ASPIRATION TRAPS

AVIAD HEIFETZ AND ENRICO MINELLI

ABSTRACT. Fundamental choices, like location or education, affect the attitudes and beliefs with which the individual will analyze future day to day decision problems. These effects cannot be assumed to be transparent to the individual. To restore methodological discipline in the analysis of such choices, we propose a solution concept based on an idea of consistency: the individual should not regret her fundamental choices after her preferences and beliefs have adjusted thereof. We show that even single person decision problems admit multiple, Pareto-ranked solutions: the individual might be stuck in an aspiration trap.

1. INTRODUCTION

The capacity to aspire. According to the standard economic model, an individual is defined by her preferences over a relevant set of alternatives. Most generally, alternatives are stochastic streams of commodity bundles that the individual is about to consume from the present day onwards. Typically, only a subset of these alternatives are feasible for the individual. The individual is said to be rational if her preferences satisfy some consistency conditions across various choice problems that she could potentially be facing, where a choice problem is defined by the feasible set of alternatives. Individual A is poorer than individual B if the actual feasible set available to A is smaller than the one available to B.

The anthropologist Arjun Appadurai (2004) presents a strikingly different concept of an individual, with far-reaching implications for the notion of poverty. For Appadurai, wants and wishes for 'commodities' in the wide sense of the term – physical goods, as well as marriage, work, leisure, respectability, friendship, health – are only a derivative of a more fundamental level, namely the individual's *aspirations to the good life.* These aspirations are not a feature of the individual as an atomistic entity. Rather, aspirations to the good life are formed *in the*

Date: This version: March 2006.

We would like to thank Tito Cordella, Sid Gordon, Françoise Forges, Salvatore Modica, Philippe Mongin, Heracles Polemarchakis and Philippe Van Parjis for useful comments.

thick of social life, as part of a system of ideas and beliefs about life and death, the significance of material assets over social relations, the conception of social permanence, the value of peace and war, etc.

According to Appadurai, the *capacity to aspire* is not evenly distributed in society. The poor are less likely to be conscious of the *links* between their fundamental aspirations and the available commodities, due to the combination of two related reasons. The poor have a *smaller stock of meaningful experiences* relating aspirations to commodities. Moreover, they have *fewer opportunities to experience* how a choice of a commodity influences their fundamental well-being. As a result , the poor's *'navigational map'* consists of very few combinations of nodes and pathways from aspirations to commodities. This limited capacity to form conjectures and refutations about the future is a hallmark of poverty.

This view is related to a critique raised by the development economist Esther Duflo (2003) on the idea of modeling poor individuals as rational decision makers facing a harsh budget constraint. She points out how this standard economic idea is not compatible, for instance, with the lack of large-scale use of available fertilizers in Western Kenya by many farmers who had previously experimented with these fertilizers successfully but on a very small scale. These farmers, always on the verge of subsistence, do not adopt an available option that can improve their standard of living substantially. Duflo concludes that "what is needed is a theory of how poverty influences decision-making, not only by affecting the constraints, but by changing the decision-making process itself."

At first sight, it may seem that the standard economic theory does allow for such endogenous changes of tastes as a result of experience, as in Becker (1996). The crucial difference is, of course, that in this standard theory the individual is supposed to have *perfect foresight* regarding the (possibly stochastic) change of her taste that would result from her experience.[1] It is exactly this *transparency* assumption that Appadurai and Duflo challenge.

'Consistency, not 'Transparency'. In this paper we propose an approach which does not rely on transparency of future tastes when 'big' decisions are at stake, like large-scale use of fertilizers, moving from a remote farm to a big city, or taking education. If the individual

[1]This assumption is well expressed by Roger Myerson, at the very beginning of his canonical text on Game Theory: "a player in the game is *intelligent* if he knows everything that we know about the game and he can make any inference about the situation that we can make "(Myerson (1991), p. 4).

cannot form beliefs on a set of fully specified contingencies, including her own future tastes given the 'big' choice, how is she to decide?

The contribution we offer here is an answer to this question in the form of a notion of *consistency*. We say that a 'big' decision is *self-justifying* if the individual does not regret it ex-post, once the decision is taken, and the aspirations and preferences for the consequential daily problems have adjusted accordingly.

