

ARTICLE

# Biased preferences equilibrium

Ariel Rubinstein<sup>1,\*</sup>  and Asher Wolinsky<sup>2</sup>

<sup>1</sup>School of Economics, Tel Aviv University, Tel Aviv, Israel 69978 and Department of Economics, New York University, NY NY 10012, USA and <sup>2</sup>Department of Economics, Northwestern University, Evanston, IL 60208, USA

\*Corresponding author. Email: [ariel@tauex.tau.ac.il](mailto:ariel@tauex.tau.ac.il)

(Received 23 July 2020; revised 18 October 2020; accepted 24 October 2020)

## Abstract

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

**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 allocation mechanisms. In either case, preferences are viewed as exogenously fixed and the focus is on the role of instruments – prices in decentralized markets or commands in centralized mechanisms – that determine agents' choice sets. The approach presented here is complementary to that standard framework.

In our model, agents' choice sets do not adjust nor is there any central body dictating the rules. The resolution of the fundamental scarcity problem is achieved by means of a systematic adjustment of individuals' preferences. These changes give rise to an equilibrium that achieves harmony between agents' choices and the social feasibility constraints. It is reminiscent of Aesop's 'The Fox and the Grapes' fable where the fox's preferences change when it cannot reach the grapes.

If agents' preferences can adjust arbitrarily, then any feasible allocation can be trivially established as an 'equilibrium'. However, we study the more interesting case in which systematic biases affect the preferences of all agents in the same way, which can be interpreted as reflecting an adjustment of social values or

norms. Obviously, any change in preferences will be somewhat arbitrary. Since marginal rates of substitution provide a primitive description of preferences, we model changes in preferences as common biases in the marginal rates of substitution. That is, a change in preferences is achieved by multiplying all agents' marginal rates of substitution by the same positive number. In order to avoid technical issues that are less relevant for this purely conceptual discussion, we restrict attention to separable preferences, such that the bias is captured by the weights by which the different sub-utilities are multiplied.

Individual choice sets are fixed. The profile of choices have to be feasible in the sense specified in the model. We introduce a new equilibrium concept – Biased Preferences Equilibrium – in which the adjustment of all agents' preference relations is determined endogenously so that the profile of individually optimal choices will be feasible. The concept is applied in three examples of economic environments in which limitations of various types preclude equilibrium in the absence of preferences adjustment.

The equilibrium concept is reminiscent of competitive equilibrium. As in the standard competitive equilibrium model, an agent's choice set and preferences are independent of the behaviour of other agents, and the combined choices must satisfy overall feasibility constraints. The endogenous adjustment of the equilibrating weights is analogous to the equilibrating price adjustment in competitive equilibrium. The conceptual exercise we carry out illustrates the dual role played by 'adjustment of preferences' and 'adjustment of choice sets'.

**The following paragraph expresses Ariel's view only.** The model should not be viewed as 'normative' – it is not being claimed that economic conflicts should be left to resolve themselves through the adjustment of preferences. Neither is it 'positive' in the sense that we do not provide evidence that real-world phenomena are described by the model. The analysis is a purely 'conceptual endeavour' and its goal (like that of other papers mentioned in section 4) is to demonstrate that non-price-based social mechanisms can also bring order to classical economic situations.

## 2. The model and the solution concept

An economy is a tuple  $\langle N, Y, (X^i)_{i \in N}, F, P, (\succsim^i)_{i \in N}, \Lambda, T \rangle$  where:

$N$  is the set of agents;

$Y \subseteq R^K$  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  $i$  (his 'choice set');

$F \subseteq \prod_{i \in N} X^i$  is the set of *feasible* profiles of agents' choices;

$P$  is the set of preferences over  $Y$  that are represented by a function of the type  $\sum_{k=1, \dots, K} v_k(y_k)$  with positive functions  $v_k$ , and, whenever relevant,  $v_k$  is an increasing, differentiable and concave function;

$(\succsim^i)_{i \in N}$  is a profile of preference relations in  $P$ .

Hereafter, we write  $(z^i)$  instead of  $(z^i)_{i \in N}$ .

Up to this point, the model is fairly standard. The non-standard elements of the model are  $\Lambda$  and  $T$ :

$\Lambda = R_+^K$  is the set of social-value states;

$T : P \times \Lambda \rightarrow P$  encodes the effect of a social-value state on preferences. If a preference relation  $\succsim$  is represented by  $\sum_{k=1, \dots, K} v_k(y_k)$ , then  $T(\succsim, \lambda)$  is represented by  $\sum_{k=1, \dots, K} \lambda_k v_k(y_k)$ . We use the notation  $\succsim_\lambda$  for  $T(\succsim, \lambda)$ .

