INFORMATION-BASED RELATIVE CONSUMPTION EFFECTS*

Larry Samuelson
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, Wisconsin 53706-1320 USA

July 26, 2001

Abstract Relative consumption effects, in which agents' preferences depend upon others' consumption, are often said to be the result of contests to secure resources that are allocated on the basis of one's status. This paper argues that Nature may induce relative consumption effects in order to compensate for incomplete environmental information. Status-based and information-based relative consumption effects can lead to quite different comparative static properties and quite different policy prescriptions.

Contents

*This paper originated in joint work with Jeroen Swinkels. I thank Jeroen, Andrew Postlewaite, John Kennan, Phil Reny, and Bill Sandholm for helpful comments and discussions. Financial support from the National Science Foundation is gratefully acknowledged.

INFORMATION-BASED RELATIVE CONSUMPTION EFFECTS

by Larry Samuelson

1 Introduction

Economists typically assume that agents' preferences depend only on their own consumption. In contrast, a variety of evidence points to the existence of relative consumption effects, in which person i's preferences regarding i's consumption depend upon person j's consumption. "Happiness" studies, in which the correlation between measures of consumption and measures of happiness are weaker than expected, are often interpreted as evidence that utility depends largely upon how one's consumption compares to that of one's peers.[1] Neumark and Postlewaite [34] find evidence consistent with relative consumption effects in the labor-supply decisions of sisters. Frank [24] offers examples of consumption that appears to be motivated by relative considerations.[2]

Why would one care about relative consumption effects? Because they can lead to inefficient allocations. The standard foundation for relative consumption effects supposes that "prizes" (such as access to good schools or mates) are allocated to those who rank high on a measure of status. If one's rank is particularly sensitive to the conspicuous consumption of certain goods (such as housing), then consumption will be inefficiently concentrated on these goods at the expense of others.[3] Frank [24] argues that such considerations lead Americans to undersave, overconsume luxury goods, and underconsume leisure and public goods.[4] Cole, Mailath and Postlewaite [14]

[1] For example, Blanchflower and Oswald [9], Easterlin [20, 21], Inglehart [28], Oswald [35], Veenhoven [41].

[2] The relative income hypothesis (Dusenberry [19], Pollak [36]) similarly explains aggregate consumption data by positing a link between desired consumption and a reference level based on previous consumption.

[3] Inefficiency is not implied by the mere existence of what Hirsch [26] called positional goods, i.e., goods in fixed supply that can be consumed only by a (perhaps high ranking) subset of an economy's agents. Competitive market prices capture the externality that agent i cannot consume such a good if agent j does.

[4] The claim that Americans underconsume leisure often begins with an argument that hours worked have increased in recent decades, fueled primarily by increased female labor supply (because "we both have to work to make ends meet"). Cox and Alm [16] argue that leisure time has actually *increased* once one accounts for the technology-induced reduction in time required for domestic production.

model the ability of status considerations to boost consumption of status-producing activities.

Concerns about relative consumption effects are heightened by the fact that income inequality has worsened in the United States in recent decades (Katz and Autor [30]). Does the resulting consumption pattern exacerbate the inefficiency caused by relative consumption effects? If so, are there appropriate policy responses?

This paper pursues an argument of Samuelson and Swinkels [39]: that relative consumption effects may have nothing to do with status-allocated goods. Instead, relative consumption effects may be a tool Nature uses to equip us with preferences that are more effectively matched to the uncertainties of our environment.

Nature must design us, through the trial-and-error process of evolution, to make decisions in a wide variety of circumstances. Giving us preferences over absolute consumption levels can be risky, causing us to exert too much effort in low-yield environments and too little effort in high-yield environments. Nature responds by designing us to condition our behavior on information gleaned from our environment. But the consumption of one's contemporaries is often an important source of environmental information. High consumption levels reflect a relatively bountiful environment in which one might be well advised to increase one's own consumption. Low consumption levels reflect meager environments in which fanatical attempts at consumption are ill advised. Relative consumption effects, in the form of a tendency to push consumption levels in the direction of those observed in one's contemporaries, thus make optimal use of scarce information.

In the absence of constraints, Nature could induce such relative consumption effects by coupling ordinary utility functions, defined only over one's own consumption, with a sufficiently sophisticated ability to process information. Information would then be used optimally. The policy implications of relative consumption effects could consist entirely of measures to ensure that information is also produced optimally.