This idea can be interpreted as capturing a basic feature of the socialization process. Social norms regarding appropriate major choices are internalized by the young without a full-fledged forward-looking analysis. If a decision is self-justifying then, when old, the individual will still recommend the same choices to the next generation, even if her views and tastes have evolved and matured. Thus, only self-justifying decisions can become social norms.

We show that, even in the absence of strategic distortion, a fundamental choice problem may admit multiple Pareto-ranked self - justifying decisions. An individual might be stuck in an *aspiration trap*: she might fail to foresee that were she to alter her fundamental choice, she would both not regret it ex - post and ultimately be happier. Moreover, all self-justifying choices might be sub-optimal, due to the inability of individuals to foresee the potential evolution of their own preferences.

Preference-formation mechanisms. For the sake of concreteness, we focus in this paper on a particular class of mechanisms by which preferences evolve as a consequence of the 'big' decision. More and more it is recognized that individual preferences over outcomes in a given situation incorporate and reflect moods, attitudes, beliefs, life views of which the individual is not fully or always conscious. The psychology literature suggests that these moods and attitudes may be purposeful in serving deeper needs and wants of the individual: they are determined unconsciously so that the conscious choice of actions will serve these deeper needs.

Recently, several papers have integrated this insight into economic models. Bewley (1999) proposes a model in which the effective effort of an employee at work is a function of a conscious decision and of an unconsciously determined mood. The conscious decision aims at the optimal trade-off between effort and wage, taking the mood as given. The unconscious side of the person dictates the mood so as to maximize an overall utility, taking into account additional factors like fairness concerns, loyalty to the employer etc. Brunnermeier and Parker (2005) study dynamic choice under uncertainty. They put forward the idea that the optimal beliefs of an individual need not be the objectively correct ones, but rather those which, taking into account

the induced decisions, maximize the expected happiness of the individual given these beliefs. We can think of the beliefs as being chosen by the unconscious side of the individual so as to optimally balance the higher utility coming from an optimistic view of the situation and the cost of the wrong choices this might induce. Similar ideas on belief - dependent utility have been discussed by Geanakoplos, Pearce and Stacchetti (1989), Rabin (1993), Yariv (2001) and Bracha (2004).

All these approaches share the feature that attitudes are determined at equilibrium, in conjunction with the conscious decisions, so as to maximize some unconscious deeper needs of the individuals. This idea underlies the following model in section 2, which we then substantiate in section 3 with an example featuring an aspiration trap. Section 4 concludes. The appendix extends the model from the single-person setting to a game-theoretic one[2].

2. Model

In period $t = 0$ an individual has to make some fundamental choice, $e \in E$, at a cost $c(e)$. For a given choice e, the game of life, to be played in $t = 1$, is described by a tuple $G^e = (X, u^e, B, \beta)$, where the strategy set of the individual is X, and her payoff function is $u^e : X \times B \to \mathbb{R}$. Differently from a standard decision theoretic problem, the utility of the individual depends also on her attitude (beliefs, aspirations) $b \in B$. When choosing a strategy $x(e, b)$ in the game of life so as to maximize u^e, the individual takes as given her attitude $b \in B$ (and, of course, also the fundamental choice e she has already taken at $t = 0$). In reality, given e the attitude b is determined, jointly with $x(e, b)$, by some preference formation mechanism $\beta : E \to B$, as follows:

Definition. *For a given e, a* solution of the game of life G^e *is a tuple* $(x(e, b), b)$ *(denoted* $(x(e), b)$ *for short) satisfying:*
1. $x(e, b)$ *maximizes* $u^e(\cdot, b) : X \to \mathbb{R}$,
2. $b = \beta(e)$.

Under the 'transparency' assumption, the individual would 'see through' the preference-formation mechanism. At $t = 0$ she would then choose e so as to maximize

$$u_i^{\cdot}(x(\cdot), \beta(\cdot)), \beta(\cdot)) - c(\cdot) \tag{2.1}$$

anticipating both the implied attitude $\beta(e)$ and the strategy $x(e, \beta(e))$. As explained in the introduction, we propose instead to identify fundamental choices which satisfy a notion of *consistency*:

[2]This extension relies on the growing literature on endogenous preferences in strategic settings (see e.g. the overview by Samuelson (2001)).

