The Endowment Effect as a blessing

Frenkel, Sivan and Heller, Yuval and Teper, Roee

Hebrew University of Jerusalem, University of Oxford, Department of Economics, Univeristy of Pittsburgh

24 August 2013

The Endowment Effect as a Blessing*

Sivan Frenkel† Yuval Heller‡ Roee Teper§

December 2, 2014

Abstract

We study the idea that seemingly unrelated behavioral biases can coevolve if they jointly compensate for the errors that any one of them would give rise to in isolation. We pay specific attention to barter trade of the kind that was common in prehistoric societies, and suggest that the "endowment effect" and the "winner's curse" could have jointly survived natural selection together. We first study a barter game with a standard payoff-monotone selection dynamic, and show that in the long run the population consists of biased individuals with two opposed biases that perfectly offset each other. In this population, all individuals play the barter game as if they were rational. Next we develop a new family of "hybrid-replicator" dynamics. We show that under such dynamics, biases are stable in the long run even if they only partially compensate for each other and despite the fact that the rational type's payoff is strictly larger than the payoffs of all other types.

Keywords: Endowment Effect, Winner's Curse, Bounded Rationality, Evolution.

JEL Classification: C73, D82, D03

1 Introduction

The growing field of Behavioral Economics has frequently identified differences between the canonical model of rational decision making and actual human behavior. These differences,

*A previous version of the paper was titled "Endowment as a Blessing". We thank Eddie Dekel, John Duffy, Alan Grafen, Richard Katzwer, Shawn McCoy, Erik Mohlin, Thomas Norman, Luca Rigotti, Larry Samuelson, Lise Vesterlund and various seminar audiences for valuable discussions and suggestions, and to Sourav Bhattacharya for the query that initiated this project.

†Faculty of Management, Tel Aviv University. frenkels@post.tau.ac.il. URL: https://sites.google.com/site/sivanfrenkel/.

‡Department of Economics and Queen's College, University of Oxford. yuval.heller@economics.ox.ac.uk. URL: https://sites.google.com/site/yuval26/.

§Department of Economics, University of Pittsburgh. rteper@pitt.edu. URL: http://www.pitt.edu/∼rteper/.

usually referred to as "anomalies" or "biases," have been identified through controlled experiments in the laboratory, as well as in field experiments (see, e.g., Kagel & Roth, 1997; Harrison & List, 2004). Such behavior is puzzling to economists, who are trained to think that competitive forces in our society and economy select optimal over suboptimal behavior. In this paper, we argue that sets of biases may persist because they jointly compensate for the errors that any one of them would give rise to in isolation. Thus, biases may coevolve as a "shortcut" solution that leads in specific important environments to behavior that is approximately optimal.[1] While the majority of the existing literature studies behavioral biases separately, our results suggest that one can gain a better understanding of different behavioral biases by analyzing their combined effects.

We specifically investigate a possible relation between the endowment effect and the winner's curse, though our framework and insights are also relevant to the study of relations between other biases. The *endowment effect* (Thaler, 1980) refers to an individual's tendency to place a higher value on a good once he owns it.[2] The *winner's curse,* or *cursedness* (Eyster & Rabin, 2005), is the failure of an agent to account for the informational content of other players' actions. Cursed agents underestimate the effect of adverse selection, and thus tend to overbid in common value auctions and bilateral-trade interactions.[3]

Model. We demonstrate that these two seemingly unrelated biases compensate for each other in barter-trade interactions such as were common among prehistoric societies.[4] In our model, each of two traders owns a different kind of indivisible good and considers whether to participate in trade or not. The value of each good depends on an unobservable property that is known to the owner of the good but not to his trading partner. The potential gain of the traders also depends on additional conditions, which are known to both players before they engage in trading but can change from one instance of trade to another. Goods are exchanged if both traders agree.

Each agent in the population is endowed with a pair of biases. The level of these biases determines his type, and the agent with the minimal level of both biases (i.e., the unbiased one) is "rational." The first bias is *cursedness*: the extent to which an agent underestimates

[1] Cesarini *et al.*, 2012 present experimental evidence suggesting that many common behavioral biases (and in particular, loss aversion, which is closely related to the endowment effect) are partially heritable.

[2] See Kahneman *et al.* (1991) for a survey on the endowment effect, and see Knetsch *et al.* (2001); Genesove & Mayer (2001); Bokhari & Geltner (2011); Apicella *et al.* (2014) for recent experimental evidence.

[3] See Kagel & Levin (2002, Chapter 1) for a survey on the winner's curse, and see Grosskopf *et al.* (2007); Massey & Thaler (2013) for recent experimental support and field evidence.

[4] Evidence from anthropology suggests that trade between groups, based on localization of natural resources and tribal specializations, was common in primitive societies (see Herskovits, 1952; Polanyi, 1957; Sahlins, 1972; Haviland *et al.*, 2007). Moreover, "[t]he literature on trade in nonliterate societies makes clear that barter is by far the most prevalent mode of exchange" (Herskovits, 1952, p. 188).

the relation between the partner's agreement to trade and the quality of this partner's good. In the trade game, agents in general choose to trade goods that are not very valuable and keep valuable goods for themselves. A cursed trader does not pay enough attention to this fact and overestimates the value of his partner's good conditional on trade (in the extreme case, a fully cursed agent simply expects to get a good of ex-ante value). As a result, a more cursed agent will agree to trade goods with a greater personal value. Thus, cursedness leads agents to trade too much, and higher cursedness results in more trade.

The second bias is a *perception bias:* a function,ψ, that distorts an agent's subjective valuation of his own good. If the good is worth x, an agent believes it to be worth $\psi(x)$. If $\psi(x) > x$, we say that the agent exhibits the endowment effect, but we allow agents to have also other perception biases as well. An agent with the endowment effect does not agree to trade goods with low values since he believes those goods to be more valuable than they actually are, and thus he loses profitable transactions. Agents with the endowment effect trade too little, and traders with a higher level of endowment effect trade less.

Results. We analyze the interaction among a large population of traders with different levels of biases (types). Agents are randomly matched and play the barter game. We assume that agents do not observe the types of their partners and in each period they best-reply to the aggregate behavior. Their best reply, however, is distorted by their own biases. We show that there is a set Γ of types that exhibit both the winner's curse and the endowment effect, such that the two biases compensate each other. The set Γ includes not only the rational unbiased type, but also types of all levels of cursedness, and types who are more cursed present a greater endowment effect.

Our first result (Proposition 1) shows that a distribution of types is a Nash equilibrium of the population game if and only if its support is a subset of Γ. Moreover, all agents in the population present the same "as-if-rational" behavior on the equilibrium path (their trading strategy is identical to that of a rational trader), and any type outside Γ achieves a strictly lower payoff if it invades the population. In a dynamic setting where the payoff of the barter game determine the agents' fitness and the frequency of types evolves according to a payoff-monotone selection (e.g., the replicator dynamic; see Taylor & Jonker, 1978), stable distributions of types are exactly those with a support in Γ.

We then extend our analysis to the case where fitness is determined not only by the outcome of barter trade, but is also a result of other activities, in which the biases typically do not compensate for each other. We assume that while agents interact most of the time in barter trade, with some probability they play other games in which biases lead to strictly lower payoffs. In this setup the rational type, who has a strict advantage, is the sole survivor

of a standard replicator dynamic.

We show, however, that under a wide collection of plausible selection dynamics, the above result is no longer true and types with both biases are stable. We present the family of *hybrid-replicator dynamics* in which, in contrast to the replicator dynamic, a newborn agent does not simply replicate the type of an incumbent agent. In such dynamic, an agent inherits with some probability each bias from a different incumbent, and with the remaining probability inherits both biases from a single incumbent. One plausible interpretation of hybrid-replicator dynamics is a biological evolutionary process where each offspring's genotype is a mixture of his parents' genes. Another interpretation is social learning in which some agents may learn different strategic aspects from different "mentors."

The hybrid-replicator dynamic is not payoff-monotone because only a fraction of the agent's offspring share his type, while the other offspring have "hybrid" types. Consider for example a population composed of a single type in Γ, that is, a type where the two biases compensate for each other. Now assume that this population is invaded by a small group of "mutant" rational agents. Such agents, by definition, have a higher fitness due to their advantage in non-barter activities. However, only a fraction of the rational agent's offspring are rational and this "dilutes" their relative fitness advantage. The remaining hybrid offspring have low fitness, because their single bias is not compensated by the other bias in the barter interaction. As a result, the biased incumbent is stable against unbiased mutants.

We formalize this observation and obtain two key results. Proposition 2 shows a global convergence towards Γ; that is, any type outside Γ can be eliminated by a mutant with the same cursedness level and a perception bias that is strictly closer to Γ. The second result (Proposition 3) shows that each type in Γ is stable against each "mutant" type t' if the barter interaction is sufficiently frequent. We interpret these results as follows. First, the population converges relatively quickly from any initial state towards a type close to Γ. The particular type depends on the initial state and it is generally a biased type. Once the population is near Γ, it remains for a long time in each state, and it slowly drifts between types near Γ such that each type is close to its precedent but has a slightly lower level of biases (giving it an advantage in non-barter activities). This drift is extremely slow, as only mutants that are "close" to the incumbent types can invade. We would therefore expect the population to be composed of agents with both cursedness and the endowment effect for a long period of time. Only in the "ultra-long" run will the population eventually reach the rational type. Previous studies have also discussed stability in shorter horizons than the "ultra-long" run, and suggest that populations can be found in locally optimal states that are not stable in the

"ultra-long" run (see, e.g., Hammerstein, 1996).[5]

Related Literature. The closest paper to ours is Waldman (1994), which shows that evolutionary dynamics with sexual inheritance can yield stable "second-best" adaptations, in which biases approximately compensate for each other. We follow the basic intuitions of Waldman regarding the effect of several parents on the stability of second-best adaptations. However, Waldman's results apply only for the case in which each bias has a finite set of feasible values (Waldman, 1994, p. 488). In many applications this assumption is too restrictive, as the set of feasible types is likely to be rich and convex. Our model is an obvious example: if two cussedness levels are feasible, it seems likely that intermediate levels will be feasible as well. Thus, all that one can infer from Waldman's analysis to our case is that no type except the rational type is asymptotically stable. While this is also a result of our analysis, our new family of hybrid-replicator dynamics allows us to distinguish between the long and ultra long run. We show that, even with a convex set of types, second-best adaptations can survive in the population for a much longer time than other types, and thus may be observed even today. In addition to this new prediction, the introduction of a new dynamic allows us to give a more tractable and complete description of the evolutionary dynamics. In the context of the current paper we show (Proposition 2) a global convergence to the set of second-best adaptations (Γ), while Waldman Waldman (1994) only deals with local stability. Finally, the hybrid replicator dynamics could be useful in future research of various evolutionary processes.