Unfortunately, it is beyond our capabilities to process all of the useful information we receive. As a result, Nature must make some compromises in designing our preferences and information-processing rules. Samuelson and Swinkels [39] examine Nature's ability to mitigate information problems with utility functions that seem anomalous from a classical point of view. Building relative consumption effects directly into preferences is one of these complexity-cutting shortcuts, substituting an approximate but simple decision rule for a more precise but complicated inference problem.

Two risks arise in working with such preferences. For Nature, there is the danger that her agents will sometimes react to others' consumption levels without good informational grounds for doing so. Nature must balance this danger against the complexity costs of more precise information processing. In the course of doing so, relative consumption effects spill into utility functions, giving rise to policy implications that go beyond the mere production and use of information. For an analyst modeling the preferences, there is the danger of building an ever-increasing menagerie of arbitrary effects or "just so stories" into preferences, a new one for every behavioral anomaly (cf. Postlewaite [37]). A useful safeguard against abuse in this respect is to provide an evolutionary foundation for such preference effects, as we do in Sections 2–3.

This paper concentrates on information-driven relative consumption effects. Status is undoubtedly also important in resource allocation. But status and information can have quite different implications that can be untangled only by carefully modeling each.

Section 2 examines a simple model of the problem Nature faces when designing agents to live in a fluctuating environment. Section 3 derives Nature's optimal solution, notes that information-processing constraints may prompt Nature to implement this solution by building relative consumption effects into the utility function, and presents an example. Section 4 contrasts the implications of information-based relative consumption effects with those arising out of concerns for status. Concentrating the distribution of productivities alleviates relative-consumption distortions in the first case, but amplifies them in the second. Section 5 concludes with a very brief discussion of how information-based and status-based relative consumption effects might be empirically distinguished.

2 Information and Consumption

This section presents a model in which agents' decisions are guided by inferences they draw from observations of others' decisions. Beginning with the work of Banerjee [5] and Bikhchandani, Hirshleifer and Welch [7], a variety of herding models have captured similar considerations. In a typical herding model, each agent uses Bayes' rule to assess the signal he has observed directly and the information contained in his observations of his predecessors' decisions. Crucial to the latter assessment is the agent's ability to infer the Bayesian decision problems faced by each of his predecessors.

In our case, this updating process is complicated by two features of Nature's design problem. First, agents in our model observe their predecessors through the filter of natural selection. Agents who have chosen strategies well-suited to the current environment are more likely to survive than are others, biasing the mix of survivors whose decisions can be observed by subsequent agents. An agent's observed behavior thus mixes clues about the agent's information with clues about his evolutionary experience, both of which enter the observer's inference problem. The incorporation of the latter effect makes the problem more similar to that of Banerjee and Fudenberg [6] and Ellison and Fudenberg [22, 23] than to herding models.

Secondly, the state of the environment itself fluctuates in our model, forcing the agents to draw inferences about a moving target. This disrupts the simple recursive structure typical of herding models. Instead, we examine a stationary equilibrium in which the optimal decision rule shapes the inference and decision problems faced by one's predecessors, which in turn affects the current inference problem and hence the optimal decision rule. A sequence of recursive optimization problems is thus replaced by a fixed-point problem.

2.1 The Environment

Let time be divided into discrete periods. At the beginning of each period t, the environment is characterized by a variable $\theta_t \in \{\underline{\theta}, \overline{\theta}\}$. The events within a period proceed as follows:

1. Each member of a continuum of surviving agents gives birth, with each surviving agent giving birth to the same, exogenously fixed number of offspring. Each offspring is characterized by a parameter ϵ. The realized values of ϵ are uniformly distributed on $[0, 1]$. We interpret each offspring as (not necessarily independently) drawing ϵ from a uniform distribution.

2. Each newly-born agent observes n randomly selected surviving agents from the previous generation, discerning whether each chose $\underline{z}$ or $\overline{z}$, and also observes a signal $\xi \in \mathbb{R}$ drawn from a distribution $G(\xi; \theta_t)$ with density $g(\xi; \theta_t)$.

3. All parents then die. Each member of the new generation chooses a consumption strategy $z \in \{\underline{z}, \overline{z}\}$.

4. Nature then conducts survival lotteries, where $h(z; \epsilon; \theta) : \{\underline{z}, \overline{z}\} \times [0, 1] \times \{\underline{\theta}, \overline{\theta}\} \to [0, 1]$ gives the probability that an agent with strategy z and characteristic ϵ survives when the state of the environment is θ. Again, we assume no aggregate uncertainty.