Definition. *A fundamental choice $e \in E$ is* self - justifying *if, at a solution $(x(e), \beta(e))$ of the game of life G^e, the individual does not regret her fundamental choice:*

$$u^e(x(e, \beta(e)), \beta(e)) - c(e) \geq u^{e'}(x(e', \beta(e)), \beta(e)) - c(e')$$

for all $e' \in E$.

When considering in retrospect a deviation e', the individual understands correctly the effect it would have on her utility function $u^{e'}$, but she takes as given her already-acquired attitude $\beta(e)$. That's why, typically, a self-justifying e does not maximize (2.1).

We can identify conditions that guarantee the existence of a self-justifying choice.

A1. The sets E, B and X are nonempty compact convex subsets of a Euclidian space
A2. For all $e \in E$, $b \in B$ the function $u^e(x, b)$ is strictly quasi-concave and continuous in x, and for all $b \in B$ the indirect utility $u^e(x(e, b), b) - c(e)$ is continuous and quasi-concave in e.
A3. $\beta : E \to B$ is continuous.

The following proposition is a standard application of Kakutani's fixed point theorem.

Proposition. *Under assumptions A1, A2 and A3, self-justifying choices exist.*

Proof: Let $\eta : B \to E$ be defined as

$$\eta(b) = argmax_{e \in E} u^e(x(e, b), b) - c(e)$$

Under $A1$ and $A2$, η is non empty and convex valued, and it has a closed graph. Under $A1, A2, A3$, the product correspondence $\eta \times \beta$ from $E \times B$ to itself satisfies the condition of Kakutani's fixed point theorem. At a fixed point (e, b), e is a self-justifying choice for the solution $(x(e, b), b)$ of the game of life. □

3. Example

The game of life involves the choice of a production strategy under uncertainty. There are two possible states of nature, $s = \{good, bad\}$, one of which will materialize after the individual has chosen her strategy. An individual has one unit of time to allocate between two activities, one safe and one risky. The safe activity yields one unit of consumption independently of the realized state. The risky activity yields two units of consumption if the state is $s = good$, and $e < 1$ units of consumption if the state is $s = bad$.

We could interpret this as a very stylized representation of the following situation. The first activity represents work in the traditional sector, which guarantees a safe return of one unit. The second activity offers a prospect of a higher return if things go well, but is subject to risk. In the game of life the individual has to choose how to allocate her time between the two activities. When things do not go well, the payoff of the risky activity is indexed by e, a variable which the individual has to choose before she enters the game of life. We may interpret it as the level of education, modeling the fact that people with better education have higher reservation utility if things go wrong in the risky activity.

For a given belief $b \in [0,1]$ over the occurrence of the good state $s = \textit{good}$ and a given utility for consumption $v(\cdot)$, the individual chooses the share of time to spend in the risky activity, $x \in [0,1]$, to maximize

$$u^e(x,b) = bv(2x + (1-x)) + (1-b)v(ex + (1-x)).$$

We denote her optimal choice for given values of e and b by $x(e,b)$.

We now have to specify the preference-formation mechanism. In this example, we adopt the one proposed by Brunnermeier and Parker (2005). Their starting point is the observation that beliefs affect the expected utility of the individual in two ways: indirectly, by influencing her choices, and directly, as weights on possible future contingencies. For example, an optimistic belief may lead to a biased choice, but its direct effect on the individual perception is an increase in well-being. Brunnermeier and Parker define optimal beliefs as those that would be unconsciously chosen to maximize the expected sum of utility as perceived today and tomorrow (see below for a formal definition in the context of our example). They interpret the mechanism as the result of social forces, like the fact that parents induce children to have a positive view of the world, or that happier individuals tend to be healthier.

