Contributions

Aviad Heifetz* and Enrico Minelli

Aspiration Traps

Abstract: Fundamental non-recurrent choices, like location or education, affect the attitudes and beliefs with which the individual consequently analyzes 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 self-justification and *consistency*: the individual should not regret her fundamental choice after her preferences and beliefs regarding day-to-day decisions have adjusted thereof. We show that even single-person fundamental choice problems admit multiple, Pareto-ranked solutions: the individual might be stuck in an aspiration trap.

Keywords: aspiration, fundamental choice
JEL Classifications: D03, D83

DOI 10.1515/bejte-2013-0120

1 Introduction

Classical economists were well aware that participation in economic activity influences the development of the individuals. Discussions of how social interactions affect individual preferences can be found in Smith, Marx and other classical writers (see, e.g., the brief survey in Becker 1996, ch. 1). This point of view remains central in sociological thought (Bourdieu 2003) and is well represented in other social sciences, like anthropology (Douglas and Ney 1998) or law (Sunstein 1997).

In economics it has been overshadowed by the acceptance of the rational choice paradigm, based on the analytical separation of preferences and

***Corresponding author: Aviad Heifetz,** Department of Economics and Management, The Open University of Israel, Ra'anana, Israel, E-mail: aviadhe@openu.ac.il
Enrico Minelli, Department of Economics and Management, University of Brescia, Brescia, Italy, E-mail: enrico.minelli@unibs.it

opportunities: the individual is supposed to have clear preferences on a choice space *before* being faced with the problem of choosing from a particular set of available opportunities. Rationality is then defined as a consistency condition across different problems.

Thanks to its very abstract nature, this paradigm encompasses situations that could, at first sight, be interpreted as involving endogenous changes in tastes. An important example is choice under uncertainty: the model can easily incorporate learning over time as a function of previous actions, in particular learning over one's own preferences. Another example is the approach adopted by Gary Becker in many writings (see, e.g., the collection of essays in Becker (1996)). He admits that preferences over economic goods are influenced by personal experience, or even by social parameters like peers' histories or the structure of social networks, but he considers the individual as maximizing an overall utility, taking into account the effects of individual choices on all parameters.

In both of these applications of the rational choice paradigm, whatever effects personal choices or social interaction may have on one's preferences, these effects are fully understood by the individual before she has to take any action.

The crucial feature of the rational choice paradigm is not that the effects of personal or social history on preferences cannot be discussed, but rather that they are discussed under a very specific methodological premise: whatever is put in the model is also perfectly transparent to the agents in the model.[1]

As already noted by Koopmans (1964), this assumption of "transparency" is particularly strained when we want to analyze "big" choices, major decisions in life – like the choice of education or location – which have far-reaching consequences for the daily choice problems that the individual is about to face henceforth: "the individual choosing an education for a profession does not spell out sequences of commodity bundles he expects to consume for each choice of profession." One reason for this is that "Tastes evolve with experience. A model that freezes preferences by the adoption at an initial point in time of an ordering of programs for a future period of indefinite duration is likely to become an unacceptable straight jacket as time proceeds."

When trying to develop an analysis of this type of "big," non-recurrent choices, like taking education, or moving from a remote farm to a big city, or

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, 4).

adopting large-scale use of fertilizers[2] we are confronted with a particularly thorny problem: in the absence of perfect understanding of the implications of her decisions on the future adaptation of her tastes and views, how is the individual to decide?

If we abandon the standard Bayesian approach, which assumes that the individual can form beliefs on a set of fully specified contingencies, including his own future tastes, how are we to impose some discipline on our modeling exercise?

The contribution we offer in this paper 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.

In Section 2 we model the idea of self-justifying choices in the context of an individual decision problem, and we discuss an example of an aspiration trap. In Section 3 we extend the model from the single-person setting to a game-theoretic one,[3] and again we illustrate our ideas in the context of a specific

2 We refer here to the important work by the development economist Esther Duflo (2006). She points out how the standard economic idea of poor individuals as rational decision makers facing a harsh budget constraint is not compatible 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."

3 This extension relies on the literature on endogenous preferences in strategic settings (see, e.g., the overview by Samuelson (2001), Heifetz, Shannon, and Spiegel (2007a, 2007b)).

example. Section 4 concludes the paper by relating our paper to the notions of "the capacity to aspire" discussed by the anthropologist Appadurai (2004).

2 Individual choice

2.1 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 1 *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

$$u_i^{(\cdot)}(x(\cdot, \beta(\cdot)), \beta(\cdot)) - c(\cdot), \qquad [1]$$

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*:

Definition 2 *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 $\beta(e)$.[4] That's why, typically, a self-justifying e does not maximize eq. [1].