5. Nature draws a value $\theta_{t+1} \in \{\underline{\theta}, \bar{\theta}\}$.

We interpret $\underline{z}$ and $\bar{z}$ as denoting low-consumption and high-consumption strategies. It may appear as if higher consumption levels, by reducing the chance of falling below a starvation barrier, must be preferred in a biological context (with free disposal obviating the danger of excess consumption). However, the high-consumption strategy may demand more effort and energy and may expose the agent to greater danger, opening the possibility that the low-consumption alternative is superior.

We allow the ranking of these strategies to depend upon both individual characteristics and the state of the environment. Some agents may be better-endowed with the skills that reduce the risk of procuring consumption than others. Some environments may feature more plentiful and less risky consumption opportunities than others. These effects appear in the specification of the survival probabilities $h(z;\varepsilon;\theta)$, given by

$$h(\underline{z};\varepsilon;\theta) = \tfrac{1}{2}$$

$$h(\bar{z};\varepsilon;\bar{\theta}) = \tfrac{1}{2} + b(\varepsilon - q) \qquad (1)$$

$$h(\bar{z};\varepsilon;\underline{\theta}) = \tfrac{1}{2} + b(\varepsilon - (1 - q)); \qquad (2)$$

where $0 < q < 1/2$. Figure 1 depicts these probabilities. The low-consumption strategy $\underline{z}$ yields a survival probability of $\frac{1}{2}$, regardless of the agent's characteristic or state of the environment. The high-consumption strategy $\bar{z}$ yields a higher survival probability for agents with higher values of ε and yields a higher survival probability when the state is $\bar{\theta}$. To ensure the probabilities are well defined, we assume

$$0 < b < 1/(2(1 - q)):$$

The environmental parameter θ follows a Markov process, with transition rule:

	$\bar{\theta}$	$\underline{\theta}$
$\bar{\theta}$	$1-\tau$	τ
$\underline{\theta}$	τ	$1-\tau$

(3)

where

$$\tau < \tfrac{1}{2};$$

so that next period's state is most likely to match this period's.[5]

[5] The symmetry of this Markov process simplifies the calculations, but is not crucial for the results.

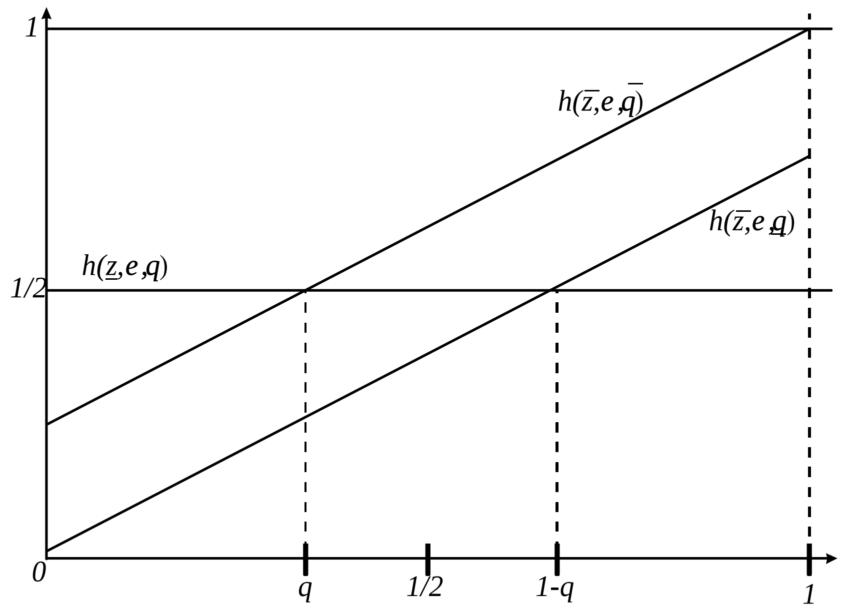

Figure 1: Survival probabilities

We assume that the density $g(\xi;\theta)$ is continuous and strictly positive on $\mathbb{R}$, for both $\underline{\theta}$ and $\bar{\theta}$, and satisfies the monotone likelihood ratio property (Milgrom [33]):

$$\xi'' > \xi' \implies \frac{g(\xi'';\bar{\theta})}{g(\xi'';\underline{\theta})} > \frac{g(\xi';\bar{\theta})}{g(\xi';\underline{\theta})}, \qquad (4)$$

so that higher values of ξ lead the agent to believe that state $\bar{\theta}$ is more likely. We assume that "proportion" $G(\xi';\theta)$ of the agents characterized by each value of ε draw values of $\xi \le \xi'$.[6]

2.2 Full-Information Optimum

We assume throughout that each agent knows the value of his characteristic ε. Much then hinges on what the agents and Nature can observe concerning the state θ.