In our example, if $\pi \in [0,1]$ is the objective probability of $s = \textit{good}$, the optimal belief b maximizes

$$\begin{aligned}&[bv(2x(e,b) + (1 - x(e,b))) + (1-b)v(ex(e,b) + (1 - x(e,b)))]\\ +&[\pi v(2x(e,b) + (1 - x(e,b))) + (1-\pi)v(ex(e,b) + (1 - x(e,b)))].\end{aligned}$$

Optimal beliefs for a given level of e are denoted by $\beta(e)$. The preceding expression can be interpreted as the expectation (using the objective probability) of the sum of two terms: the expected utility as perceived by the individual at the beginning of the game of life, when she chooses x, and her actual expected utility over the possible states of nature. The definition of optimal beliefs takes into account their influence on the choice of x, and therefore their consequences in terms of actual utility in each state of nature, but also their direct effect on the individual's perception of the situation.

Before the game of life starts, the individual chooses a level of education e, at a cost $c(e)$. She is fully aware of the nature of the game of life. In particular she knows how her choice of e today will affect the payoff of the risky activity, and she correctly anticipates her own optimal strategy $x(e,b)$ tomorrow for a given belief b. But she is not aware of the way in which her own beliefs will evolve.

Her choice of education $e \in E$ is self - justifying if, for all e',

$$\begin{gathered}\beta(e)v[2x(e,\beta(e))+(1-x(e,\beta(e)))]+\\(1-\beta(e))v[ex(e,\beta(e))+(1-x(e,\beta(e)))]-c(e)\geq\\\beta(e)v[2x(e',\beta(e))+(1-x(e',\beta(e)))]+\\(1-\beta(e))v(ex(e',\beta(e))+(1-x(e',\beta(e)))]-c(e').\end{gathered}$$

That is, when contemplating in retrospect an alternative fundamental choice e', the individual maintains her already-acquired belief $\beta(e)$.

To fix ideas, let $v(\cdot)=log(\cdot)$, $\pi=0.5$, $E=\{0,0.1,0.33\}$, $c(0)=0$, $c(0.1)=0.01$, $c(0.33)=0.11$. Optimal beliefs turn out to be $b(0)=0.5$, $b(0.1)=0.7$, $b(0.33)=0.78$, with associated optimal choices of strategy $x(0,b(0))=0$, $x(0.1,b(0.1))=0.47$, $x(0.33,b(0.33))=0.95$.

One can check that both $e=0$ and $e=0.1$ are self - justifying choices. If she chooses to take some education, $e=0.1$, the individual ends up investing more in the risky activity, adopting more optimistic beliefs, and being happier: the utilities are $u^0(x(0,b(0)),b(0))-c(0)=0$ and $u^{0.1}(x(0.1,b(0.1)),b(0.1))-c(0.1)=0.09$, respectively.

The choice $e=0$ is thus an *aspiration trap.* In a society in which $e=0$ is the established norm, a social planner who would force or tempt the individuals into choosing $e=0.1$ would be opposed by the old (with the already acquired belief $b(0)=0.5$), but would eventually be thanked by the young when they mature and acquire the belief $b(0.1)=0.7$. The choice $e=0.1$ would then be sustained as self-justifying, with no need for further policy intervention. With $e=0.1$, individuals in this society would be mildly optimistic, invest to some extent also in the risky activity, and would be happier than under the previous regime of no-education $e=0$ and its consequential despairing realism.

In fact, an individual who would have taken an even higher level of education, $e=0.33$, would have invested even more in the risky activity, $x(0.33,b(0.33))=0.95$, become even more optimistic, $b(0.33)=0.78$, and happier, $u^{0.33}(x(0.33,b(0.33)),\,b(0.33))-c(0.33)=0.19$. The choice $e=0.33$ is not a self - justifying choice, though: being so optimistic the individual would regret, ex post, to have spent so much on education.

That's why even an omniscient planner who cares about happiness cannot design a policy which would eventually make $e = 0.33$ the established norm. If $e = 0.33$ were continuously forced, individuals would live under the impression that such an intense level of education is exaggerated, and that they would be happier with less education. The impression is false, but this cannot be comprehended by the individuals, who cannot actually imagine their conception of the world and the way they would live their life had they been less educated.

4. Discussion and Conclusion

This simple example manifests the stark difference in conclusions when we replace 'transparency' by 'consistency'. Under the standard assumption of transparency, an uneducated person should be eager to study once supplied with the information about the return to education. If the facilities to study are available, the social planner need only worry about the adequate dissemination of this information, and in many societies the relevant information is indeed, by and large, freely available in the media. Under this paradigm, an individual who fails nevertheless to pursue education simply manifests her revealed preference, i.e. that her idiosyncratic cost of studying turns to be abnormally high. Such a person can be pitied, but cannot be helped by forcing her to take the available education.

