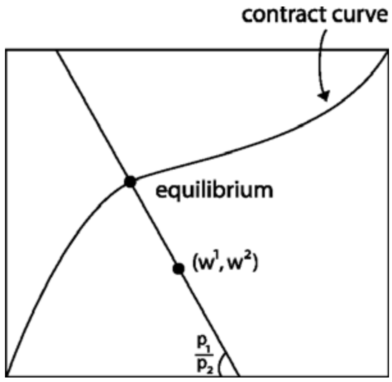

Figure 1

Let E be the intersection point of the contract curve and the line sloped $-p_1/p_2$ through the initial endowment. Let (e^1,e^2) be the allocation at E. Since the contract curve is continuous and connects the two opposing corners of the box, such an intersection point exists and obviously $(e^1,e^2) \in F$. Denote by $MRS(u)(x)$the marginal rate of substitution of the function u at the bundle x. Then, since the allocation (e^1,e^2) is efficient $MRS(u^1)(e^1) = MRS(u^2)(e^2) = \mu$. Let $(\widehat{\lambda}_1,\widehat{\lambda}_2) = (\frac{p_1}{\mu p_2},1)$ and $\widehat{x}^i = e^i$. Then, $MRS(u^i_{\widehat{\lambda}})(\widehat{x}^i) = p_1/p_2$ and thus the pair $< (\widehat{x}^i),\widehat{\lambda} >$ is an equilibrium.

For the general case, Balasko (1979) showed that for every price vector p, there is a Pareto-efficient allocation (e^i) such that $pe^i = pw^i$ for all i. By the Pareto efficiency of (e^i), there is a vector $\mu = (\mu_1,\mu_2,\ldots,\mu_K)$ such that $\nabla u^i(e^i) = c^i\mu$ for each agent i. Let $\widehat{\lambda}$ be a vector such that $\widehat{\lambda}_k\mu_k = p_k$ and $\widehat{x}^i = e^i$. Then, $\nabla u^i_{\widehat{\lambda}}(\widehat{x}^i) = c^i p$ for all i and therefore, $\widehat{x}^i$ maximizes u^i_{λ} over X^i. Thus, $< (\widehat{x}^i),\widehat{\lambda} >$ is an equilibrium. Pre-efficiency follows from the choice of $(\widehat{x}^i)$. ■

In the case of $K = 2$ this result can be extended to preferences represented by any differential utility function (not just separable). This is because, for any function u and for any positive number λ, there is a utility function u_λ such that $MRS(u_\lambda)(x) = \lambda MRS(u)(x)$. However, this is not generally true when $K > 2$.

Our (fixed price) equilibrium outcome might be inferior to the initial endowment for some agent in terms of the original preferences. Obviously, this is in contrast to the Pareto superiority of any competitive equilibrium outcome to the initial endowment (in terms of the fundamental preferences). That is, while the point E in the $n = K = 2$ case presented in the beginning of the proof is pre-efficient, it might not be in the core.

Figure 2 depicts an illustration of the equilibrium for the case of $n = K = 2$ and with linear fundamental preference $\alpha^i x_1 + x_2$. We present one possible configuration in which $p_1/p_2 > \alpha^1 > \alpha^2 > 0$. The Pareto-efficient allocations are all points on the south and east edges of the Edgeworth box.

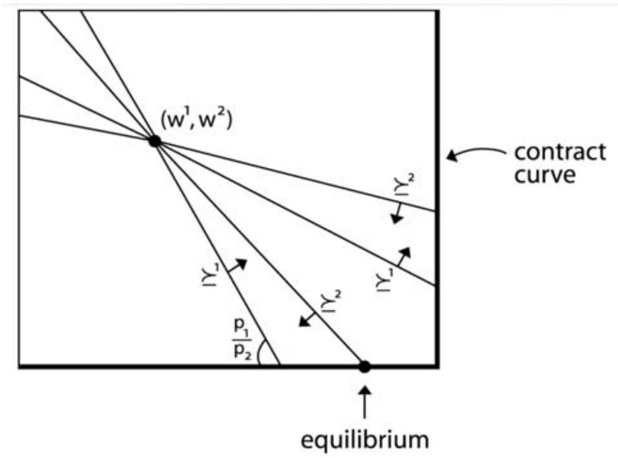

Figure 2

To obtain an equilibrium, the relative value of good 1 needs to increase. An equilibrium λ transformation of the preferences rotates both indifference curves clockwise making them steeper so that Agent 2's indifference curve exactly coincides with the budget line and hence Agent 1's indifference curve is even steeper. In the resulting equilibrium allocation, Agent 1 gets only good 1 and Agent 2 gets all of good 2 and the remainder of good 1. Note that in this equilibrium, Agent 1 is worse off in terms of the original preferences.