Note that although agent  $i$  chooses from the set  $X^i$ , his preferences are defined over the larger set  $Y$ .

An *equilibrium* is a pair  $\langle (\hat{x}^i), \hat{\lambda} \rangle$  such that  $(\hat{x}^i) \in F$  and  $\hat{x}^i \succsim_\lambda^i y$  for all  $i$  and all  $y \in X^i$ . That is, an equilibrium is a profile of choices and a social-value state such that: (i) the profile  $(\hat{x}^i)$  is feasible; and (ii) each agent's choice  $\hat{x}^i$  is individually optimal with respect to the endogenously determined biased preferences  $\succsim_\lambda^i$ .

The standard Pareto efficiency notion with respect to the fundamental preferences will be referred to as pre-efficiency. A profile  $(x^i) \in F$  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, there is no interesting content to efficiency ex-post here since in equilibrium each individual makes his own optimal choice from his fixed choice set.

### 3. Three examples

#### 3.1. An exchange economy with fixed prices

There is a large literature on models of equilibrium with fixed prices (see for example Benassy (1986) and the references therein). The novelty in our setting lies in the approach that it is the adjustment of preferences that equilibrate the market, rather than rationing schemes.

Consider an exchange economy with  $K$  goods and fixed prices. Agent  $i$ 's initial endowment is  $w^i$ , and  $w$  denotes the total endowment. Trade can take place only at the fixed prices  $p = (p_k)$ . In this case,  $Y = R_+^K$  (the set of all bundles),  $X^i = \{x \in Y \mid px = pw^i\}$  for all  $i$ , and  $F = \{(x^1, \dots, x^n) \in \Pi_i X^i \mid \sum_i x^i = w\}$ . Let  $P$  be the set of all preference relations represented by a utility function of the form  $u(y) = \sum_{k=1, \dots, K} v_k(y_k)$  where each  $v_k$  is a differentiable, strictly increasing and concave function. Given that  $\succsim$  is represented by such  $u$ , the preferences  $T(\succsim, \lambda)$  are represented by  $u_\lambda(y) = \sum_{k=1, \dots, K} \lambda_k v_k(y_k)$ .

**Claim 1.** (i) *The model has an equilibrium.* (ii) *All equilibria are pre-efficient.*

**Proof.** (i) 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 result, consider first the  $n = K = 2$  case depicted by the Edgeworth box in Figure 1:

Let  $(e^1, e^2)$  be the allocation corresponding to the intersection point of the contract curve (the set of all Pareto efficient allocations) and the line with slope  $-p_1/p_2$  through the initial allocation. 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 a 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  $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2) = (\frac{p_1}{\mu p_2}, 1)$ . Then,  $MRS(u_\lambda^i)(e^i) = p_1/p_2$  and thus the pair  $\langle (e^i), \hat{\lambda} \rangle$  is an equilibrium.

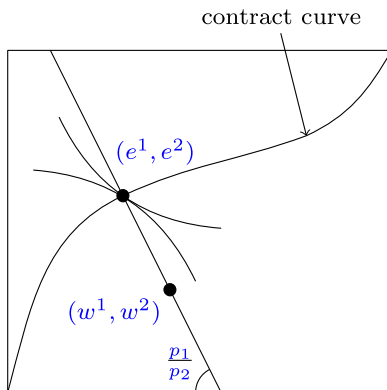


Figure 1. Equilibrium in Edgeworth Box.

For the general case, Keiding (1981) (following Balasko 1979) showed that for every price vector  $p$ , there is a vector  $\mu = (\mu_1, \mu_2, \dots, \mu_K)$  and an allocation  $(e^i)$  such that  $pe^i = pw^i$  for all  $i$  and if  $x^i \succ^i e^i$  then  $\mu x^i \succ pe^i$ . Therefore, for each  $i$  there is a positive number  $c^i$  such that  $\nabla u^i(e^i) = c^i \mu$ . Let  $\hat{\lambda} = (p_k/\mu_k)$ . Then,  $\nabla u^i_{\hat{\lambda}}(e^i) = c^i p$  for all  $i$  and therefore,  $e^i$  maximizes  $u^i_{\hat{\lambda}}$  over  $X^i$ . Thus,  $\langle (e^i), \hat{\lambda} \rangle$  is an equilibrium.