In contrast, revealed preference might be a poor indication for optimal behavior under the alternative approach that we propose here. This approach admits that individuals not only learn new facts over time, but are *inherently* influenced by their life experience, by their encounter with new people and new ideas, in ways they cannot foresee in advance. Interpersonal influence – by peers and leaders – is a genuine part of the socialization process. It is far from limited to the dissemination of information. It can sometimes be beneficial and allow for pulling individuals out of aspiration traps. But unlike in the classical paradigm, even in the best of all worlds it need not lead to sustainable optimal choices.

Embeddedness. Abstractly, a self-justifying choice is a "best reply" to one's attitude, which has itself been optimally shaped *given* the self-justifying choice that the individual made when she was still young and lacking a clear vision of the repercussions of her choice. Formally, therefore, we have here a Nash equilibrium between a mature individual and the unconscious preference-formation mechanism of the young.

This non-standard use of the Nash equilibrium notion is a particular way to model the individual as *embedded* in society, in-between the two

extreme models of the individual that can be encountered in the social sciences. Unlike in the standard Economic model, the individual is not atomistic, and her preferences are not well defined and do not form outside a social context. On the other hand, unlike in some Sociological models, the individual is not a leaf in the swaying winds of social norms and forces.

In the proposed model, the two approaches are interwoven. Regarding 'big' choices the individual is subject to the socialization process, while she is a standard economic maximizer concerning day-to-day alternatives. Embeddedness is expressed by the idea that these two dimensions are at equilibrium with one another.

Inertia revisited with a new vocabulary. Our approach suggests a new perspective on problems of inertia with respect to economic change and the way to confront it. In the classical setting, such inertia has been imputed to traditional culture, habits, beliefs. In game theoretic terms : 'Past cultural beliefs provide focal points and coordinate expectations, thereby influencing equilibrium selection' (Greif (1994), p. 915). A related idea is that of group identity: poor people, or more generally individuals belonging to a socially marginal group, would pay a psychological cost in terms of identity loss if they were to adopt the pattern of behavior of the rich/dominant group (Akerlof and Kranton (2000)).

Both ideas recognize that individual preferences are influenced by social factors, and treat these social factors as given from the point of view of the individual, grounded in past aggregate experience. We do subscribe to the same premises. However, unlike in the two models quoted above, individuals stuck in an aspiration trap are not powerless: a different choice may put them on a different path, in which their aspirations change, thereby engendering a self-sustaining switch in their choice. This might involve considerable individual pain, but need not require any major (and therefore discouragingly difficult) *coordination* effort.

Let us try to clarify this point by means of an example. Children of poor families are less likely to take higher education than children of richer families. Models invoking 'culture', or 'identity' explain this observation by arguing that if a child of a poor family takes high education she pays a cost in terms of dissonance or lost identity. We think a more basic explanation has to do with the fact that the meaning of the sentence 'taking higher education', and the impact of the corresponding choice on future outcomes are hard to evaluate from the point of view of someone whose closer relatives and friends never experienced anything similar. This lack of capacity to aspire is the burden that the individual inherits from her origins. But if she were to try

a new course of action, her preferences would evolve, and sustain her choice. In this interpretation, the role of the social group is not that of a punishing device for those who put into question a static and immutable group identity or an established social equilibrium. Its role is more that of a source of a meaningful stock of experiences, which may influence choices and thereby the formation of one's own preferences. Only direct individual experience has the power to induce a change in the capacity to aspire. Change does not have to wait for aggregate transformations in culture or identity, but it will not come about spontaneously: individuals might remain stuck in an aspiration trap.