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

A1. The sets E, B and X are non-empty 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 3 *Under assumptions A1, A2 and A3, self-justifying choices exist.*

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

$$\eta(b) = \text{argmax}_{e \in E} u^e(x(e, b), b) - c(e).$$

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.[5] □

Remark 1. Abstractly, a self-justifying choice is a "best reply" to one's attitude, which is itself optimally shaped by the preference formation mechanism.

Related notions of equilibrium in decision problems or games have been discussed by Geanakoplos, Pearce, and Stacchetti (1989), Rabin (1993), Yariv (2005), Dalton and Ghosal (2012) and Bracha and Brown (2012).

---

4 One possible interpretation of β may be as the evolutionary viable "genotype," which for any given fundamental choice e gives rise to the phenotypic expression $b = \beta(e)$. However, this interpretation of β need not be constrained to a physical, biological reading, as in genoeconomics, because the preference formation mechanism β may evolve under cultural pressures and not merely under physical ones.

5 Notice that despite of the continuity of β, the eventual day-to-day decision $x(e, \beta(e))$ may be discontinuous in e. This is compatible with the observation that at times, small changes in initial conditions may bring about abrupt changes in individual (and collective) behavior.

In models with belief-dependent utilities, however, the "preference formation mechanism" implies that beliefs should be correct at equilibrium.

Here we allow for a rather general interpretation of the "preference formation mechanism" and of the notion of consistency (in the example of this section b is indeed a belief, but it need not be correct at equilibrium; in the example in the next section b represents more generally a "preference type," and the preference formation mechanism reflects evolutionary stability in a strategic context).

For us, the distinction between "big" choices and the day-to-day choices in the game of life is crucial. The standard argument to justify the assumption of perfect foresight or rational expectations refers to the repetition of similar choices in a stable environment, and this assumption is untenable for self-transforming "big" choices. One possible interpretation is that of a self-justifying choice as a non-recurrent "big" choice that the individual made when she was still young and lacking a clear vision of the repercussions of her choice.

Formally, we have a Nash equilibrium between a mature individual and the unconscious preference-formation mechanism of the young, but the use of the Nash equilibrium notion is just 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, the individual is not a "cultural dupe," 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.

2.2 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, capturing the idea that people with better education may 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 $s = good$ 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 $s = good$, 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',

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

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.

Remark 2. For the sake of concreteness, we focused in this example on a particular preference formation mechanism proposed by Brunnermeier and Parker (2005). In our application this mechanism has the implication that individuals with better education end up being too optimistic, that is, that education reduces realism but enhances exploration and risk tolerance. One could refer here to the phenomenon of "depressed realism," that is, to the phenomenon of people in difficult situations becoming more focused on the concrete possibilites available to them (see, e.g., Svenson, 1981; Taylor and Brown, 1988),[6] but we do not want to insist on any particular interpretation of this example, which is meant only as an illustration of a more general idea. Independently from the specific form of the Brunnermeier and Parker mechanism, what is relevant for our discussion is the fact that 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. 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, and several papers have integrated this insight into economic models. We just mention the seminal contribution by Bewley (1999), who 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.

Whatever the particular form of the preference formation mechanism, if the working of this mechanism cannot be assumed to be transparent to the individual our equilibrium notion is relevant, and the type of aspiration trap illustrated in the example is a possible outcome.

---

6 We thank one referee and the editor for attracting our attention to this issue.

3 Strategic interaction

3.1 Model

Individuals are $i = 1, 2, \ldots, I$. 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 4 *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 5 *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 $\beta(e)$.

3.2 Example

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

$$x_i(e_i, e_j, b_i, b_j), x_j(e_i, e_j, b_i, b_j). \qquad [2]$$

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.[7] Equilibrium attitudes $\beta(e_i, e_j) = \left(b_i^*(e_i, e_j), b_j^*(e_i, e_j)\right)$ are such that

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

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 $\beta(e_i, e_j)$ need not be unique, that is, 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

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

7 This assumption expresses the idea that no Nash equilibrium in the eventual game (after b_i, b_j have been determined given the fundamental choices e_i, e_j) is a priori excluded.

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 all $e_i' \in E^i$, 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 $c_i(0) = 0$, $c_i(2) = \delta > 0$, 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, prosocial) 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 prisoner's dilemma: (D,D) is the equilibrium among materialistic players.

On the other hand, when a prosocial 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 prosocial gets the same material payoff, 3, as does the materialistic.

Things change when two prosocial 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 prosocial 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 prosocial players is therefore larger than 3.

We conclude that in a population of educated and hence prosocial players, education is self-justifying. Indeed, fix prosocial 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. Prosocial 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.

Remark 3. In this example, $e_i = 0$ for all the individuals i, and $e_i = 2$ 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:

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

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

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

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 $e_i = 4.24$.

4 Concluding comments

Even if abstract and simplistic, the two examples of educational choice that we discussed illustrate the stark difference in conclusions when we replace "transparency" by "consistency."