(ii) Let  $\langle (\hat{x}^i), \lambda \rangle$  be an equilibrium. For each  $i$ , the bundle  $\hat{x}^i$  maximizes the function  $u^i_{\lambda}(x) = \sum_{k=1, \dots, K} \lambda_k v^i_k(x_k)$  over  $X^i = \{x | px = pw^i\}$ . Given the differentiability of the utility function  $u^i$ , the bundle  $\hat{x}^i$  maximizes the function  $u^i(x) = \sum_{k=1, \dots, K} v^i_k(x_k)$  over  $\{x | p^{\lambda} x = p^{\lambda} \hat{x}^i\}$  where  $p^{\lambda}_k = p_k/\lambda_k$  for every commodity  $k$ . Thus,  $(\hat{x}^i)$  is pre-efficient. ■

**Individual Rationality:** Although the equilibrium outcome is pre-efficient, it might be inferior to the initial endowment according to the fundamental preferences of some agent. This is in contrast to the Pareto superiority of any competitive equilibrium outcome over the initial endowment.

Figure 2 depicts an illustration of the equilibrium for the case of  $n = K = 2$  with linear fundamental utility functions  $\alpha^i x_1 + x_2$ . 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.

To obtain an equilibrium, the relative value of good 1 needs to increase. An equilibrium  $\lambda$ -transformation of the preferences rotates both indifference curves clockwise, thus increasing their slopes, such 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 receives only some of good 1 and Agent 2 receives all of good 2 and the remainder of good 1. In terms of the original preferences, agent 1 is worse off in this equilibrium than with his initial endowment.

If the budget line does not connect the initial allocation to the allocation  $z$  where agent 1 receives the entire stock of good 1 and agent 2 receives the entire stock of good 2, then this is the only equilibrium  $\lambda$ . In the special case that the budget set contains  $z$ , there is a continuum of  $\lambda$  values that support the allocation  $z$ .

**Exchange economies with non-convex preferences:** Exchange economies with non-convex preferences may not have a competitive equilibrium. Nevertheless,

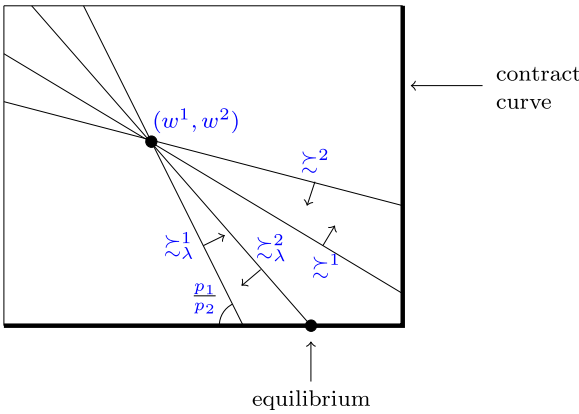


Figure 2. Equilibrium in Edgeworth Box with linear indifference curves.

there may exist a set of prices and a social value state for which an equilibrium exists. Thus, a price system and social values that distort the preferences systematically may serve as a means of attaining stability in markets where prices on their own cannot.

Consider, for example, the standard model of an exchange economy with two agents who have initial bundles  $(1, 2)$  and  $(3, 2)$  respectively and identical non-convex preferences represented by the function  $u(x_1, x_2) = x_1^2 + x_2^2$ . This exchange economy does not have a competitive equilibrium since, whatever the price is, each agent’s optimal bundle will have only one good in it. Therefore, in equilibrium the agents must be indifferent between the two corners of the budget line, which means that the prices must be  $(1, 1)$ . This implies that Agent 1’s consumption is either  $(3, 0)$  or  $(0, 3)$  and agent 2’s is either  $(5, 0)$  or  $(0, 5)$ . Therefore, their consumption bundles do not sum up to the total bundle  $(4, 4)$ .

In contrast, the vector of prices  $(2, 1)$  combined with  $\lambda = (4, 1)$  is an equilibrium in our model. This  $\lambda$  transforms the agents’ preferences into preferences represented by the utility functions  $4x_1^2 + x_2^2$  (with the indifference curves transformed from quarter-circles into quarter-ellipses). In this case, both individuals are indifferent between the corners of their budget lines (the bundles  $(2, 0)$  and  $(0, 4)$  for Agent 1 and the bundles  $(4, 0)$  and  $(0, 8)$  for Agent 2). Thus, assigning  $(0, 4)$  to Agent 1 and  $(4, 0)$  to Agent 2 completes the description of an equilibrium.