5. Appendix: Extension to Strategic Interaction

5.1. **Model.** Individuals are $i = 1, 2, \dots I$. In period $t = 0$ each individual i has to make some fundamental choice, $e_i \in E_i$, at a cost $c_i(e_i)$. For a given profile $e \in E = \times_i E_i$, the game of life, to be played in $t = 1$, is a tuple $G^e = ((X_i, u_i^e, B_i,)_{i \in I}, \beta)$. The strategy set of each individual is X_i, her set of potential attitudes is B_i, and her payoff function is $u_i^e : X \times B_i \to \mathbb{R}$, where $X = \times_i X_i$. When choosing strategies in the game of life, individuals take attitudes as given. Attitudes are actually determined, jointly with individual strategies, by a preference formation mechanism $\beta : E \to B$, where $B = \times_i B_i$. Definition 1 thus generalizes to:

Definition. *For a given e, a* solution of the game of life G^e *is a tuple $(x(e,b), b)$ (denoted $(x(e), b)$ for short) satisfying*
1. $x(e,b)$ is a Nash equilibrium of the game with payoff functions $u_i^e(\cdot, b) : X \to \mathbb{R}$,
2. $b = \beta(e)$.

Under the classical 'transparency' assumption, we could find such a solution by backward induction. At $t = 0$ individual i would choose e_i so as to maximize

$$u_i^{(\cdot, e_{-i})}(x(\cdot, e_{-i}), \beta_i(\cdot, e_{-i})), \beta_i(\cdot, e_{-i})) - c_i(\cdot)$$

given the equilibrium choices e_{-i} of the other individuals. In contrast, our alternative notion of 'consistency' in definition 2 generalizes as follows:

Definition. *A profile of fundamental choices $e \in \times_i E_i$ is* self - justifying *if, at a solution $(x(e), \beta(e))$ of the game of life G_e, no individual regrets her fundamental choice:*

$$u_i^e(x(e, \beta(e)), \beta_i(e)) - c_i(e_i) \geq u_i^{(e_i', e_{-i})}(x(e_i', e_{-i}, \beta(e)), \beta_i(e)) - c_i(e_i')$$

for all $e_i' \in E_i$, and all $i \in I$.

When considering a deviation e_i', the individual understands correctly its effect on the structure of the game of life, and on the equilibrium behavior there, but she takes as given the other individuals' fundamental choices *and* the attitudes $\beta(e)$.

5.2. **Examples.** The game of life consists of pairwise interactions. Individuals randomly meet to play in pairs (i,j) a symmetric game with material payoffs Π indexed by the levels of fundamental choices ('education') $(e_i,e_j)\in E\subset\mathbb{R}$, to be decided beforehand:

$$\Pi(x_i,x_j;e_i).$$

Individuals maximize perceived payoffs

$$u_i^{e_i}(x_i,x_j,b_i)$$

where $b_i\in\mathbb{R}$ is an unconscious attitude, and $u_i^{e_i}(x_i,x_j,0)=\Pi(x_i,x_j;e_i)$ (with no bias, $b_i=0$, the individual's perceived payoff coincides with her material payoff). For given levels of education and attitudes, a Nash equilibrium of the game defined by $\left(u_i^{e_i}(\cdot,\cdot,b_i),u_j^{e_j}(\cdot,\cdot,b_j)\right)$ is

$$x_i(e_i,e_j,b_i,b_j),\ x_j(e_i,e_j,b_i,b_j). \tag{A.1}$$

If there are multiple equilibria, we assume that across the pairwise interactions with parameters (e_i,e_j,b_i,b_j), these equilibria are played according to a distribution that assigns positive probabilities to all of them. Equilibrium attitudes $\beta\left(e_i,e_j\right)=\left(b_i^*\left(e_i,e_j\right),b_j^*\left(e_i,e_j\right)\right)$ are such that

$$b_i^*\left(e_i,e_j\right)\in\arg\max_{b_i}E(\Pi_i(x_i(e_i,e_j,b_i,b_j^*\left(e_i,e_j\right)),x_j(e_i,e_j,b_i,b_j^*\left(e_i,e_j\right));e_i))$$

where the expectation is taken over the distribution of play of Nash equilibria (A.1) with opponents that have chosen e_j. We note that $\beta\left(e_i,e_j\right)$ need not be unique, i.e. β may be a correspondence.