Our paper is also related to the "indirect evolutionary approach" literature (Guth & Yaari, 1992), dealing with the evolution of preferences that deviate from payoff (or fitness) maximization. [6] A main stylized result in this literature (see Ok & Vega-Redondo, 2001; Dekel *et al.*, 2007) is that biases can be stable only if types are observable, so a player can condition his behavior on an opponent's type.[7] We show, in contrast, that even with the "conventional" replicator dynamic, stable states may contain biased players who play as if they were rational on the equilibrium path (however, off the equilibrium path their "mistakes" can be observed).[8] Moreover, when considering hybrid-replicator dynamics, players can also play suboptimally on the equilibrium path.

Finally, Our paper is related to the literature that explains how behavioral biases may

[5] One famous example of this in the human population is Sickle cell disease, which occurs when a person has two mutated alleles. This disease is relatively frequent, especially in areas in which malaria is common, due to the heterozygote advantage: a person with a single mutated allele has a better resistance to malaria.

[6] See Remark 1 below for a discussion on extending this literature to dealing with biases.

[7] A related example is Huck *et al.* (2005) and Heifetz & Segev (2004) who show that an endowment effect observed by others can evolve in populations that engage in bargaining, through its use as a "commitment" device. See also Heller (2013) that shows a related result for limited foresight.

[8] See a related result in Sandholm (2001) in a setup of preference evolution.

evolve. A majority of these papers deal with a single bias. Few papers have dealt with the possibility that evolution will create two biases that are significantly different and yet complementary. Heifetz *et al.* (2007) develop a general framework in which natural selection may lead to perception biases, and show that if preferences are observable, then, generically, non-material preferences will be stable. In a non-evolutionary context, Kahneman & Lovallo (1993) suggest that two biases, excessive risk aversion and the tendency of individuals to consider decision problems one at a time, partially cancel each other out. Recently, Ely (2011) demonstrated that in evolutionary processes improvements tend to come in the form of "kludges," that is, marginal adaptations that compensate for, but do not eliminate, fundamental design inefficiencies. Finally, Herold & Netzer (2011) show that, different biases that are postulated in prospect theory partially compensate for each other, and Johnson & Fowler (2011) show that overconfidence arise naturally in a setup where agents are not fitness-maximizing and use a non-Bayesian decision making heuristic. To the best of our knowledge, this paper is the first to tie between cursedness and the endowment effect.

The paper is organized as follows. Section 2 presents the model, which is analyzed in Section 3. In Section 4 we introduce the hybrid-replicator dynamics and study stability when the biases only partially compensate for each other. We conclude with a discussion in Section 5. All proofs appear in the Appendix.

2 Model

We present a model of barter trade, in which a population consists of a continuum of agents that are randomly matched to engage in a trading interaction. Each agent in the population is endowed with a type that determines his biases. Agents do not observe the types of their trading partners. We describe below the different components of the interaction between each pair, and then move on to describe the induced population game.

2.1 Barter Interaction

A barter interaction is composed of two agents matched as trading partners. Each agent $i \in \{1,2\}$ in the pair owns a different kind of indivisible good, and obtains a private signal $\mathbf{x}_i$ regarding the value of his own good. We assume $\mathbf{x}_1, \mathbf{x}_2$ are continuous, independent, and identically distributed with full support over $[L,H]$, where $0 < L < H$. Let $\mu \equiv E(\mathbf{x}_i)$ be the (ex-ante) expected value of $\mathbf{x}_i$ and let $\mu_{\leq y} \equiv E(\mathbf{x}_i | \mathbf{x}_i \leq y)$ be the expected value of $\mathbf{x}_i$ given that its value is at most y.

Both traders receive a public signal $\alpha \geq 1$, which is a "surplus coefficient" of trade: the

good owned by agent $-i$ is worth $\alpha \cdot \mathbf{x}_{-i}$ to agent i. High α represents better conditions for trade independently of the quality of the goods. For example, if both parties have a great need for the commodity they do not own, then α will be high. Given that α denotes trade conditions other than quality, which is represented by $\mathbf{x}_1$ and $\mathbf{x}_2$, we assume that α, $\mathbf{x}_1$, and $\mathbf{x}_2$ are all independent. The coefficient α can have any continuous distribution with support $\left[1, \frac{H}{L}\right]$.[9] The agents interact by simultaneously declaring whether they are willing to trade. The goods are exchanged if and only if both agents agree to trade.

2.2 Biases / Types

Each agent has a pair of biases, and their specific levels are denoted by its type, $t = (\chi, \psi)$. The first bias is *cursedness* à la Eyster & Rabin (2005). A trader of type $\chi \in [0,1]$ best-replies to a biased belief that the expected value of his partner's good is $\alpha \cdot (\chi \cdot \mu + (1-\chi) \cdot \mu_\alpha)$, where μ_α is the expected value of his partner's good when the partner agrees to trade and the trade coefficient is α. Thus, a cursed trader only partially takes into account the informational content of the other trader's action (a rational agent has $\chi = 0$; a "fully cursed" trader with $\chi = 1$ believes he always gets an "average" object). Notice that if $\mu_\alpha < \mu$ (as we show below), then a χ-cursed trader with $\chi > 0$ overestimates the quality of his partner's good.

The second component, ψ, is a trader's *perception bias* regarding his own good. We assume that $\psi \in \Psi$, where Ψ is the set of continuous and strictly increasing functions from $[L, H]$ to itself. A trader with perception bias ψ best-replies to a biased belief that his own good's value is $\psi(x)$ (rather than x). If $\psi(x) > x$ for all $x \neq H$ we say that the trader exhibits an *endowment effect*. Given two perception biases, ψ_1 and ψ_2, we say that ψ_1 is *more biased* than ψ_2, denoted by $\psi_1 \succeq \psi_2$, if for all $x \in [L, H]$ either $\psi_1(x) \leq \psi_2(x) \leq x$ or $\psi_1(x) \geq \psi_2(x) \geq x$. We write $\psi_1 \succ \psi_2$ if $\psi_1 \succeq \psi_2$ and $\psi_2 \not\succeq \psi_1$. Letting $I \in \Psi$ denote the identity function, $I(x) \equiv x$, type $(0, I)$ is the unbiased, or the "rational" type. Denote by $T \equiv [0,1] \times \Psi$ the set of all types.

2.3 Strategies and Configurations

A general strategy for trader i is a function from α to values of $\mathbf{x}_i$ for which the trader agrees to trade. If an agent expects a positive surplus from trading an object of value x_i, then he

[9] Full support over $\left[1, \frac{H}{L}\right]$ is needed for uniqueness. Specifically, it implies that for each fixed cursedness level, any two different perception biases induce different threshold strategies. Without this property our results would not change qualitatively: the equilibrium we define below would include more types, but the observed behavior of these types would be the same. Finally, the results are similar if we allow for smaller or greater α's, or if we assume full support only over $\left[1, \frac{H}{\mu}\right]$; however, this makes the notation cumbersome.

expects a positive surplus also from trading objects with a value less than x_i. Thus, we can restrict our attention to threshold strategies. An agent's *pure threshold strategy* (or simply, strategy) is a continuous function $s : \left[1, \frac{H}{L}\right] \to \mathbb{R}^+$ that determines, for each α, the maximal value of x for which the agent accepts trade.[10] That is, an agent who follows strategy s accepts trade if and only if $x \leq s(\alpha)$. Note that $\mathbf{x}_i$ is continuous and so there is always a unique best response (see Equation. 1 below), hence the focus on pure strategies is without loss of generality. Since the actual value of the trader's good belongs to the interval $[L, H]$, we interpret thresholds $s(\alpha) < L$ and $s(\alpha) > H$ as strategies where the trader never agrees to trade and always agrees to trade, respectively. Let S denote the set of strategies.

In what follows we assume that the number of types in the population is finite. Specifically, let $\Delta(T)$ be the set of type distributions with finite support (we slightly abuse notation and denote by t degenerate distributions with a single type t). A state of the population, or configuration, is formally defined as follows:

Definition 1. A *configuration* is a pair (η, b), where $\eta \in \Delta(T)$ is a distribution of types and $b : \text{supp}(\eta) \to S$ is a *behavior* function assigning a strategy to each type.

Observe that the definition implies that an agent does not observe his opponent's type. He best-replies while taking into account the value of a random traded good, which is determined jointly by the distribution of types (biases) in the population and the strategy that each type uses in the game.[11] Given a configuration (η, b), let $\mu_\alpha(\eta, b)$ be the mean value of a good of a random trader, conditional on the trader's agreement to trade when the trade coefficient is α (formally defined in the Appendix in (5)). Notice that this mean value is determined jointly by the distribution of types (biases) in the population and the strategy each type employs in the game. Let $s_t^*(\alpha)(\eta, b)$ denote the best-reply threshold of a trader of type $t = (\chi, \psi)$ who is facing a surplus coefficient α and configuration (η, b). Formally, $s_t^*(\alpha)(\eta, b)$ is the unique value in $[L, H]$ that solves the equation

$$\psi\left(s_t^*(\alpha)(\eta, b)\right) = \alpha \cdot \left(\chi \cdot \mu + (1 - \chi) \cdot \mu_\alpha(\eta, b)\right). \tag{1}$$

The interpretation of (1) is as follows. The RHS describes the value of the good a trader expects to receive from a trade, given his cursedness level χ. The LHS describes the value the trader attaches to a good of value $s_t^*(\alpha)(\eta, b)$, given his perception bias ψ. A trader strictly prefers to trade if and only if his good's perceived value is less than the expected value of

[10] Without the mild assumption of continuity in α there may be some Nash equilibria that differ from elements of Γ only with respect to values of x that do not serve as a threshold for any type.

[11] In Remark 1 below we discuss how such a best reply may be a result of a simple learning process.

his partner's good, conditional on this partner agreeing to trade. The unique threshold for which the trader is indifferent between trading and not trading is $s_t^*(\alpha)(\eta, b)$.

We conclude by describing the influence of biases on behavior, as implied by (1):

1. A cursed trader ($\chi > 0$) overestimates the good of his partner and therefore uses a threshold that is too *high* (since we deal with threshold strategies $\mu_\alpha \leq \mu$). That is, cursed traders trade too much.

2. A trader with an endowment effect ($\psi(x) > x$) overestimates the quality of his own good and therefore uses a threshold that is too *low* and trades too little.

2.4 Equilibrium Configurations and the Population Game

A configuration is an equilibrium if each type best-replies in the manner presented above.

Definition 2 (Equilibrium Configuration)**.** A configuration (η, b) is an *equilibrium* if for each type $t = (\chi, \psi) \in \text{supp}(\eta)$ and for every $\alpha \in \left[1, \frac{H}{L}\right]$, $b(t)(\alpha) = s_t^*(\alpha)(\eta, b)$.

Our first result shows the existence of equilibrium configurations. Formally:

Lemma 1 (Existence of Equilibrium Configurations)**.** For every distribution of types $\eta \in \Delta(T)$, there exists a behavior b such that (η, b) is an equilibrium configuration.