### 3.2. The Shapley–Shubik assignment model without money transfer

Our second example reinterprets the Shapley and Shubik (1971) assignment model, where  $n$  houses have to be allocated to  $n$  individuals with possible conflicting preferences. Harmony is achieved by the assignment of prices to houses so that each individual prefers a different house. We show that this society can instead achieve harmony through systematic bias of the individuals’ valuations of the houses rather than through prices.

In the model,  $N$  is a set of  $n$  individuals;  $H = \{1, \dots, n\}$  is a set of houses;  $Y = H \times R$  is the set of all house-money transfer pairs; for each  $i$ ,  $X^i$  is the set of all pairs in  $Y$  with zero money transfers, i.e.  $X^i = \{(h, 0) | h \in H\}$ ; and  $F$  is the

set of profiles  $(x^i) \in \prod_{i \in N} X^i$  such that for any house  $h$  there is a unique  $i$  for which  $x^i = (h, 0)$ . Thus, an element of  $F$  is an allocation of  $H$  among the individuals with no money transfers. As in the Shapley–Shubik model, the preferences are defined over  $Y$ , a larger domain than  $\prod_{i \in N} X^i$  since it is used to describe the preference bias function  $T$ .

Any preference relation in  $P$  is represented by a utility function of the form  $v(h) + m$  where  $v(h) > 0$  for all  $h \in H$ . Thus, every preference relation is characterized by a vector  $(v(h))_{h \in H}$ . Let  $\Lambda = R_+^n$ . The transformation  $T(v, \lambda)$  maps the preferences characterized by  $(v(h))_{h \in H}$  to the preferences characterized by  $(\lambda(h)v(h))_{h \in H}$ .

Notice that this model fits into our general framework as follows: Let  $e_h$  denote the vector that has 1 in the  $h$  coordinate and 0 elsewhere and let  $E$  denote the set of these unit vectors. We identify a house  $h$  with the vector  $e_h \in H$ . Therefore,  $Y = E \times R$ ;  $X^i = \{(e_h, 0) | h \in H\}$ ;  $F = \{x \in \prod_{i \in N} X^i \mid \sum x^i = (1, 1, \dots, 1, 0)\}$ . Any preference relation in  $P$  is represented by a utility function of the form  $U(y) = \sum_{h \in H} y_h v(h) + y_{n+1}$ . The transformation  $T(U, \lambda)$  is given by a vector  $(\lambda(h))_{h \in H}$  as in the general framework (setting  $\lambda_{n+1} = 1$  w.l.o.g):  $T(U, \lambda)$  is a preference relation represented by  $\sum_{h \in H} y_h v(h) \lambda(h) + y_{n+1}$ .

**Claim 2.** (i) *An equilibrium exists.* (ii) *All equilibria are pre-efficient.*

**Proof:** (i) Consider the Shapley–Shubik house allocation setting with preferences characterized by  $u^i(h) = \ln v^i(h)$ . By Shapley–Shubik (1971), there is a vector of monetary transfers  $(p_h)_{h \in H}$  and an allocation  $(h^i)$  such that, for each agent  $i$ , the assigned house  $h^i$  maximizes  $u^i(h) + p_h$  over the set  $B(p) = \{(h, p_h) | h \in H\}$ . 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 a Biased Equilibrium.

(ii) Let  $((h^i), (\lambda(h))_{h \in H})$  be a Biased Equilibrium in the economy with  $i$ 's preferences being represented by  $(w^i(h))$ . Assume that  $(h^i)$  is not pre-efficient. Then, there is a feasible allocation  $(x^i)$  such that  $w^i(x^i) \geq w^i(h^i)$  for all  $i$  with at least one strict inequality and then  $\prod_{i \in N} w^i(x^i) > \prod_{i \in N} w^i(h^i)$ . Since both  $(h^i)$  and  $(x^i)$  are allocations of the  $n$  houses,  $\prod_{i \in N} \lambda(x^i) = \prod_{i \in N} \lambda(h^i)$ . Thus,  $\prod_{i \in N} \lambda(x^i) w^i(x^i) = \prod_{i \in N} \lambda(x^i) \prod_{i \in N} w^i(x^i) > \prod_{i \in N} \lambda(h^i) \prod_{i \in N} w^i(h^i) = \prod_{i \in N} \lambda(h^i) w^i(h^i)$  which implies that for at least one agent  $i$ ,  $\lambda(x^i) w^i(x^i) > \lambda(h^i) w^i(h^i)$  contradicting  $h^i$  being the maximizer of  $\lambda(h) w^i(h)$  over all houses. ■