If the budget line does not coincide with the (negatively sloped) diagonal, then this is the only equilibrium preference transformation. In the special case that the budget set coincides with the minor diagonal, there is a continuum of rotations that support an allocation in which each agent gets the entire stock of a distinct good.

The idea of equilibrium with fixed prices is the subject of a large literature (see for example Benassy (1986) and the references there). The novelty in our model is that preferences, in contrast to rationing schemes, adjust to equilibrate the market.

5. Giving with Pride and Receiving with Shame

The stock of a single consumption good is distributed across a society consisting of a unit mass (continuum) of agents. A public fund is dedicated to redistribution in this society. Agents voluntarily contribute to the fund and can freely withdraw from it. Donors experience pride while donees experience shame. The fund has to be balanced, i.e. contributions must equal withdrawals.

The initial ownership of the good is described by a distribution function G on the interval $[0,M]$ with strictly positive density g. The set Y is the set of pairs (c,d) where c is a non-negative amount of consumption and d is the transfer to (positive d) or from (negative d) the public fund. An agent with wealth w faces the choice set $X^w = \{(c,d) \mid c+d = w\}$. The set F is the set of all profiles $(c^w,d^w)_{w\in[0,M]}$ such that $\int_0^M d^w dG(w) = 0$. The set P contains all preferences represented by a utility function of the form $U_\alpha(c,d) = \alpha u(c) + r\max[d,0] + s\min[d,0]$, where u is strictly increasing and concave, $u'(0) = \infty$, $u'(M) = 0$ and $\alpha,r,s > 0$. The coefficient r captures the pride of donating and the coefficient s captures the shame associated with withdrawing. It is assumed that $s \geq r$, which makes it suboptimal for an agent to both give and take. All agents have the preference relation represented by the utility function U_1.

Let $\Lambda = (0,\infty)$ and let $T(U,\lambda)$ transform $U_1(c,d)$ to $U_\lambda(c,d)$. Thus, λ captures the degree of sensitivity to pride and shame. All of U_λ's marginal rates of substitution between consumption and the mental cost of giving or receiving are λ multiplies of the corresponding magnitudes of U_1.

Claim 3. (**i**) The model has a unique equilibrium $< (\widehat{c}^w,\widehat{d}^w)_{w\in[0,M]},\widehat{\lambda} >$.

(**ii**) Let $\underline{w}$ and $\overline{w}$ satisfy $\widehat{\lambda}u'(\underline{w}) = s$ and $\widehat{\lambda}u'(\overline{w}) = r$, respectively. The equilibrium profile is

$$\widehat{c}^w = \begin{cases} \underline{w} & \text{if} \quad w \leq \underline{w} \\ w & \text{if} \quad \underline{w} < w < \overline{w} \\ \overline{w} & \text{if} \quad w \geq \overline{w} \end{cases}$$

and $\widehat{d}^w = w - \widehat{c}^w$.

(**iii**) The equilibrium is pre-efficient.

Proof: (**i**) & (**ii**). Given $w \in [0,M]$ and $\mu \in \Lambda$, the assumptions on u imply that the agent's maximization problem has a unique optimal consumption choice $c(w,\mu)$. Let $\Psi(\mu) = \int_0^M [c(w,\mu) - w]dG(w)$, i.e. the excess demand for consumption given μ. Since $c(w,\mu)$ is continuous and increasing in μ, so is $\Psi(\mu)$. For μ near 0, $\Psi(\mu) < 0$ and for large μ, $\Psi(\mu) > 0$. Let $\widehat{\lambda}$ be the unique value such that $\Psi(\widehat{\lambda}) = 0$. The assumptions on u guarantee the existence of unique values $\underline{w}$ and $\overline{w}$ satisfying $\widehat{\lambda}u'(\underline{w}) = s$ and $\widehat{\lambda}u'(\overline{w}) = r$, respectively. Then, $< (\widehat{c}^w, \widehat{d}^w), \widehat{\lambda} >$ where $\widehat{c}^w = c(w,\lambda)$ and $\widehat{d}^w = w - \widehat{c}^w$ is an equilibrium.

(**iii**) Consider an equilibrium $< (c^w, d^w), \lambda >$ with $\lambda > 1$. Given the fundamental preferences, each agent w weakly prefers to consume less than $c(w)$ and depending on w prefers to receive less or donate more. Therefore, a Pareto-superior allocation would have to weakly lower everybody's consumption. But this contradicts the feasibility that requires exact budget balance of the public fund (and precludes disposal). An analogous argument holds for the case of $\lambda < 1$.■

Note that when $r = s$, in the unique equilibrium all agents consume the same amount $\int_0^M wdG(w)$.