There may be more than one equilibrium configuration with the same underlying distribution. In what follows, we assume that for any distribution, one specific equilibrium is always played. Formally, assume an arbitrary function b^* that assigns to each type distribution $\eta \in \Delta(T)$ a behavior such that $\left(\eta, b_\eta^*\right)$ is an equilibrium configuration. One example of b^* is a function that chooses the equilibrium with maximal trade, but any function will do. Since we assume that for each distribution η a specific equilibrium is played, in what follows we can focus solely on the distribution of types η.

The barter-trade interaction together with the equilibrium selection function b^* define a population game $G_0 = (T, u)$, where T is the set of types (as defined above), and $u : T \times \Delta(T) \to [L, H]$ is the payoff function that describes the expected value of a good obtained by a type t agent that best-replies to a trade with a random partner from configuration $\left(\eta, b_\eta^*\right)$; a formal definition of $u(t, \eta)$ appears in Appendix A.2.

Remark 1 (Learning to Best-Reply). Our notion of population game applies the "evolutionary indirect approach" of a two-layer evolutionary process: a relatively slower process according to which the distribution of types evolves, and a faster process according to which agents learn to "subjectively" best-reply to the aggregate behavior in the population. Most of the existing

literature studies a setup in which types represent subjective preferences (see, e.g., Guth & Yaari, 1992; Ok & Vega-Redondo, 2001; Dekel *et al.*, 2007). It is reasonable to ask whether the assumption of subjective best-replying is still appropriate when dealing with behavioral biases such as cursedness, where agents make "mistakes."[12] That is, one may criticize the implicit assumption that agents perfectly understand the aggregate biased behavior of other agents, but then resort to a biased best reply. Observe, however, that the only information an agent requires for best-replying is an estimator to the mean value of a traded good. An agent can form such an estimate simply from observing the value of goods in several past interactions (idiosyncratic errors due to finite samples do not have any qualitative effect on our results). Thus, all we need to assume in order to obtain equilibrium behavior is partial observability of the past, and agents who best reply to an estimate created by a very simple learning process: Non-cursed agents estimate the mean value based only goods that were offered to trade, while cursed agents also take into account the remaining goods.

3 Compensation of Biases in Barter Trade

In this section we analyze the equilibrium of the population game described above.

3.1 Rational and As-If Rational Behavior

As a first step of analyzing G_0, consider the case where the entire population is unbiased; that is, the threshold of each agent is determined by the indifference condition

$$x^*(\alpha) \equiv s^*_{(I,0)}(\alpha)(I,0) = \alpha \cdot \mu_\alpha(I,0),$$

which is derived by substituting $\psi(x) = x$ and $\chi = 0$ into (1). It is easy to show that if all agents play homogeneously in such a way, then in a Nash equilibrium the "rational" threshold, denoted by x^*, must be a solution to the equation

$$x^*(\alpha) = \alpha \cdot \mu_{<x^*(\alpha)}.$$

Next, for each $\chi \in [0,1]$ associate the perception bias $\psi^*_\chi \in \Psi$ defined by

$$\psi^*_\chi(x) \equiv \chi \cdot \frac{\mu}{\mu_{\leq x}} \cdot x + (1-\chi) \cdot x. \tag{2}$$

[12] We focus on cursedness since the endowment effect has a natural interpretation as subjective preferences.

Now, let $\Gamma = \left\{\left(\chi, \psi^*_\chi\right) : \chi \in [0,1]\right\} \subset T$ be the set of all such types. Observe that: (1) $\psi^*_0(x) \equiv I$. Thus the unbiased type $(0, I)$ is in Γ; (2) all other types in Γ are cursed ($\chi > 0$) and exhibit the endowment effect ($\psi^*_\chi(x) > x$ for all $x < H$); and (3) Types in Γ who are more cursed also have a larger endowment effect: $\chi_1 > \chi_2$ implies that $\psi^*_{\chi_1} \succ \psi^*_{\chi_2}$.

Assume that the population behaves according to the "rational Nash equilibrium" described above, and therefore $\mu_\alpha = \mu_{<x^*(\alpha)}$. Then, the threshold chosen by types in Γ is

$$\chi \cdot \frac{\mu}{\mu_{\leq s^*_t(\alpha)}} \cdot s^*_t(\alpha) + (1-\chi) \cdot s^*_t(\alpha) = \alpha \cdot (\chi \cdot \mu + (1-\chi) \cdot \mu_\alpha), \text{ or}$$

$$s^*_t(\alpha) = \alpha \cdot \mu_{<x^*(\alpha)} = x^*(\alpha).$$

That is, when all other agents in the population behave "rationally," the behavior of types in Γ is indistinguishable from the rational type. The endowment effect and cursedness compensate for each other, and the $\left(\chi, \psi^*_\chi\right)$ agent behaves (on the equilibrium path) as if he is rational.

3.2 Equilibrium and Stability

We now formalize the ideas that were presented informally above. First, let us define a Nash equilibrium of the population game, as a distribution of types that is a best reply to itself.

Definition 3. Distribution $\eta \in \Delta(T)$ is a *Nash equilibrium* in game $G_0 = (T, u)$ if $u(t, \eta) \geq u(t', \eta)$ for all $t \in supp(\eta)$ and $t' \in T$.

The following definition assures that a set of types has two properties: (1) all types have the same payoff (internal-equivalence) and it is therefore internally stable; (2) the set is immune to an invasion of a small number of "mutant" types, since those types underperform types in the set (external strictness).

Definition 4. A set of types $Y \subseteq T$ is *internally equivalent* and *externally strict* in game $G_0 = (T, u)$ if $u(t, \eta) = u(t', \eta) > u_p\left(\hat{t}, \eta\right)$ for all $\eta \in \Delta(Y)$, $t, t' \in Y$ and $\hat{t} \in T \backslash Y$.

Using these definitions, we can explicitly formulate our first main result:

Proposition 1. *A distribution $\eta \in \Delta(T)$ is a Nash equilibrium of G_0 if and only if $supp(\eta) \subseteq \Gamma$. Moreover, Γ is an internally equivalent and externally strict set.*

Suppose now a dynamic game where agents are randomly matched each period and play the trade game. Agents' payoffs determine their fitness and therefore their frequency in the population in the next stage; that is, the type distribution η evolves through a payoff-monotone selection dynamic, such as the replicator dynamic (Taylor & Jonker, 1978). Given

Proposition 1, we can rely on existing results to relate the static equilibria with dynamically stable distributions in such a dynamic. We sketch below the argument why the set Γ is dynamically stable, and types outside Γ are unstable.

A distribution of types η is *Lyapunov stable* if after any sufficiently small invasion of a mutant distribution of types, the population composition remains close to η at all future generations.[13] A set of Lyapunov-stable distributions is *asymptotically stable* if, after any small enough invasion of a mutant distribution to any of the distributions in the set, the population reverts back to the set in the long run. An asymptotically stable set is *minimal* if no strict subset is asymptotically stable. For brevity we omit the formal definitions and the formal arguments, which are quite standard.

Recall that we restrict attention to distributions of types with finite support (both for incumbents and for mutants), and thus we can apply the result by Nachbar (1990) that any Lyapunov stable distribution is a Nash equilibrium. Thomas (1985) defines a notion of an *evolutionarily stable set* and shows that it implies asymptotic stability in the replicator dynamic in a finite strategy space (Cressman, 1997, extends this result to a large set of payoff-monotone dynamics). Norman (2008, Theorem 1) extends this result to infinite strategy spaces. Specifically, he shows that if the set of all distributions over a Borel set is evolutionary stable with a uniform invasion barrier, then it is asymptotically stable.

Let $\Delta_\Gamma \subset \Delta(T)$ be the set of distributions over Γ. The fact that Γ is internally equivalent and externally strict implies that Δ_Γ is evolutionarily stable. It is also relatively straightforward to see that the fully-cursed type $(1, \psi_1^*) \in \Gamma$ has a uniform invasion barrier (i.e., there is $\bar{\epsilon} > 0$ such that the incumbent type $(1, \psi_1^*)$ strictly outperforms any mutant type outside Γ with mass $0 < \epsilon < \bar{\epsilon}$,) and that this invasion barrier also holds for any other distribution in Δ_Γ. Thus, one can adapt the result of Norman (2008) to the current setup and conclude that the set Δ_Γ is asymptotically stable.[14] This implies the following corollary of Proposition 1 which characterizes stable distributions in the replicator dynamic.

Corollary 1. *In game G_0 with an underlying replicator dynamic, (1) A distribution η is Lyapunov stable iff $\eta \in \Delta_\Gamma$, and (2) the set Δ_Γ is a minimal asymptotically stable set.*

[13] Note that we only consider perturbations in which an incumbent population is invaded by a small group of mutants (that is, we consider the variational norm when assessing relevant nearby perturbations; see, e.g., Bomze, 1990). We do not consider "continuous" perturbations in which a large group of incumbents slightly change their types as in Eshel & Motro (1981). This is because: (1) a coordinated change in the type of many incumbents seems less plausible in our setup; and (2) the existing results regarding "continuous" stability (e.g., Oechssler & Riedel, 2002) hold only when the set of strategies is a subset of $\mathbb{R}$, and an analysis with the set of strategies in our population game is intractable.

[14] Norman (2008) formally deals with strategy spaces that are subsets of $\mathbb{R}^n$, but it seems that all the arguments in his proof can be extended to the current setup.

4 Preservation of Biases with Additional Activities

In this section we present the "hybrid-replicator dynamics" and show that biased types are stable even when additional activities are present.

4.1 Introducing Additional Activities

The main drawback of the above analysis is the assumption that fitness is a result of a single interaction. Obviously, agents engage in many activities that determine their fitness, and it is reasonable to assume that in many of these activities biases will have negative impact on the biased agent's payoff. For example, in a trade of goods with publicly observable quality, or goods with a private value, cursedness will not be relevant, and the endowment effect will result in an average loss. In what follows we allow agents to engage in additional activities, and assume that in such activities biases are harmful. In this section we show that in this case the rational type has an advantage over all the other types, including those in Γ. In the next section we show that with some plausible assumptions on the dynamics, biased types can be stable.

We model additional activities as follows. With some probability p agents take part in other activities, in which biases are detrimental. Formally, for $0 \leq p \leq 1$, agents play a population game $G_p = (T, u_p)$ with a payoff function $u_p : T \times \Delta(T) \to [L, H]$ defined as

$$u_p(t, \eta) = (1 - p) \cdot u(t, \eta) + p \cdot v(t, \eta), \tag{3}$$

where $v(t, \eta)$, the payoff in other activities, is *bias-monotone:* larger biases yield lower payoffs. Formally, for every $(\chi, \psi), (\chi', \psi') \in T$ and $\eta \in \Delta(T)$:

$$\chi < \chi' \text{ and } \psi \prec \psi' \Rightarrow v((\chi', \psi'), \eta) < v((\chi, \psi), \eta).$$

In addition, we assume that $v(t, \eta)$ is *Lipschitz-continuous.*

By definition, in any game G_p the rational type $(I, 0)$ has maximal v. In the previous section we have shown that in G_0 there are types that behave as if they were rational in some configurations, but clearly this cannot be the case when $p > 0$. In such a case, the rational type will always have an advantage. Therefore, the only Nash equilibrium of G_p (for $p > 0$) is a population of unbiased agents, which implies that under payoff monotonic dynamics, as soon as additional activities are introduced, even when they are infrequent (i.e., with a low probability), only the rational type $(0, I)$ is stable.