### 3.3. Giving with pride and receiving with shame

The third example is a society in which individuals give and take freely from a public fund, which is used to achieve redistribution. Agents voluntarily contribute to the fund and can freely withdraw from it. Donors experience pride while withdrawers experience shame. The fund has to be balanced, i.e. contributions must equal withdrawals. We will see that balance can be achieved by  $1, \dots, n$  adjustment of the weights assigned to pride and shame.

In the model, the stock of a single consumption good is distributed across a society consisting of agents. Agent  $i$  initially owns  $w^i$  and  $w^1 < w^2 < \dots < w^n$ .

The set  $Y = R_+ \times R$  is the set of pairs  $(c, d)$  where  $c$  is a non-negative amount of consumption and  $d$  is a donation to the public fund. A negative  $d$  means a withdrawal from the fund. An agent faces the choice set  $X^i = \{(c, d) \in Y \mid c + d = w^i\}$ . The set  $F$  is the set of all profiles  $(c^i, d^i)_{i \in N}$  such that  $\sum_i d^i = 0$ .

The set  $P$  contains all preferences represented by a utility function of the form  $U(c, d) = v_1(c) + v_2(d)$  where  $v_2(s) = r \max[d, 0] + s \min[d, 0]$ ,  $v_1'(0) = \infty$ ,  $v_1'(M) = 0$  and  $\alpha, r, s > 0$ . The coefficient  $r$  captures the pride of donating while the coefficient  $s$  captures the shame of withdrawing. It is assumed that  $s \geq r$  and that all agents have the same preference relation over  $Y$  (but their choice sets are not identical). Given that  $Y$  is two dimensional, we can assume that  $\Lambda = (0, \infty)$  and that  $T(U, \lambda)$  transforms  $U(c, d) = v_1(c) + v_2(d)$  into  $U_\lambda(c, d) = \lambda v_1(c) + v_2(d)$ . Thus,  $\lambda$  captures the degree of sensitivity to pride and shame.

We now show that there is a unique equilibrium in which the relatively poor agents complement their initial wealth by taking from the public fund up to the level of consumption  $\underline{w}$ ; the relatively rich contribute to the fund and are left with consumption level  $\bar{w}$ ; and the ‘middle class’ remain with their initial wealth. It will be shown that when  $r = s$ , all agents consume the same amount in the unique equilibrium.

**Claim 3.** (i) *The model has a unique equilibrium  $\langle (\hat{c}^i, \hat{d}^i), \hat{\lambda} \rangle$ . Let  $\underline{w}$  and  $\bar{w}$  satisfy  $\hat{\lambda} v_1'(\underline{w}) = s$  and  $\hat{\lambda} v_1'(\bar{w}) = r$ , respectively. Then, the equilibrium profile is:*

$$\hat{c}^i = \begin{cases} \underline{w} & \text{if } w^i \leq \underline{w} \\ w & \text{if } \underline{w} < w^i < \bar{w} \\ \bar{w} & \text{if } w^i \geq \bar{w} \end{cases}$$

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

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

**Proof:** (i) Given  $w \in (0, M]$  and  $\lambda \in \Lambda$ , the assumptions on  $v_1'$  imply that the agent’s maximization problem has a unique optimal consumption choice  $c(w, \lambda)$ . Let  $\Psi(\lambda) = \sum_i [c(w^i, \lambda) - w^i]$ , i.e. the excess demand for consumption given  $\lambda$ . Since  $c(w^i, \lambda)$  is continuous and increasing in  $\lambda$ , then so is  $\Psi(\lambda)$ . For  $\lambda$  near 0,  $\Psi(\lambda) < 0$  and for large  $\lambda$ ,  $\Psi(\lambda) > 0$ . Let  $\hat{\lambda}$  be the unique value such that  $\Psi(\hat{\lambda}) = 0$ . The assumptions on  $v_1$  guarantee the existence of unique values of  $\underline{w}$  and  $\bar{w}$  satisfying  $\hat{\lambda} v_1'(\underline{w}) = s$  and  $\hat{\lambda} v_1'(\bar{w}) = r$ , respectively. Then,  $\langle (\hat{c}^w, \hat{d}^w), \hat{\lambda} \rangle$  is an equilibrium since  $\hat{c}^w = c(w, \hat{\lambda})$  and  $\hat{d}^w = w - \hat{c}^w$ .