Before the game of life starts, each individual i chooses a level of e_i at a cost $c(e_i)$. Individuals understand the effects of the choice of e_i on payoffs Π, but are not aware of the possible effect on the equilibrium attitudes, since they are unable to go through the mental exercise regarding the future formation of their attitudes. If ε is the level of education chosen by everybody, and $\beta(\varepsilon,\varepsilon)=(b^*,b^*)$ are the associated attitudes, an individual i believes that if she were to choose e_i' while everybody else still chose ε, the corresponding equilibrium strategies in an interaction in which she is assigned role i would be

$$x_i(e_i',\varepsilon,\beta(\varepsilon,\varepsilon)),\ x_j(e_i',\varepsilon,\beta(\varepsilon,\varepsilon)).$$

Given that everybody else is choosing ε, a level of education ε is self-justifying if

$$u_i^{\varepsilon}(x_i(\varepsilon,\varepsilon,b^*,b^*),x_j(\varepsilon,\varepsilon,b^*,b^*),b^*)\geq u_i^{e_i'}(x_i(e_i',\varepsilon,b^*,b^*),x_j(e_i',\varepsilon,b^*,b^*),b^*)$$

for $i = 1, 2$.

To fix ideas, let the game of life be a game with material payoffs as follows:

	C	D
C	$2 + e_1, 2 + e_2$	$0, 5$
D	$5, 0$	$3, 3$

fundamental choices, $e_i \in \{0, 2\}$ (no-education, education) affect the material payoff of cooperation, (C, C).

Education is chosen before the game of life starts, and it has a cost $c_i(0) = 0, \quad c_i(2) = \delta > 0$, small.

In any given pairwise interaction, individuals maximize perceived payoffs

	C	D
C	$2 + e_1 + b_1, 2 + e_2 + b_2$	$b_1, 5$
D	$5, b_2$	$3, 3$

where $b_i \in \{0, 2\}$ (materialistic, pro-social) is an unconscious attitude. Consider first a scenario where the population playing the game of life is composed of educated individuals. With $e_1 = e_2 = 2$, material payoffs are

	C	D
C	$4, 4$	$0, 5$
D	$5, 0$	$3, 3$

This is a prisoners dilemma: (D, D) is the equilibrium among materialistic players.

On the other hand, when a pro-social row player meets a materialistic column player, their utilities are

	C	D
C	$4 + 2, 4$	$0 + 2, 5$
D	$5, 0$	$3, 3$

Again, the unique equilibrium is (D, D), and the pro-social gets the same material payoff, 3, as does the materialistic.

Things change when two pro-social players meet. In this case their utilities are

	C	D
C	$4 + 2, 4 + 2$	$0 + 2, 5$
D	$5, 0 + 2$	$3, 3$

The game has two equilibria – (C, C) and (D, D). If at some of these meetings the efficient equilibrium (C, C) is played, the pro-social attitude is dominant in the attitude - game and therefore it is the unique

equilibrium in that game. The expected material payoff in the game of life among educated and pro-social players is therefore larger than 3.

We conclude that in a population of educated and hence pro-social players, education is self-justifying. Indeed, fix pro-social attitudes, and consider a player contemplating the consequences of not taking education. Her payoff as a column player meeting an educated row player would be:

	C	D
C	$4+2, 2+2$	$0+2, 5$
D	$5, 0+2$	$3, 3$

D would be a dominant strategy for her, leading to a payoff of 3 in the unique Nash equilibrium of this game, smaller than her payoff with education.

We now show that no-education is also a self-justifying choice. With $e_1 = e_2 = 0$, material payoffs are

	C	D
C	$2, 2$	$0, 5$
D	$5, 0$	$3, 3$

and utilities are

	C	D
C	$2+b_1, 2+b_2$	$0+b_1, 5$
D	$5, 0+b_2$	$3, 3$

Whatever $b_1, b_2 \in \{0, 2\}$ are, D is a dominant strategy. Pro-social and materialistic attitudes lead to the same material payoff, and any (mixture) of the two could be the outcome of the attitude selection game.

Regardless of the mixture of attitudes prevailing in the population, $e_i = 0$ by all individuals i is a self-justifying choice. Indeed, consider a player contemplating the consequences of taking education. Her payoff as a column player meeting an uneducated row player would be:

	C	D
C	$2+b_1, 2+2+b_2$	$0+b_1, 5$
D	$5, 2+b_2$	$3, 3$

Now the *row* player has a dominant strategy, D, whatever his attitude, and the equilibrium outcome is again (D, D). Education gives no advantage: it is not worth its cost.