4.2 Hybrid-Replicator Dynamics

We model the evolutionary selection process in discrete time with each period τ representing a generation. The dynamics are represented by a deterministic transition function $g : \Delta(T) \to \Delta(T)$ describing the distribution of types in the next generation as a function of the distribution in this generation. The family of hybrid-replicator dynamics is characterized by a *recombination rate* $r \in [0,1]$ that describes the probability that each offspring is randomly assigned to two incumbent agents ("parents") and copies a single trait from each one of them; under the complementary probability each offspring is assigned to a single incumbent and copies both of its traits. As in the replicator dynamic, these random assignments are distributed according to the incumbents' fitness. Below we provide a semi-formal description of the transition function. Formal definitions appear in Appendix A.4.

As is common in the literature, we assume that the expected number of offspring of each agent is equal to his game payoff plus a positive constant that reflects background factors that are unrelated to the game. Let $f_\eta(t) \in R^+$ be the relative fitness of type t in population $\eta \in \Delta(T)$, that is, this agent's fitness divided by the average fitness in the population. Denote by $\eta(\chi)$ and $\eta(\psi)$ the total frequency of types in η with cursedness level $\chi \in [0,1]$ and perception bias $\psi \in \Psi$, respectively. Now, let $f_\eta(\chi)$ and $f_\eta(\psi)$ be the expected relative fitness of types in η with cursedness level χ and perception bias ψ, respectively. The transition function is defined as follows:

$$g(\eta)((\chi,\psi)) = (1-r)\cdot\eta((\chi,\psi))\cdot f_\eta((\chi,\psi)) + r\cdot\eta(\chi)\cdot f_\eta(\chi)\cdot\eta(\psi)\cdot f_\eta(\psi). \qquad (4)$$

The family of hybrid-replicator dynamics extends the standard replicator dynamic (Taylor & Jonker, 1978; reformulated in Weibull, 1997, Chapter 4.1) for which $g(\eta)((\chi,\psi)) = \eta((\chi,\psi))\cdot f_\eta((\chi,\psi))$. Observe that a hybrid-replicator dynamic coincides with the replicator dynamic if either $r = 0$ or one of the biases has the same level in the entire population.[15]

One interpretation of the hybrid-replicator dynamics is biological heredity. If one assumes that each trait (i.e., bias) is controlled by a different locus (position in the DNA sequence), then the probability that each child inherits each trait from a different parent is equal to the biological recombination rate between these loci. This parameter is equal to 0.5 if the two loci are on two different chromosomes, and it is strictly between 0 and 0.5 if the two loci are in different locations in the same chromosome. A hybrid-replicator dynamic is an exact description of a selection process in a haploid population in which each individual carries

[15] If all types have the same cursedness, χ, then $\eta(\chi) = 1$, $\eta(\psi) = \eta((\chi,\psi))$, $f_\eta(\chi) = 1$, and $f_\eta(\psi) = f_\eta((\chi,\psi))$, which implies that this dynamic coincides with the replicator dynamic. The same happens when there is a single ψ.

one copy of each chromosome, and it is a stylized description that captures the important relevant properties of a selection process in a diploid population, like the human population, in which each individual carries two copies of each chromosome (see, e.g., Maynard Smith 1971). One can show that the hybrid-replicator dynamics have the same implications for asymptotic stability as a more detailed description of diploid populations.[16]

An additional interpretation is that of a social learning process. The parameter r determines the frequency of new agents that independently choose two "mentors" and imitate a single trait from each; each of the remaining new agents chooses a single "mentor" and imitates both his traits.

4.3 Instability of Heterogeneous Populations

In order to facilitate the formal analysis presented in Section 4.4, we focus on studying the stability of homogenous populations that include a single incumbent type against the invasion of a single mutant type, rather than studying the stability of heterogeneous distributions against an invasion of heterogeneous groups of mutant types. The restriction to a single mutant type reflects the assumption that mutants are rare; it can be relaxed without affecting the results (but it makes for more cumbersome notation and definitions). The restriction to a single incumbent type is explained in this section where we informally sketch why heterogeneous populations are not stable in our setup.

We begin by demonstrating why a heterogeneous distribution η that includes two types in Γ (t and t') is unstable. Observe that the support of a fixed point of a hybrid-replicator dynamic (with $r > 0$) must be a product set (i.e., $\mathcal{X}_\eta \times \Psi_\eta$), and thus η also includes the two hybrid types that combine a trait from each of these types:

$$\left\{t = \left(\chi, \psi^*_\chi\right), t' = \left(\chi', \psi^*_{\chi'}\right), \left(\chi, \psi^*_{\chi'}\right), \left(\chi', \psi^*_\chi\right)\right\}.$$

For simplicity, assume that η is symmetric: $\eta(t) = \eta(t')$, and $\eta\left(\left(\chi, \psi^*_{\chi'}\right)\right) = \eta\left(\left(\chi', \psi^*_\chi\right)\right)$, and denote $\bar{\chi} = 0.5 \cdot (\chi + \chi')$. Distribution η is not stable against an external mutant with type $\bar{t} = \left(\bar{\chi}, \psi^*_{\bar{\chi}}\right)$ for low p's because: (1) type $\bar{t}$ has approximately the same fitness as types t and t', and (2) the hybrid types that share one of the traits of the mutant type $\bar{t}$, $\left(\bar{\chi}, \psi^*_\chi\right)$, $\left(\bar{\chi}, \psi^*_{\chi'}\right)$,$\left(\chi, \psi^*_{\bar{\chi}}\right)$, and $\left(\chi', \psi^*_{\bar{\chi}}\right)$, are substantially closer to Γ than are the hybrid incumbent types $\left(\chi, \psi^*_{\chi'}\right)$ and $\left(\chi', \psi^*_\chi\right)$. Thus they achieve higher payoff in the barter trade. This implies that $\bar{\chi}$ and $\bar{\psi}$ have higher average fitness than the other biases (i.e., $f_\eta(\bar{\chi}) > f_\eta(\chi), f_\eta(\chi')$

[16] In particular, assuming that each trait is influenced by a different chromosome and that the "mutant" type can be "dominant" in both loci yields $r = 0.5$. Other plausible assumptions (e.g., "recessive" mutants, and the two loci being in different parts of the same chromosome) yield a value strictly between 0 and 0.5.

and $f_\eta\left(\bar{\psi}\right) > f_\eta(\psi), f_\eta(\psi'))$. By (4), a hybrid-replicator dynamic implies that mutant $\bar{t}$ succeeds in invading the population (his fitness is as high as that of the the incumbents, and the fitness of each of his traits is strictly higher than that of the incumbents' traits).

This argument can be extended to general heterogeneous distributions. Observe that the payoff of the barter trade is strictly concave in a trader's thresholds (see part 2 in the proof of Proposition 1, and assume trade occurs with positive probability). In addition, the thresholds of the traders are strictly increasing (decreasing) in the cursedness level (perception bias). These two observations imply that "intermediate" types use intermediate thresholds; that is, if each bias of type t is a mixed average of the respective biases of types t_1 and t_2, then its threshold strategy is strictly between the threshold strategies of types t_1 and t_2. If a heterogeneous population, η, is a fixed point of the dynamic, then different types must use different thresholds (because the support of the population is a product set, and two types that differ in only one of the traits must have different threshold strategies). Intuitively, due to these observations, there is a "mean" type $\bar{t}$ with biases that are weighted averages of the incumbents' biases, such that it uses thresholds that are weighted averages of the incumbents' thresholds, and a similar property holds for his "hybrid" offspring. (The explicit expression of $\bar{t}$ involves technical difficulties and this is why we only sketch an intuitive argument.) Finally, due to strict concavity, mutant agents with type $\bar{t}$ and their "hybrid" offspring would achieve, on average, strictly higher payoffs in the barter trade relative to the incumbents. For a sufficiently low p, this implies that such mutants could successfully invade the population.

4.4 Stability of Types

In this section we analyze the stability of types given an underlying hybrid-replicator dynamic, which is perturbed by allowing at each generation a small probability that a small mass of "mutant" types invades the population.

Our first result shows *global convergence to* Γ from any initial type. Any type $t = (\chi, \psi)$ outside Γ, provided that p is sufficiently low, will be eliminated by any mutant with the same cursedness level as t and a perception bias strictly closer to Γ but not too far from ψ. Formally, let $g^\tau(\eta)$ be the induced distribution of type τ generations after an initial distribution η (i.e, $g^2(\eta) = g(g(\eta))$, etc.).

Definition 5. Type $t' \in T$ *eliminates* $t \in T$ if $\forall\, \epsilon \in (0,1)$: $lim_{\tau\to\infty} g^\tau(\epsilon \cdot t' + (1-\epsilon)\cdot t) = t'$.

Proposition 2. *For every type* $t = (\chi, \psi) \in T \backslash \Gamma$ *there exists* $\bar{p}, \bar{\beta} > 0$ *such that for* $0 \leq p \leq \bar{p}$ *and* $0 < \beta \leq \bar{\beta}$, *type* $t_\beta = \left(\chi, \beta \cdot \psi^*_\chi + (1-\beta)\cdot\psi\right)$ *eliminates type* t.

The sketch of the proof is as follows. For each surplus coefficient α, the unique optimal threshold is $\alpha \cdot \mu_\alpha$. Due to Proposition 1, the fact that $\psi \neq \psi^*_\chi$ implies that there is an interval of α's for which type t uses suboptimal thresholds. Moreover, one can show that for each α for which type t uses a too low (high) threshold, type $\left(\chi, \psi^*_\chi\right)$ uses a too high (low) threshold. Thus, for a sufficiently low β, the mixed perception bias $\beta \cdot \psi^*_\chi + (1-\beta) \cdot \psi$ induces thresholds that are strictly between type t's thresholds and the optimal thresholds. This implies that type t_β achieves a strictly higher payoff against type t, and against any mixed population of these two types. Finally, since hybrid-replicator dynamics are payoff-monotone when all types share the same cursedness level, type t_β eliminates type t.

Given Proposition 2, we can expect that any population outside Γ will converge to Γ in several steps. Our second result shows that all types in Γ are stable in the following sense: for each incumbent type $t \in \Gamma$ and mutant type $t' \in T$, if p is sufficiently small then after a small invasion of mutants the population never moves far away from its pre-entry state t, and it converges back to the pre-entry state in the long run. Formally:

Definition 6. Type $t \in T$ *is asymptotically stable* against type $t' \in T$ if for every $\lambda \in (0,1)$ there exists ϵ such that for every $\epsilon' \in (0, \epsilon)$:

1. *(*Lyapunov stability) $g^\tau\left(\epsilon' \cdot t' + (1-\epsilon') \cdot t\right)(t) > 1 - \lambda$ for every $\tau \in \mathbb{N}$; and
2. $lim_{\tau \to \infty} g^\tau\left(\epsilon' \cdot t' + (1-\epsilon') \cdot t\right) = t$.