(ii) Consider an equilibrium  $\langle (c^w, d^w), \lambda \rangle$  with  $\lambda > 1$ . Given the fundamental preferences, each agent  $w$  weakly prefers to consume less than  $c^w$  and depending on  $w$  prefers to withdraw less or donate more. Therefore, a Pareto-superior allocation would have to weakly lower everyone’s consumption. However, this contradicts feasibility which requires the exact balancing of the public fund (and precludes disposal). An analogous argument holds for the case of  $\lambda < 1$ . ■

## 4. Discussion

### 4.1. A game model

In our model, each agent optimally chooses his action while ignoring those of others. The social norms expressed by adjust to ensure that agents' choices satisfy the feasibility constraint. This approach is in the spirit of competitive equilibrium, but is also applicable in situations of strategic interaction. A game version of the model differs from the 'competitive' one in that each agent's preferences are over the profiles of actions of all agents. The set  $F$  can be viewed as a collection of acceptable profiles in the sense that a profile that lies outside  $F$  generates pressure on the preferences to adjust. The equilibrium concept is modified accordingly. An equilibrium is a pair  $(\hat{x}^i_{i \in N}, \hat{\lambda})$  such that:  $(\hat{x}^1, \dots, \hat{x}^N) \succ_{\hat{\lambda}}^i (\hat{x}^1, \dots, y^i, \dots, \hat{x}^N)$  for all  $y^i \in X^i$  and  $(\hat{x}^i) \in F$ . Thus, equilibrium in this model imposes a standard Nash equilibrium condition on the profile of actions with respect to the endogenously determined preferences and the equilibrium profile is in the set of acceptable profiles.

### 4.2. A hybrid model

A more general version of the model would allow the agents' choice sets to be endogenous. In that case, the parameter  $\lambda$  would systematically affect not only preferences but also the choice sets. For example, in an exchange economy environment,  $\lambda$  might consist of a preference bias parameter, as before, and a price vector. The preferences  $\succ_{\lambda}^i$  would be the biased preferences as before, and the set  $X^i(\lambda)$  would then be the agent's budget sets as determined by his endowment and the price vector.

### 4.3. Pre-Efficiency

For each of the three examples, we have shown that the biased equilibrium is pre-efficient. The proofs differ but all rely on the convexity of the preferences and the structure of the choice sets (linear with a common slope). The choice of the transformation function  $T$  is also critical to the result. It is not surprising that this results requires strong conditions. The analogous fundamental First Welfare Theorem of competitive equilibrium is also a knife-edge result that depends on the convexity of the preferences, the linearity and uniformity of prices, and the lack of externalities. An example of an alternative preference bias transformation that does not yield efficiency is given in section 3.2. Similar examples can easily be constructed for the other two examples as well.

### 4.4. Separability

Perhaps the most fundamental description of preferences is as a collection of local marginal rates of substitution. With this in mind, we chose to model changes in preferences as common biases of the marginal rates of substitution. To make this as clear as possible and to avoid technical issues that are less relevant for this conceptual discussion, we restrict attention to separable preferences.

To understand the sort of difficulties that might arise with non-separable preferences, recall the exchange economy example in section 3.1. The function



$T(\tilde{z}, \lambda)$  distorts  $\tilde{z}$  as represented by  $u$  into a preference relation with a utility function  $v$  such that  $\nabla v(x) = (\lambda_k \nabla u(x)_k)$  for every  $x$ . In the case of non-separable preferences and  $K = 2$ , the bias operator can be defined by this equality since for any function  $u$  and for any positive number  $\lambda$ , there is a utility function  $u_\lambda$  such that  $MRS(u_\lambda)(x) = \lambda MRS(u)(x)$  for all  $x \in X$ . However, without the separability assumption, such a  $T$  function would not be well-defined when  $K > 2$ , since given an arbitrary differentiable function  $u : R^K \rightarrow R$  and a vector  $\lambda \in R_+^K$  there might not be a function  $v$  satisfying  $\nabla v(x) = (\lambda_k \nabla u(x)_k)$  for every  $x$ .