Going beyond the specific examples, we end the paper with some observations on the new perspective that our approach suggests on the issue of education policy, and, more generally, of inertia with respect to economic change and a way to address it.

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, that is, 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.

An observed feature of situations of oppression and deprivation is the lack of a "capacity to aspire" (Appadurai 2004). Oppressed people come to see their situation as inescapable, and sometimes seem to lack the motivation to adopt alternative courses of action, even when the available infrastructure renders such alternatives accessible.

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, 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. And a policy maker with the appropriate foresight can incentivize individuals to pull themselves out of an aspiration trap into another (self-justifying) fundamental choice. This interpretation of social change as the outcome of a self-transformation induced by individual and communal *experience* echoes the view so forcefully put forward by Appadurai on the basis of field experiences in India.[8]

For 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 higher education she pays a cost in terms of dissonance or lost identity. In contrast, in terms of the model proposed here, 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, reflected and multiplied by a similar lack of her peers. 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 may be able to

8 See also Ray (2006) for a different formalization of Appadurai's idea of the capacity to aspire.

free themselves from an aspiration trap only if they actually try a different course of action and live through the personal changes that this entails.

Acknowledgment: We would like to thank Tito Cordella, Sid Gordon, Françoise Forges, Salvatore Modica, Philippe Mongin, Heracles Polemarchakis and Philippe Van Parjis for useful comments. We also thank the editor and four reviewers for their careful reading and suggestions.

References

Akerlof, G., and R. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115:715–53.

Appadurai, A.. 2004. "The Capacity to Aspire: Culture and the Terms of Recognition." In *Culture and Public Action*, edited by V. Rao and M. Walton, 59–84. Redwood City, CA: Stanford University Press. http://www.cultureandpublicaction.org

Becker, G.. 1996. *Accounting for Tastes*. Cambridge, MA: Harvard University Press.

Bewley, T.. 1999. "Work Motivation." Cowles Foundation DP 1209, Yale University.

Bourdieu, P.. 2003. *The Social Structures of the Economy*. Cambridge: Polity Press.

Bracha, A., and D. Brown. 2012. "Affective Decision-Making: A Theory of Optimism Bias." *Games and Economic Behavior* 75:67–80.

Brunnermeier, M., and J. Parker. 2005. "Optimal Expectations." *American Economic Review* 95(4):1092–118.

Dalton, P., and S. Ghosal. 2012. "Decisions with Endogenous Frames." *Social Choice and Welfare* 38:585–600.

Douglas, M., and S. Ney. 1998. *Missing Persons: A Critique of Personhood in the Social Sciences*. Berkeley, CA: University of California Press.

Duflo, E.. 2006. "Poor but Rational?" In *Understanding Poverty*, edited by A. V. Banerjee, R. Benabou, and D. Mookherjee, 367–78. Oxford: Oxford University Press.

Geanakoplos, J., D. Pearce, and E. Stacchetti. 1989. "Psychological Games and Sequential Rationality." *Games and Economic Behavior* 1:60–79.

Greif, A. 1994. "Cultural Beliefs and the Organization of Society." *The Journal of Political Economy* 102:912–49.

Heifetz, A., C. Shannon, and Y. Spiegel. 2007a. "What to Maximize if You Must." *Journal of Economic Theory* 133(1):31–57.

Heifetz, A., C. Shannon, and Y. Spiegel. 2007b. "The Dynamic Evolution of Preferences." *Economic Theory* 32(2):251–86.

Koopmans, T. 1964. "On the Flexibility of Future Preferences." In *Human Judgements and Optimality*, edited by M. Shelly and G. Bryan, 243–254. New York: Wiley & Sons.

Myerson, R. 1991. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.

Rabin, M. 1993. "Incorporating Fairness Into Game Theory and Economics." *American Economic Review* 83:1281–302.

Ray, D. 2006. "Aspirations, Poverty and Economic Change." In *Understanding Poverty*, edited by A. V. Banerjee, R. Benabou, and D. Mookherjee, 409–21. Oxford: Oxford University Press.

Samuelson, L. 2001. "Introduction to the Evolution of Preferences." *Journal of Economic Theory* 97:225–30.

Sunstein, C. 1997. *Free Markets and Social Justice*. Oxford: Oxford University Press.

Svenson, O. 1981. "Are We All Less Risky Drivers and More Skillful Than Our Fellow Drivers?" *Acta Psychol*. 47(2):143–8.

Taylor, S. E., and J. D. Brown. 1988. "Illusion and Well-Being: A Social Psychological Perspective on Mental Health." *Psychological Bulletin* 103:193–210.

Yariv, L. 2005. "I'll See It when I Believe It: A Simple Model of Cognitive Consistency", mimeo, Caltech