Thus, the choice of no-education by everybody is an *aspiration trap*. Notice that in some sense, this trap is even more severe than the trap of a Pareto-inferior Nash equilibrium of a standard coordination game.

Here, when everybody is materialistic, low-cost education seems unworthy to each individual *even under a hypothetical scenario by which everybody would have chosen education.* Even a *joint* move from no-education to education cannot be perceived as beneficial when individuals are unconscious of the attitude-formation mechanism.

In this example, $e_i = 0$ for all the individuals i, as well as $e_i = 2$ for all i, are both sub-game perfect equilibria under the classical assumption that the mechanism of attitude formation is transparent to the individuals. Indeed, in this classical setting, the choice of education at $t = 0$ simply boils down to a coordination game with multiple, Pareto-ranked equilibria. However, it is easy to find other examples in which there is a unique self-justifying choice, which would not be a sub-game perfect equilibrium if individuals could see through the mechanism of attitude formation. For instance, let the game of life be a game with strategic complementarities:

$$\Pi_i(x_i, x_j; e_i) = (1 + e_i + \frac{1}{2}x_j - x_i)x_i.$$

with the cost of the 'big choice' e_i being

$$c\,(e_i) = \frac{1}{2}e_i^2.$$

For given investment choices e_i and attitudes b_i, the utilities in the game of life have the form

$$u_i^e(x_i, x_j, b_i) = (1 + e_i + b_i + \frac{1}{2}x_j - x_i)x_i$$

The unique self-justifying pair of investment choices turns out to be $e_i = 3.46$ for all the individuals. However, if individuals were aware of the effect of investment on the formation of attitudes in the game of life, a higher level of investment would result. If everybody were choosing $e_j = 3.46$, each individual i, if she could anticipate correctly the working of the preference-formation mechanism, would rather choose $e_i = 4.24$.

References

[1] Akerlof G. and R. Kranton (2000) "Economics and Identity", *The Quarterly Journal of Economics*, **115**, 715 - 753.

[2] Appadurai A. (2004) "The Capacity to Aspire: Culture and the Terms of Recognition", in Rao V. and M. Walton (eds.) *Culture and Public Action*, Stanford University Press, http://www.cultureandpublicaction.org

[3] Becker, G., (1996), *Accounting for tastes* , Harvard University Press, Cambridge, MA.

[4] Bewley T. (1999) "Work motivation", Cowles Foundation DP 1209, Yale University.

[5] Bracha, A. (2004), Affective Decision Making in Insurance Markets, Yale ICF Working Paper No. 04-03.

[6] Brunnermeier M. and J. Parker (2005) "Optimal Expectations", *American Economic Review* **95(4)**, ? . Available at http://www.wws.princeton.edu/ econdp/
[7] Duflo, E. (2003), "Poor but Rational?", mimeo, MIT.
[8] Geanakoplos J., D. Pearce and E. Stacchetti (1989) "Psychological Games and Sequential Rationality." *Games and Economic Behavior*, **1**, 60 - 79.
[9] Greif A. (1994) "Cultural Beliefs and the Organization of Society ", *The Journal of Political Economy*, **102**, 912 - 949.
[10] Myerson R. (1991) *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge MA.
[11] Rabin M. (1993) "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, **83**, 1281-1302.
[12] Ray D. (2003) "Aspirations, Poverty and Economic Change," mimeo, New York University. Available at http://www.econ.nyu.edu/user/debraj/
[13] Samuelson, L. (2001), "Introduction to the Evolution of Preferences", *Journal of Economic Theory* **97**, pp. 225-230.
[14] Yariv L. (2001) "Believe and Let Believe: Axiomatic Foundations for Belief Dependent Utility Functionals", mimeo, UCLA

Aviad Heifetz: The Department of Economics and Management, The Open University of Israel.

E-mail address: avaidhe@openu.ac.il

Enrico Minelli: Economics Department, University of Brescia, Italy.

E-mail address: minelli@eco.unibs.it