Otherwise, we say that type t is asymptotically unstable against t'.

Proposition 3. *Assume that $r > 0$ and let $t \in \Gamma$ and $t' \in T$ where $t \neq t'$. Then, there exists $\bar{p} > 0$ such that for every $p \leq \bar{p}$, type t is asymptotically stable against type t'.*

The sketch of the proof is as follows. First, we adapt a related result of Waldman (1994), and show that type t is asymptotically stable against t' if the fitness of t' is not more than $\frac{1}{1-r}$ times the fitness of t, and in addition the fitness of the hybrid types that share one trait of t and one trait of t' is lower than the fitness of t. The intuition for this is that as long as type t' and the hybrid types are rare, then at each generation the frequency of offspring of type t' is approximately $(1-r)$ times its relative fitness, while its remaining offspring are hybrid (i.e., share one of its traits). Thus, if $(1-r)$ times the fitness of t' and the fitness of each hybrid type is smaller than the fitness of t, then the frequency of all types other than type t converges to 0. Then we observe that Proposition 1 implies that type t best-replies to itself, and that any type that shares one of the traits with t achieves a strictly smaller payoff against type t. Combining the lemma and the observation immediately completes the proof.

An important feature of hybrid-replicator dynamics (for $r > 0$) is that in these dynamics the rational type $(0, I)$ does *not* eliminate all other types when $p > 0$. A simple adaptation of the proof of Proposition 3 yields the following:

Corollary 2. *Any type that is sufficiently close to* Γ *and sufficiently far from* $(0, I)$ *is asymptotically stable against* $(0, I)$ *for a sufficiently small* $p > 0$.

The reason for this is that while type $(0, I)$ strictly outperforms the incumbent type t, its relative fitness is less than $\frac{1}{1-r}$ (assuming the t is sufficiently close to Γ and $p > 0$ is sufficiently small), and the hybrid types with one unbiased trait are strictly outperformed by the incumbent type that has two biases that partially compensate for each other.

Note that parameter $\bar{p}$ in Proposition 3 depends on t and t', and converges to 0 when $t' \to t$. When p is small but positive, type $t \in \Gamma$ is asymptotically stable against all mutant types except those that are very close to t and have slightly smaller biases (such a close mutant, t', can invade because it obtains almost the same payoff in the barter interaction and a strictly higher payoff in the non-barter activity). By continuity, a similar property holds for incumbent types that are very close to Γ. In other words, a type t that is close to Γ may be invaded by some other type t' only if the latter type is close both to Γ and to t. Due to the Lipschitz continuity of v, one can show that each small neighborhood of Γ is an asymptotically stable set for a sufficiently small $p > 0$, and that each type in this neighborhood is asymptotically unstable only against other types that are very close.

Proposition 3 shows that a biased type close enough to Γ can survive as a unique incumbent type for a long time (that is, in the "long run"). Only the rational type, however, is asymptotically stable against all types, and therefore if we wait long enough the population will eventually be composed of that type (that is, the rational type is the only one stable in the "ultra-long run"). The latter implication is a result of the continuity of types: for each $p > 0$ there exists a sufficiently close type that can invade the population. Restricting the set of feasible types to a large finite set (as in Waldman, 1994) would yield asymptotic stability of each type that is close enough to Γ (i.e., there will exist $\bar{p} > 0$ such that for every $p < \bar{p}$, each feasible type close enough to Γ will be stable against all feasible types).

5 Discussion

Random Traits and Empirical Prediction For simplicity, we have described agents with a particular type as having the same cursedness level and the same perception bias. However, the results remain qualitatively the same if the biases are only partially heritable;

that is, if cursedness and perception bias that an agent of type (χ, ψ) exhibits are $\chi + \epsilon_\chi$ and $\psi + \epsilon_\psi$, respectively, where ϵ_χ and ϵ_ψ are random factors unrelated to the agent's type.

Combining this interpretation of our model with the instability of heterogeneous populations (Section 4.3) yields the following empirical prediction. In a specific society agents are expected to have on average a positive cursedness level and a positive endowment effect, and these traits should be uncorrelated (as society should be concentrated around a single type in Γ, and the heterogeneity reflects random type-unrelated "noise"). If one can compare different societies that have independent evolutionary histories, then these societies are expected to be at different states in Γ, and therefore one expects a strong positive correlation between the average cursedness and the average endowment effect across societies. We are unaware of any existing experimental evidence that can either validate or falsify this prediction.

Richer Environments The model described in Section 4.1 can capture (with slight modifications) an environment in which agents randomly engage in one of *many* barter-trade interactions, which are relatively similar to each other (rather than playing one game with high probability and a completely different activity with low probability, as in the Section 4). For concreteness, consider the following example. Assume that each period traders' private signals are drawn from a different distribution. The distribution is chosen i.i.d. each period according to a normal random variable $\nu_\sigma \sim N(0, \sigma^2)$. Given ν_σ, the private value of each trader is distributed according to the CDF $F_{\nu_\sigma}(x) = \left(\frac{x-L}{H-L}\right)^{\exp(\nu_\sigma)}$. Both agents publicly observe the realization of ν_σ before they decide whether or not to trade (the case in which ν_σ is unobservable is equivalent to playing a fixed barter trade). This environment can be embedded in our model. Let the function u be defined as the payoff function in the barter trade with the distribution of private signals $F_0(x)$, and the function $v(\sigma)$ be defined as the expected (with respect to ν_σ) between the payoff of the game with $F_{\nu_\sigma}(x)$ and the payoff in the game $F_0(x)$. Now let $u_\sigma = u + v(\sigma)$ be the expected payoff in such an environment. Note that the standard deviation σ replaces p as the magnitude of the perturbation in the model. Observe that: (1) the function $u_\sigma(t, \eta)$ describes the expected payoff of type t in population η in this environment; (2) $v(\sigma) \sim O(\sigma)$ for a sufficiently small σ; and (3) function $v(\sigma)(t, \eta)$ is Lipschitz continuous and bias-monotone. Due to these observations, all our stability results can be extended to this setup in a relatively straightforward way when σ is sufficiently small. This example can be generalized to any family of ex-ante distributions of private values, as long as the variance among these distributions is sufficiently small.

A Technical Appendix

A.1 Proof of Lemma 1 (Existence of Equilibrium Configurations)

Denote by F the absolutely continuous cumulative distribution function of $\mathbf{x}_i$. The explicit formula for the expected value of the partner's good conditional on his agreement to trade , $\mu_\alpha(\eta, b)$, is

$$\mu_\alpha(\eta, b) = \frac{\sum_{t\in\text{supp}(\eta)} \eta(t) \cdot F(b(t)(\alpha)) \cdot \mu_{\leq b(t)(\alpha)}}{\sum_{t\in\text{supp}(\eta)} \eta(t) \cdot F(b(t)(\alpha))}, \tag{5}$$

and if the denominator equals 0 (i.e., no agent ever agrees to trade given coefficient α), let $\mu_\alpha(\eta, \beta) = L$.

Let $\{t_i = (\chi_i, \psi_i)\}_{i\leq n}$ be the finite set of types in population η (n is the arbitrary number of types). Substituting (5) in (1), configuration (η, b) is an equilibrium if and only if for each $\alpha \in \left[1, \frac{H}{L}\right]$ and for each type $i \leq n$:

$$\psi(b(t_i)(\alpha)) = \alpha \cdot \left(\chi_i \cdot \mu + (1-\chi_i) \cdot \frac{\sum_{j\leq n} \eta(t_j) \cdot F(b(t_j)(\alpha)) \cdot \mu_{\leq b(t_j)(\alpha)}}{\sum_{j\leq n} \eta(t_j) \cdot F(b(t_j)(\alpha))}\right). \tag{6}$$

Fix any $\alpha \in \left[1, \frac{H}{L}\right]$. Let $x_i = b(t_i)(\alpha)$ and $\eta_i = \eta(t_i)$. Then (6) is reduced to the following set of n equations:

$$\forall i \leq n \;\; \psi(x_i) = \alpha \left(\chi_i \cdot \mu + (1-\chi_i) \cdot \frac{\sum_{j\leq n} \eta_j \cdot F(x_j) \cdot \mu_{\leq x_j}}{\sum_{j\leq n} \eta_j \cdot F(x_j)}\right). \tag{7}$$

For each $i \leq n$, let $g_{i,\alpha} : [L, \mu] \to [L, \alpha \cdot \mu]$ be a function that assigns, for any expected value of an object that is traded in the population v, the threshold of agent with type t_i. Formally:

$$g_{i,\alpha}(v) = \psi^{-1}(\alpha \cdot (\chi_i \cdot \mu + (1-\chi_i) \cdot v)). \tag{8}$$

Notice that in equilibrium $x_i = b(t_i)(\alpha) = g_{i,\alpha}(\mu_\alpha(\eta, b))$. Now let $h : [L, \alpha \cdot \mu]^n \to [L, \mu]$ be the function that assigns, to a profile of thresholds that are used in population η, $(x_1, ..., x_n)$, the expected value of an object that is traded in that population. Formally:

$$h_\eta(x_1, ..., x_n) = \frac{\sum_{j\leq n} \eta_j \cdot F(x_j) \cdot \mu_{\leq x_j}}{\sum_{j\leq n} \eta_j \cdot F(x_j)}.$$

Notice that $\mu_\alpha(\eta, b) = h_\eta[b(t_1)(\alpha), ..., b(t_n)(\alpha)]$. Finally, let $f : [L, \mu] \to [L, \mu]$ be defined as follows:

$$f_\alpha(v) = h_\eta(g_{1,\alpha}(v), ..., g_{n,\alpha}(v)).$$

Observe that any solution to the equation $v = f_\alpha(v)$ induces thresholds that can be used as part of an equilibrium when the coefficient surplus is equal to α; that is, $v = \mu_\alpha(\eta, b)$ and $b(t_i)(\alpha) = g_{i,\alpha}(v)$. We now show that there is at least one solution for $v = f_\alpha(v)$. First observe that since we assume that F and ψ are continuous functions then g, h, and f are a continuous. Second, notice that for each $\alpha \in \left[1, \frac{H}{L}\right]$ $f_\alpha(L) \geq L$ and $f_\alpha(\mu) \leq \mu$. Thus $f_\alpha(v) = v$ at some $v \in [L, \mu]$.

Finally, we have to show that there exists a *continuous* function $v^* : \left[1, \frac{H}{L}\right] \to [L, H]$ such that for each $\alpha \in \left[1, \frac{H}{L}\right]$ $f_\alpha(v^*(\alpha)) = v^*(\alpha)$.[17] Let $v^*(\alpha)$ be the largest fixed point of f_α (i.e, for each α $v^*(\alpha) = \arg\max_{v \in \left[1, \frac{H}{L}\right]} f_\alpha(v) = v$). The fact that ψ^{-1} and h are continuous and increasing functions and the linear dependency in α in (8) implies the continuity of $v^*(\alpha)$ with respect to α.