6. Discussion

A. **Game version**. In the model and the examples, each agent chooses his action optimally ignoring the actions of others. The social norms expressed by λ adjust to assure that agents' choices satisfy the feasibility constraint. This approach is in the spirit of competitive equilibrium, but the idea of equilibrating adjustment of preferences is of course just as relevant for environments with strategic interaction. A game version of the model differs from the "competitive" version in that the preferences of each agent are over profiles of actions of all agents. The set F may be viewed here as a collection of acceptable or desirable profiles (e.g., Pareto optimal or equitable) in the sense that a play of profiles outside F generate pressure on the preferences to adjust. The equilibrium concept is modified accordingly. An equilibrium is a pair $< (\widehat{x}^i)_{i \in N}, \widehat{\lambda} >$ such that: $(\widehat{x}^1, \ldots \widehat{x}^N) \succsim^i_{\widehat{\lambda}} (\widehat{x}^1, \ldots y^i, \ldots, \widehat{x}^N)$ for all $y^i \in X^i$ and $(\widehat{x}^i) \in F$. Thus, equilibrium imposes a standard Nash equilibrium condition on the agents' action profile with respect to the endogenously determined preferences and the equilibrium is in the set of acceptable profiles.

B. **Comments on the Literature**. The paper is related to the literature on endogenous evolution of preferences (e.g., Dekel, Ely and Yilankaya (2007) and Alger and Weibull (2016)). Whereas that literature looks at dynamic evolution of preferences conducive to social interaction, we look at static equilibrium resolution of resource allocation problems.

The paper is also somewhat related to the literature on the use of honors to incentivize agents (e.g., Benabou and Tirole (2003), Tirole (2016) and Dubey and Geanakoplos (2017)) or the role of status in the allocation of labor across occupations (Fershtman, Murphy and Weiss (1996)). These models feature agents who value the attainment of "status" within their peer group. Their preferences over status and other goods are exogenously fixed as in conventional economic models. It is the status associated with the different actions that is determined within these models, either via the emergence of an equilibrium convention or via deliberate design by a principal. In contrast, in our approach, the preferences adjust endogenously (in a specific uniform fashion) to equilibrate the economy. This role of the preferences as endogenous equilibrating instruments is the main qualitative difference between our model and theirs.

The idea of endogenous preference change has been considered in different areas of economics. Since these literatures are vast, diverse and quite distant from our work, we avoid throwing in a few random references. But let us discuss in general terms the relationship between our approach and the main ideas of those other approaches. One strand of this literature looks at the causes and consequences of preference changes induced by, for example, addiction, habitual behavior, advertising, and fashions. Roughly speaking, the "technology" that induces the preference change at the individual level is taken as given (e.g., the effect of addictive substance consumption) and the outcomes that emerge in their presence are examined. In contrast, in our approach the preference change is a social phenomenon that responds to resource pressures and serves in a harmonizing role reminiscent of the role of prices in a competitive market.

Another strand looks at deliberate strategic manipulation of the preferences that is inherent to the two-stage game model. The main element of these models is each player's strategic selection (in the first stage) of her own second stage preferences to affect favorably the equilibrium, anticipating the selections made by others. This is obviously different from the nature and role of the preference change in our model, where the changes in preferences are not chosen by the agents, they affect all the agents in the same way and they emerge in response to pressures to equilibrate a social allocation problem.

References

Alger, Ingela and Jorgen W. Weibull (2016), "Evolution and Kantian Morality," *Games and Economic Behavior,* 98, 56–67.

Balasko, Yves (1979), "Budget-constrained Pareto-efficient allocations," *Journal of Economic Theory* 21, 359-379.

Benabou, R. and J. Tirole (2003), “Intrinsic and Extrinsic Motivation,” *Review of Economic Studies*, 70, 489 - 520.

Benassy, Jean-Pascal (1986), *Macroeconomics: An Introduction to the Non-Walrasian Approach*, Elsevier.

Dekel, Eddie., Ely, Jeffrey. and Okan Yilankaya (2007), "Evolution of Preferences," *Review of Economic Studies*, 74, 685–704

Dubey, Pradeep K. and John Geanakoplos (2017), "Money and Status: How Best to Incentivize Work," Cowles Foundation Discussion Paper No. 1954R2.

Fershtman, Chaim, Kevin M. Murphy, and Yoram Weiss (1996). "Social Status, Education, and Growth," *Journal of Political Economy*, 104, 108-132.

Richter, Michael and Ariel Rubinstein (2015), "Back to Fundamentals: Convex Geometry and Economic Equilibrium," *American Economic Review*, 105, 2570-2594.

Richter, Michael and Ariel Rubinstein (2018), "Normative Equilibrium: The permissible and the forbidden as devices for bringing order to economic environments," Working paper.

Shapley, Lloyd and Martin Shubik (1971), "The assignment game I: The Core," *International Journal of Game Theory*, 1, 111–130.

Tirole, Jean (2016), "On Some Social determinants of behavior and Institutions," Notes for Nemmers Short Course, Northwestern University.