#### 4.5. Comments on the literature

Two previous papers in which non-price adjustment achieves harmony in a society are Richter and Rubinstein (2015, 2020). In the first, a common ordering of all alternatives, interpreted as prestige, is adjusted and in equilibrium the profile of choices has the property that each agent's choice is optimal from among all alternatives that are not more prestigious than the one he chooses. In the second, it is a partition of the alternatives into permitted and forbidden actions that is adjusted until the vector of optimal choices is feasible, and any loosening of the forbidden set would result in any optimal vector of choices not being feasible. The innovation of the model presented here is the role of preferences adjustment as an endogenous equilibrating instrument.

The paper is somewhat related to several strands of the literature. The literature on endogenous evolution of preferences (e.g. Dekel *et al.* 2007; Alger and Weibull 2016) examines the dynamic evolution of preferences conducive to social interaction, whereas we examine a static equilibrium resolution of resource allocation problems. The literatures on the use of honours to incentivize agents (e.g. Benabou and Tirole 2003; Tirole 2016; Dubey and Geanakoplos 2017) or the role of status in the allocation of labour across occupations (Fershtman *et al.* 1996) feature agents who value the attainment of 'status' within their peer group. Agents' preferences over status and other goods are exogenously fixed, as in conventional economic models. It is the status associated with agents' actions that is determined within these models, by means of either the emergence of an equilibrium convention or deliberate design by a principal. In contrast, the preferences in our model adjust endogenously in a specific uniform fashion in order to attain equilibrium.

The idea of endogenous preference change has been considered previously in various areas of economics. Since the literature is vast, diverse and not closely related to our model, we will refrain from throwing in a few random references and make do with comparing, in general terms, our approach to two strands of the literature. One of them examines the causes and consequences of preference change induced by things like addiction, habits, advertising or fashion. Roughly speaking, the 'technology' that induces preference change at the individual level is taken as given (e.g. the effect of addictive substance). Another strand deals with the deliberate strategic manipulation of preferences that is inherent in the two-stage game model. The main element of these models is each player's strategic selection (in the first stage) of his own second-stage preferences with the goal of favourably influencing the equilibrium, while anticipating the

selections made by others. Neither strand of the literature is closely related to our model, in which the changes in preferences emerge in response to pressures to equilibrate a social situation, in a way analogous to the way that price adjustment harmonizes exchange in a competitive market.

## References

- Alger I. and J.W. Weibull** 2016. Evolution and Kantian morality. *Games and Economic Behavior* **98**, 56–67.
- Balasko Y.** 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 J.P.** 1986. *Macroeconomics: An Introduction to the Non-Walrasian Approach*. New York, NY: Elsevier.
- Dekel E., J. Ely and O. Yilankaya** 2007. Evolution of preferences. *Review of Economic Studies* **74**, 685–704
- Dubey P.K. and J. Geanakoplos** 2017. *Money and Status: How Best to Incentivize Work*. Cowles Foundation Discussion Paper No. 1954R2.
- Fershtman C., K.M. Murphy and Y. Weiss** 1996. Social status, education, and growth. *Journal of Political Economy* **104**, 108–132.
- Keiding H.** 1981. Existence of budget constrained Pareto efficient allocations. *Journal of Economic Theory* **24**, 393–397.
- Richter M. and A. Rubinstein** 2015. Back to fundamentals: convex geometry and economic equilibrium. *American Economic Review* **105**, 2570–2594.
- Richter M. and A. Rubinstein** 2020. The permissible and the forbidden. *Journal of Economic Theory* **188**, art. 105042.
- Shapley L. and M. Shubik** 1971. The assignment game I: The Core. *International Journal of Game Theory* **1**, 111–130.
- Tirole J.** 2016. *On some Social determinants of Behavior and Institutions*. Notes for Nemmers Short Course, Northwestern University.

**Ariel Rubinstein** is a Professor of Economics at New York University and Professor (Emeritus) of Economics at Tel Aviv University. His interests are in Economic Theory, Models of Bounded Rationality, Choice Theory, Game Theory and Experimental Economics. His views about economics are presented in his book *Economic Fables*. Homepage: <https://arielrubinstein.org>

**Asher Wolinsky** is a Professor of Economics at Northwestern University. His area is Economic Theory.