Note that the above $v^*(\alpha)$ is the function that selects the equilibria with maximal trade. All of our results remain the same if one chooses a different continuous function $v^*(\alpha)$, such as the one that selects the smallest fixed point of f_α (and minimizes the probability of trade).

A.2 Formal Definition of the Payoff Function of G_0

In this section we explicitly state the payoff of each type in the population game G_0. Let $u(s_1, s_2|\alpha, x_1, x_2)$ be trader's 1 payoff when his signal is x_1 and he plays s_1 while his partner, trader 2, has a signal x_2 and plays s_2, and when the public signal is α:

$$u_0(s_1, s_2|\alpha, x_1, x_2) = \begin{cases} \alpha \cdot x_2 & x_1 \leq s_1(\alpha) \text{ and } x_2 \leq s_2(\alpha) \\ x_1 & \text{else.} \end{cases}.$$

Now, let $u(s_1, s_2)$ be the expected payoff of an agent with strategy s_1 who faces a partner with s_2, where the expectation is taken WRT the values of of the signals $\mathbf{x_1}, \mathbf{x_2}$ and α.

$$u(s_1, s_2) = \int_{\alpha=1}^{H/\mu} \int_{x_1=L}^{H} \int_{x_2=L}^{H} u_0(s_1, s_2|\alpha, x_1, x_2) \, \mathrm{d}F_{\mathbf{x_2}} \, \mathrm{d}F_{\mathbf{x_1}} \, \mathrm{d}F_\alpha,$$

where F_j is the CDF of random variable $i \in \{\alpha, \mathbf{x_1}, \mathbf{x_2}\}$. Finally, given type $t \in T$ and an equilibrium configuration $(\eta, b^*(\eta))$, define $u(t, \eta)$ as the expected payoff of a type $t \in T$ agent who faces an opponent randomly selected from population η:

$$u(t, \eta) = \sum_{t' \in supp(\eta)} \eta(t') \cdot u\left[s_t^*(\alpha)(\eta, b^*(\eta)), b^*(\eta)(t')\right].$$

[17] We have to show the continuity of $v^*(\alpha)$ in order to have all types using a continuous threshold strategy as a function of surplus coefficient (as assumed in the definition of a strategy in Section 2.3).

A.3 Proof of Proposition 1 (Characterization of Equilibria in G_0)

Proof. The proof includes the following parts:

1. Definitions and notations:

   (a) *Probability of trade*: let $q(\alpha|\eta)$ be the probability that a random partner from population η agrees to trade given α. Note, that $q(\alpha|\eta) > 0$ iff $\exists t \in supp(\eta)$ with $b^*_\eta(t)(\alpha) > L$.

   (b) *Expected payoff of a threshold*: let $u(x, \alpha|\eta)$ be the expected payoff of a player who uses threshold $x \in [L, H]$, and faces a distribution of types η (which plays according to b^*_η) conditional on the coefficient beingα.

   (c) *Incumbents*: the types in the support of a given distribution η.

   (d) *Type's threshold strategy*: for each type $t = (\chi, \psi)$ let strategy $s^*_t(\alpha|\eta)$ be the threshold strategy of type t who faces population η; that is, for each $\alpha \in \left[1, \frac{H}{L}\right]$ $s^*_t(\alpha|\eta)$ is the unique solution to the equation:

   $$\psi\left(s^*_t(\alpha|\eta)\right) = \alpha \cdot \left(\chi \cdot \mu + (1-\chi) \cdot \mu_\alpha\left(\eta, b^*_\eta\right)\right).$$

   (e) *Mean threshold*: For each $\alpha \in \left[1, \frac{H}{L}\right]$ and distribution of types $\eta \in \Delta(T)$,let $\bar{x}(\alpha|\eta)$ be the unique solution to the equation: $\mu_{\leq \bar{x}(\alpha|\eta)} = \mu_\alpha(\eta)$; that is, in an homogenous population in which everyone uses threshold $\bar{x}(\alpha|\eta)$, the mean value of a traded good is the same as in population η (given α).

2. The expected payoff of a threshold is given by the following formula:

   $$u(x, \alpha|\eta) = u(\alpha \cdot \mu_\alpha, \alpha|\eta) - q(\alpha|\eta) \cdot |F(x) - F(\alpha \cdot \mu_\alpha)| \cdot E\left(|y - \alpha \cdot \mu_\alpha| \mid y \in [\alpha \cdot \mu_\alpha, x]\right), \tag{9}$$

   where, with slight abuse of notation, $[\alpha \cdot \mu_\alpha, x] = [x, \alpha \cdot \mu_\alpha]$ if $x < \alpha \cdot \mu_\alpha$. The *optimal threshold*, which induces the maximal payoff, is $x = \alpha \cdot \mu_\alpha$ because it results in trade iff the trader's good is worth less than the expected value of a trading partner. Using a different threshold x yields a wrong decision with probability $q(\alpha|\eta) \cdot |F(x) - F(\alpha \cdot \mu_\alpha)|$. Conditional on making a wrong decision and the partner's agreement to trade, the expected loss from trade is equal to $E\left(|y - \alpha \cdot \mu_\alpha| \mid y \in [\alpha \cdot \mu_\alpha, x]\right)$ (the expected difference between the value of the trader's good and the conditional expected value of his partner). Observe that $u(x, \alpha|\eta)$ is concave w.r.t. x, and strictly concave if $q(\alpha|\eta) > 0$.

3. The previous step immediately implies that type $(0, I)$ (Which always choose the optimal threshold of $\alpha \cdot \mu_\alpha$) weakly outperforms any other type (i.e., $u((0,I),\eta) \geq u(t,\eta)$ for each $t \in T$ and $\eta \in \Delta(T)$), and strictly outperforms if there is any α such that: $q(\alpha|\eta) > 0$ and $x \neq \alpha \cdot \mu_\alpha$.

4. The "if" side. Let $\eta \in \Delta(T)$ be a Nash equilibrium of G_0. We prove that $supp(\eta) \in \Gamma$.

   (a) In any Nash equilibrium all types use the optimal thresholds for all αs; that is, $\forall t \in supp(\eta)$ and $\alpha \in \left[1, \frac{H}{L}\right]$ $b^*_\eta(t)(\alpha) = \alpha \cdot \mu_\alpha(\eta)$.

   Assume in contrast that there is at least one value of α for which one of the incumbent types uses a non-optimal threshold. This set of α's is defined as

   $$A_\alpha = \left\{\alpha \mid \exists t_0 \in supp(\eta) \text{ s.t. } b^*_\eta(t_0)(\alpha) \neq \alpha \cdot \mu_\alpha(\eta)\right\}.$$

   From 3 and the continuity of the threshold strategies an incumbent type in a Nash equilibrium distribution can use a non-optimal threshold in an α only if $q(\alpha|\eta) = 0$.

   Let $\bar{\alpha}$ be the supremum of A_α. Consider first the case where $\bar{\alpha} = \frac{H}{L}$. In such a case, the assumptions that each ψ is strictly increasing and $\psi(H) \leq H$ implies that $\psi(x) < H$ for each $x < H$, and therefore $q(\alpha|\eta) > 0$ for α's sufficiently close to $\frac{H}{L}$, and $u((0,I),\eta) > u(t,\eta)$ for some type $t \in supp(\eta)$.

   Now consider the case where $\bar{\alpha} < \frac{H}{L}$. All types play the rational threshold for $\alpha > \bar{\alpha}$, and by continuity of $b^*_\eta(t)(\alpha)$ all agents play the rational threshold at $\bar{\alpha}$. Observe, that the rational thresholds are always strictly larger than L, and this implies that $q(\alpha|\eta) > 0$ for each $\alpha \geq \bar{\alpha}$. By continuity of $b^*_\eta(t)(\alpha)$, there exist an interval of αs such that: $\alpha < \bar{\alpha}$, $q(\alpha|\eta) > 0$ and $\alpha \in A_\alpha$, and this implies that η cannot be a Nash equilibrium.

   (b) If there exists $t_0 = (\chi_0, \psi_0) \in T \backslash \Gamma$ in $supp(\eta)$ then there exists an α such that type t_0 does not use the optimal threshold: i.e., $\exists \alpha_0 \in \left[1, \frac{H}{L}\right]$ $b^*_\eta(t_0)(\alpha) \neq \alpha \cdot \mu_\alpha(\eta)$.

   Assume to the contrary that for each $\alpha \in \left[1, \frac{H}{L}\right]$, all incumbents $t \in supp(\eta)$ use the "optimal" threshold $b^*_\eta(t)(\alpha) = x^*(\alpha) \equiv \alpha \cdot \mu_\alpha$. This implies that $\forall \alpha \in \left[1, \frac{H}{L}\right]$:

   $$\psi_0(\alpha \cdot \mu_\alpha) = \alpha \cdot (\chi_0 \cdot \mu + (1 - \chi_0) \cdot \mu_\alpha).$$

   The fact that all players use the optimal thresholds implies that $\mu_\alpha = \mu_{\leq x^*(\alpha)}$, and therefore $x^*(1) = L$ and $x^*\left(\frac{H}{L}\right) = H$. By continuity, $x^*(\alpha)$ obtains all values in $[L, H]$. Given $x^*(\alpha) \equiv \alpha \cdot \mu_\alpha$ we can rewrite the indifference condition above as

follows: $\forall x^* \in [L, H]$,

$$\psi_0(x^*) = \alpha \cdot \left(\chi_0 \cdot \mu + (1-\chi) \cdot \frac{x^*}{\alpha}\right) = \chi_0 \cdot \frac{\mu}{\mu_{\leq x^*}} \cdot x^* + (1-\chi) \cdot x^* = \psi^*_{\chi_0}(x^*),$$

which implies that $\psi_0(x) = \psi^*_{\chi_0}(x)$ – a contradiction to that $t_0 \in T \backslash \Gamma$.

5. The "only if" side and the "moreover" statement. Let η_0 be a distribution with $supp(\eta_0) \in \Gamma$. We prove that η_0 is a Nash equilibrium, that all incumbents use the same threshold strategy, and that each type $t' \notin \Gamma$ is strictly out-performed (i.e, $u(t', \eta) < u(t_0, \eta)$ for each $t_0 \in supp(\eta_0)$.

   (a) In Section 3.1 we have shown that if all other agents use the "rational" threshold that is defined by $x^*(\alpha) = \alpha \cdot \mu_{<x^*(\alpha)}$, then a type $t \in \Gamma$ also uses $x^*(\alpha)$. Thus, $b^* = x^*(\alpha)$ is an equilibrium behavior induced by distribution η_0. In what follows we show that there are no other equilibrium behaviors induced by distribution η_0.

   (b) Assume, in contrast, another equilibrium behavior, that induces $\bar{x}(\alpha|\eta_0) \neq x^*(\alpha)$

      i. If $\bar{x}(\alpha|\eta) > \alpha \cdot \mu_\alpha(\eta_0)$ then $s^*_t(\alpha|\eta_0) < \bar{x}(\alpha|\eta_0)$ for each $t = \left(\chi, \psi^*_\chi\right) \in \Gamma$. To see why recall that ψ^*_χ is strictly increasing, and observe that $s^*_t(\alpha|\eta_0)$ is the unique solution to the following equation:

$$\begin{aligned} \psi^*_\chi\left(s^*_t(\alpha|\eta_0)\right) &= \alpha \cdot (\chi \cdot \mu + (1-\chi) \cdot \mu_\alpha(\eta_0)) \\ &< \chi \cdot \frac{\mu}{\mu_{\leq \bar{x}(\alpha|\eta_0)}} \cdot \bar{x}(\alpha|\eta_0) + (1-\chi) \cdot \bar{x}(\alpha|\eta_0) = \psi^*_\chi\left(\bar{x}(\alpha|\eta_0)\right), \end{aligned}$$

      where the strict inequality is implied by $\bar{x}(\alpha|\eta_0) > \alpha \cdot \mu_\alpha(\eta_0)$ and $\mu_\alpha(\eta_0) = \mu_{\leq \bar{x}(\alpha|\eta_0)}$. Since $\bar{x}(\alpha|\eta)$ is the average threshold, we get a contradiction.

      ii. By analogous argument, if $\bar{x}(\alpha|\eta_0) < \alpha \cdot \mu_\alpha$ then $s^*_t(\alpha|\eta_0) > \bar{x}(\alpha|\eta_0)$ for each $t = \left(\chi, \psi^*_\chi\right) \in \Gamma$, and we get again a contraction.

      iii. Therefore, it must be that $\bar{x}(\alpha|\eta_0) = \alpha \cdot \mu_\alpha$. In such a case, by an analogous argument, $s^*_t(\alpha|\eta_0) = \bar{x}(\alpha|\eta_0) = x^*(\alpha)$ for each $t = \left(\chi, \psi^*_\chi\right) \in \Gamma$.

   (c) The previous parts imply that if $supp(\eta_0) \in \Gamma$ then all incumbent types use threshold strategy $x^*(\alpha)$ and are internally-equivalent. Moreover, this threshold strategy is the optimal one, and this implies that η_0 is a Nash equilibrium.

   (d) Finally, we can use a similar argument to 4b to show that for any α there is a positive probability of trade, and therefore any type $t' \notin \Gamma$, that uses non-optimal threshold for some α's, has a strictly lower payoff compared to the incumbents

□

A.4 Formal Definition of the Hybrid-Replicator Dynamics

In this section we formalize the definition of the transition function $g : \Delta(T) \to \Delta(T)$ that is described informally in Section 4.2. The relative fitness $f_\eta(t)$ of type t in population η is

$$f_\eta(t) = \frac{\phi + u_p(t,\eta)}{\phi + E(u_p(t,\eta))} = \frac{\phi + u_p(t,\eta)}{\phi + \sum_{t' \in supp(\eta)} \eta(t') \cdot u_p(t',\eta)}, \tag{10}$$

where $\phi \geq 0$ is the background expected number of offspring for an individual (unrelated to his payoff in the population game). Let $\mathcal{X}_\eta$ and Ψ_η be the cursedness levels and the perception biases in population η:

$$\mathcal{X}_\eta = \{\chi \in [0,1] \mid \exists \psi \in \Psi \text{ s.t. } (\chi,\psi) \in supp(\eta)\},$$

$$\Psi_\eta = \{\psi \in \Psi \mid \exists \chi \in [0,1] \text{ s.t. } (\chi,\psi) \in supp(\eta)\}.$$

For each $\chi \in \mathcal{X}_\eta$ ($\psi \in \Psi_\eta$) define $\eta(\chi)$ ($\eta(\psi)$) as the total frequency of types with χ (ψ):

$$\eta(\chi) = \sum_{\psi \in \ominus_\eta} \eta((\chi,\psi)) \quad \left(\eta(\psi) = \sum_{\chi \in \mathcal{X}_\eta} \eta((\chi,\psi))\right),$$

and define $f_\eta(\chi)$ ($f_\eta(\psi)$) as the mean relative fitness of types with cursedness χ (bias ψ):

$$f_\eta(\chi) = E(f_\eta((\chi,\psi)) \,|\psi \in \Psi_\eta) = \sum_{\psi \in \ominus_\eta} \frac{\eta((\chi,\psi))}{\eta(\chi)} \cdot f_\eta((\chi,\psi))$$

$$\left(f_\eta(\psi) = E(f_\eta((\chi,\psi)) \,|\chi \in \mathcal{X}_\eta) = \sum_{\chi \in \mathcal{X}_\eta} \frac{\eta((\chi,\psi))}{\eta(\psi)} \cdot f_\eta((\chi,\psi))\right).$$

Lastly, the transition function is (for every $(\chi,\psi) \in \mathcal{X}_\eta \times \Psi_\eta$)

$$g(\eta)((\chi,\psi)) = (1-r) \cdot \eta((\chi,\psi)) \cdot f_\eta((\chi,\psi)) + r \cdot \eta(\chi) \cdot f_\eta(\chi) \cdot \eta(\psi) \cdot f_\eta(\psi).$$

A.5 Proof of Proposition 2 (Convergence Towards Γ)

Proof. The proof uses definitions and notations that were introduced in the proof of Prop. 1. Assume an homogenous population $\eta = t = (\chi,\psi)$ where t is an arbitrary type. Define $t_\Gamma \equiv \left(\chi, \psi^*_\chi\right)$ as the type in Γ with the same cursedness as t.

1. By an analogous argument to part 4 of the proof of Prop. 1, there is an interval of α's such that: (1) type t induces trade with positive probability (i.e., $q(\alpha|t) > 0$); and (2) types t and t_Γ use different thresholds (i.e, $s_t^*(\alpha|\eta) \neq s_{t_\Gamma}^*(\alpha|\eta)$.

2. The optimal threshold is between the thresholds induced by types t and t_Γ: for each $\alpha \in \left[1, \frac{H}{L}\right]$, either $s_{t_\Gamma}^*(\alpha|t) < \alpha \cdot \mu_\alpha < s_t^*(\alpha|t)$ or $s_t^*(\alpha|t) < \alpha \cdot \mu_\alpha < s_{t_\Gamma}^*(\alpha|t)$.
Assume without loss of generality that $s_t^*(\alpha|t) > \alpha \cdot \mu_\alpha$(the proof is analogous in the opposite case). We have to show that type t_Γ does not want to trade if he has an object with value $\alpha \cdot \mu_\alpha$. Using (1), (2) and the monotonicity of ψ_χ^* we write

$$\psi_\chi^*(\alpha \cdot \mu_\alpha) \overset{?}{>} \alpha \cdot (\chi \cdot \mu + (1-\chi) \cdot \mu_\alpha) \Leftrightarrow$$

$$\chi \cdot \frac{\mu}{\mu_{\leq \alpha \cdot \mu_\alpha}} \cdot \alpha \cdot \mu_\alpha + (1-\chi) \cdot \alpha \cdot \mu_\alpha \overset{?}{>} \alpha \cdot \chi \cdot \mu + \alpha \cdot (1-\chi) \cdot \mu_\alpha \Leftrightarrow \mu_\alpha \overset{?}{>} \mu_{\leq \alpha \cdot \mu_\alpha},$$

and the last inequality is satisfied since $\mu_\alpha = \mu_{\leq s_t^*(\alpha|t)}$ and $s_t^*(\alpha|t) > \alpha \cdot \mu_\alpha$.

3. By continuity, the previous step implies that there exists $\bar{\beta} > 0$ such that type $t_{\bar{\beta}} = \left(\chi, \beta \cdot \psi_\chi^* + (1-\beta) \cdot \psi\right)$ induces threshold between t's threshold and the optimal threshold. That is, for each $\alpha \in \left[1, \frac{H}{L}\right]$ with $x(t, \alpha|t) \neq \alpha \cdot \mu_\alpha$ either $x(t, \alpha|t) < x(t_\beta, \alpha|t) < \alpha \cdot \mu_\alpha$ or $x(t, \alpha|t) > x(t_\beta, \alpha|t) > \alpha \cdot \mu_\alpha$. Note that $u\left(t_{\bar{\beta}}, t\right) > u(t, t)$.

4. By continuity, there exist $\bar{p} > 0$ such that for each $0 \leq p \leq \bar{p}$ and $0 < \beta < \bar{\beta}$, type $t_\beta = \left(\chi, \beta \cdot \psi_\chi^* + (1-\beta) \cdot \psi\right)$ satisfies $u_p(t_\beta, t) > u_p(t, t)$.

5. Observe, that all the above arguments work also in an heterogeneous population that includes a mixture of types t and t_β. That is, for each $\gamma \in (0, 1)$,

$$u_p(t_\beta, \gamma \cdot t + (1-\gamma) \cdot t_\beta) > u_p(t, \gamma \cdot t + (1-\gamma) \cdot t_\beta).$$

As noted in Section 4.2, any hybrid-replicator dynamic coincides with the replicator dynamic when all types share the same cursedness. This implies that type t_β eliminates type t because it always achieves a strictly higher payoff.

□

A.6 Stability in Hybrid-Replicator Dynamics and Prop. 3's Proof

In what follows, we first prove a lemma that characterizes stability in hybrid-replicator dynamics. The lemma shows that an incumbent type is asymptotically stable against a mutant

type if: (1) the mutant's payoff is not substantially higher than incumbent's payoff (specifically, the mutant's fitness should be less than $\frac{1}{1-r}$ times the incumbent's fitness); and (2) the hybrid types (who have one trait from type t and one trait from type t') yield lower payoffs than the incumbent. We then use this lemma to prove Proposition 3 that deals with stability of types in Γ. The lemma is a simple adaptation of Prop. 2 of Waldman (1994).

Lemma 2 (Characterization of Stability in the Hybrid-Replicator Dynamics)**.** *Let $t_1 = (\chi_1, \psi_1)$ and $t_2 = (\chi_2, \psi_2)$ denote some arbitrary types. Assume a population game G_p with a hybrid-replicator dynamic with parameters ϕ and r, and let $u_p(t, \eta)$ and $f_\eta(t)$ be the expected payoff and relative fitness of type t against population η as defined in (3) and (10).*

1. *If $(1-r) \cdot f_{t_1}(t_2) > 1$, then type t_1 is asymptotically unstable against t_2.*
2. *If $u_p((\chi_2, \psi_1), t_1) > u_p(t_1, t_1)$, then type t_1 is asymptotically unstable against (χ_2, ψ_1).*
3. *If $u_p((\chi_1, \psi_2), t_1) > u_p(t_1, t_1)$, then type t_1 is asymptotically unstable against (χ_1, ψ_2).*
4. *If (a) $(1-r) \cdot f_{t_1}(t_2) < f_{t_1}(t_1)$, (b) $u_p((\chi_2, \psi_1), t_1) < u_p(t_1, t_1)$, and (b) $u_p((\chi_1, \psi_2), t) < u_p(t_1, t_1)$, then type t_1 is asymptotically stable against t_2.*

Proof. Let $\epsilon > 0$ be sufficiently small, and let the initial distribution be: $\eta_0 = (1-\epsilon) \cdot t_1 + \epsilon \cdot t_2$.

Part (1). Assume that $(1-r) \cdot f_{t_1}(t_2) = c > 1$. By neglecting components that are $O(\epsilon^2)$, Eq. (4) implies that $g^\tau(\eta_0)(t_2) \approx \epsilon \cdot ((1-r) \cdot f_{t_1}(t_2))^\tau = \epsilon \cdot c^\tau$ and this implies instability of t_1 against t_2.

Parts (2)-(3) are immediately implied by well-known results for the replicator dynamic and the observation that when the population includes a single cursedness level (or a single perception bias), then the hybrid-replicator dynamic coincides with a replicator dynamic.

Part (4). Assume inequalities (a)-(c). Inequalities (b) and (c) imply $f_{t_1}((\chi_2, \psi_1)) < 1$ and $f_{t_1}((\chi_1, \psi_2)) < 1$. Define a constant c such that

$$\max\left\{(1-r) \cdot f_{t_1}(t_2), f_{t_1}((\chi_2, \psi_1)), f_{t_1}((\chi_1, \psi_2))\right\} < c < 1.$$

Let $\eta_\tau = g^\tau(\eta_0)$ be the distribution after τ generations. By neglecting components that are $O(\epsilon^2)$, Eq. (4) implies that for every τ in which $\eta_\tau((\chi_1, \psi_2))$ and $\eta_\tau((\chi_2, \psi_1))$ are $O(\epsilon)$: $\eta_\tau(t_2) \approx \epsilon \cdot c^\tau$ which converges to 0 at an exponential rate. Assume to the contrary, that $\eta_\tau((\chi_1, \psi_2))$ does not converge to zero at an exponential rate. Then, for a sufficiently large τ, $\eta_\tau(t_2) << \eta_\tau((\chi_1, \psi_2))$. In addition $f_{\eta_\tau}(\psi_2) \approx f_{\eta_\tau}((\chi_1, \psi_2))$ where $f_{\eta_\tau}(\psi_2)$ is the average relative fitness of all types with $\psi = \psi_2$ (defined in Section A.4) – $f_{\eta_\tau}(\psi_2)$ is a mixed average

of $f_{\eta_\tau}((\chi_1,\psi_2))$ and $f_{\eta_\tau}(t_2)$, and as the weight of the latter converges quickly to zero, the average converges to the former. Let $\epsilon' = \eta_\tau((\chi_1,\psi_2))$. Thus we obtain that:

$$\eta_{\tau+1}((\chi_1,\psi_2)) \approx (1-r)\cdot\epsilon'\cdot f_{\eta_\tau}((\chi_1,\psi_2)) + r\cdot\epsilon'\cdot f_{\eta_\tau}(\psi_2) \approx \epsilon'\cdot f_{\eta_\tau}((\chi_1,\psi_2)) < \epsilon'\cdot c$$

which implies that $\eta_\tau((\chi_1,\psi_2))$ converges to zero at an exponential rate. An analogous arguments works for (χ_2,ψ_1).

□

Proof of Prop. 3. Let the incumbent type be $t = \left(\chi,\psi^*_\chi\right) \in \Gamma$ and let the mutant type be $t' = (\chi',\psi')$. By Proposition 1, in G_0 type $t \in \Gamma$ plays optimally against itself and any type outside Γ achieves a strictly smaller payoff against type t. That is, $u(t',t) \leq u(t,t)$ with strict inequality if $t' \notin \Gamma$. Observe that $(\chi,\psi'), \left(\chi',\psi^*_\chi\right) \notin \Gamma$ because each pair of types in Γ differ in both traits. These observations imply that for each $\delta > 0$ there exists $\bar{p} > 0$ such that for each $p \leq \bar{p}$: (1) $u_p(t',t) < u_p(t,t) + \delta$, (2) $u_p((\chi,\psi'),t) < u_p(t,t)$, and (3) $u_p((\chi',\psi),t) < u_p(t,t)$. Notice that δ can be chosen to be sufficiently small so that the first inequality will imply $(1-r)\cdot f_t(t') < 1$. By Lemma 2, these inequalities imply that type t is asymptotically stable against t'.

References

Apicella, Coren L, Azevedo, Eduardo M, Fowler, James H, & Christakis, Nicholas A. 2014. Evolutionary origins of the endowment effect: evidence from hunter-gatherers. *American Economic Review*, **104(6)**, 1793–1805.

Bokhari, Sheharyar, & Geltner, David. 2011. Loss Aversion and Anchoring in Commercial Real Estate Pricing: Empirical Evidence and Price Index Implications. *Real Estate Economics*, **39**(4), 635–670.

Bomze, Immanuel M. 1990. Dynamical aspects of evolutionary stability. *Monatshefte für Mathematik*, **110**(3-4), 189–206.

Cesarini, David, Johannesson, Magnus, Magnusson, Patrik KE, & Wallace, Björn. 2012. The behavioral genetics of behavioral anomalies. *Management science*, **58**(1), 21–34.

Cressman, Ross. 1997. Local stability of smooth selection dynamics for normal form games. *Mathematical Social Sciences*, **34**(1), 1–19.

Dekel, Eddie, Ely, Jeffrey C., & Yilankaya, Okan. 2007. Evolution of Preferences. *Review of Economic Studies*, **74**(3), 685–704.

Ely, Jeffrey C. 2011. Kludged. *American Economic Journal: Microeconomics*, **3**(3), 210–231.

Eshel, Ilan, & Motro, Uzi. 1981. Kin selection and strong evolutionary stability of mutual help. *Theoretical population biology*, **19**(3), 420–433.

Eyster, Erik, & Rabin, Matthew. 2005. Cursed Equilibrium. *Econometrica*, **73**(5), 1623–1672.

Genesove, David, & Mayer, Christopher. 2001. Loss Aversion and Seller Behavior: Evidence from the Housing Market. *The Quarterly Journal of Economics*, **116**(4), 1233–1260.

Grosskopf, Brit, Bereby-Meyer, Yoella, & Bazerman, Max. 2007. On the Robustness of the Winner's Curse Phenomenon. *Theory and Decision*, **63**, 389–418.

Guth, Werner, & Yaari, Menahem. 1992. Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach. *Pages 23–34 of:* Witt, Ulrich (ed), *Explaining Process and Change: Approaches to Evolutionary Economics*. Univ. of Michigan Press, Ann Arbor.

Hammerstein, Peter. 1996. Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of mathematical biology*, **34**(5-6), 511–532.

Harrison, Glenn W., & List, John A. 2004. Field Experiments. *Journal of Economic Literature*, **42**(4), 1009–1055.

Haviland, W.A., Prins, H.E.L., Walrath, D., & McBride, B. 2007. *Anthropology: The Human Challenge*. 12, illustrated edn. Wadsworth/Thomson Learning.

Heifetz, Aviad, & Segev, Ella. 2004. The Evolutionary Role of Toughness in Bargaining. *Games and Economic Behavior*, **49**(1), 117–134.

Heifetz, Aviad, Shannon, Chris, & Spiegel, Yossi. 2007. The Dynamic Evolution of Preferences. *Economic Theory*, **32**, 251–286. 10.1007/s00199-006-0121-7.

Heller, Yuval. 2013. Three steps ahead. *https://sites.google.com/site/yuval26/3-steps.pdf*.

Herold, F., & Netzer, N. 2011. Probability Weighting as Evolutionary Second-best. *University of Zurich, Socioeconomic Institute Working Papers*.

Herskovits, Melville J. 1952. *Economic Anthropology: A Study in Comparative Economics*. New York: A.A. Knopf.

Huck, Steffen, Kirchsteiger, Georg, & Oechssler, Jorg. 2005. Learning to Like What You Have – Explaining the Endowment Effect. *The Economic Journal*, **115**(505), 689–702.

Johnson, Dominic DP, & Fowler, James H. 2011. The evolution of overconfidence. *Nature*, **477**(7364), 317–320.

Kagel, John H., & Roth, Alvin E. (eds). 1997. *The Handbook of Experimental Economics*. Princeton University Press.

Kagel, John Henry, & Levin, Dan. 2002. *Common Value Auctions and the Winner's Curse*. Princeton University Press.

Kahneman, Daniel, & Lovallo, Dan. 1993. Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. *Management Science*, **39**(1), 17–31.

Kahneman, Daniel, Knetsch, Jack L., & Thaler, Richard H. 1991. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *J. of Econ. Perspectives*, **5**(1), 193–206.

Knetsch, Jack, Tang, Fang-Fang, & Thaler, Richard. 2001. The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing? *Experimental Economics*, **4**, 257–269.

Massey, Cade, & Thaler, Richard H. 2013. The Loser's Curse: Decision Making and Market Efficiency in the National Football League Draft. *Management Science*.

Maynard Smith, John. 1971. What use is sex? *Journal of theoretical biology*, **30**(2), 319–335.

Nachbar, John H. 1990. Evolutionary selection dynamics in games: convergence and limit properties. *International journal of game theory*, **19**(1), 59–89.

Norman, Thomas WL. 2008. Dynamically stable sets in infinite strategy spaces. *Games and Economic Behavior*, **62**(2), 610–627.

Oechssler, Jörg, & Riedel, Frank. 2002. On the dynamic foundation of evolutionary stability in continuous models. *Journal of Economic Theory*, **107**(2), 223–252.

Ok, Efe A, & Vega-Redondo, Fernando. 2001. On the evolution of individualistic preferences: An incomplete information scenario. *Journal of Economic Theory*, **97**(2), 231–254.

Polanyi, Karl. 1957. The Economy as Instituted Process. *Pages 243–270 of:* Polanyi, Karl, Arensberg, Conrad M., & Pearson., Harry W. (eds), *Trade and Market in the Early Empires: Economies in History and Theory*. The Free Press.

Sahlins, Marshall David. 1972. *Stone Age Economics*. Aldine.

Sandholm, William H. 2001. Preference evolution, two-speed dynamics, and rapid social change. *Review of Economic Dynamics*, **4**(3), 637–679.

Taylor, P.D., & Jonker, L.B. 1978. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, **40**(1), 145–156.

Thaler, Richard. 1980. Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization*, **1**(1), 39–60.

Thomas, Bernhard. 1985. On evolutionarily stable sets. *J. of Math. Biology*, **22**(1), 105–115.

Waldman, Michael. 1994. Systematic Errors and the Theory of Natural Selection. *The American Economic Review*, **84**(3), pp. 482–497.

Weibull, J.W. 1997. *Evolutionary Game Theory*. The MIT press.