If the agent can observe θ (i.e., $G(\xi;\theta)$ is perfectly informative), then observations of the previous generation are irrelevant and the optimal decision

[6] Again, we assume there is no aggregate uncertainty. In each case, the fact that we do not require independence could be exploited to construct an explicit measure-theoretic foundation.

is straightforward, given by:[7]

$$\bar{z}\quad(\ )\qquad \begin{cases} \mu = \bar{\mu} \text{ and } {}^{2} > \ q \\ \mu = \underline{\mu} \text{ and } {}^{2} > 1 \ ¡ \ q \end{cases} :$$

If $\underline{\mu}$ denotes winter and $\bar{\mu}$ summer, then the environmental characteristic seems su±ciently unambiguous as to be observed by the agents. Even here, however, mistakes can occur. Plants can produce buds at an inopportune time, sometimes disastrously, on the strength of an uncharacteristically warm spell. Animals can make ill-advised migration and hibernation decisions. People can plant or harvest crops too early or too late. Hence, even with something so obvious as the season, we can expect agents' abilities to observe the state to fall short of the desired precision.

More generally, the environmental variable may be something less obvious than climate. For example, $\underline{\mu}$ and $\bar{\mu}$ may index the relative availability of edible plants and slow-moving animals, which in turn may move in noisy cycles, leaving the agents with quite imprecise beliefs about the state of the environment.

This uncertainty on the part of the agents is no problem if Nature can observe the state of the environment. But Nature \observes" the state of the environment only by allowing the trail-and-error process of evolution to produce a decision rule that is optimal given the state, and hence can only observe states that are stationary for long periods of time. If the relevant information contained in $\underline{\mu}$ and $\bar{\mu}$ is only whether the edible animals in one's environment are mostly dinosaurs or mostly mammals, then the state changes slowly enough that Nature can simply count on adapting her agents to the environment, equipping them with the $\underline{\mu}$ or $\bar{\mu}$ decision rule (and causing them to ignore ») as relevant. There will be some disruption whenever the state changes, but such transitions are su±ciently infrequent and idiosyncratic that Nature cannot endow her agents with the capability to respond.

We are interested in cases in which the state µ exhibits °uctuations which are not perfectly observed by the agents (so that G(»;µ) is nontrivial) and which are su±ciently transitory that Nature cannot observe them. We assume that the process generating the state, given by (3), remains constant for a su±ciently long time as to be observed by Nature.[8]

[7] Throughout, we ignore measure-zero cases in which decisions are arbitrary.

[8] If not, then we would embed the current process in a larger one which generates short-term transition matrices like (3) according to a rule that is sufficiently stationary as to be observed by Nature.

2.3 No-Information Optimum

Suppose $n = 0$ (no members of the previous generation can be observed) and the agents know only their characteristic ϵ and signal ξ and that the state of the environment is generated by a Markov process given by (3). The stationary distribution of this Markov process attaches a probability of $\frac{1}{2}$ to each of the states $\underline{\theta}$ and $\overline{\theta}$. Hence, the probability that the state is $\overline{\theta}$, conditional on having observed signal ξ, is

$$\rho(\overline{\theta}|\xi) = \frac{\frac{1}{2}g(\xi;\overline{\theta})}{\frac{1}{2}g(\xi;\overline{\theta}) + \frac{1}{2}g(\xi;\underline{\theta})}.$$

An agent with no additional information about the state maximizes his probability of withstanding the survival lottery he faces by choosing

$$\overline{z} \quad (\ ) \quad \rho(\overline{\theta}|\xi)h(\overline{z};\epsilon;\overline{\theta}) + (1 - \rho(\overline{\theta}|\xi))h(\overline{z};\epsilon;\underline{\theta}) > \frac{1}{2}. \qquad (5)$$

Using (1)–(2) to fill in the survival probabilities, this inequality becomes

$$\overline{z} \quad (\ ) \quad \epsilon > \rho(\overline{\theta}|\xi)q + (1 - \rho(\overline{\theta}|\xi))(1 - q).$$

For each posterior probability $\rho(\overline{\theta}|\xi) = \rho$ there is thus a critical characteristic $\epsilon^*(\rho) = \rho q + (1-\rho)(1-q)$, with larger values of ϵ prompting a choice of $\overline{z}$ and lower values prompting $\underline{z}$. The function $\epsilon^*(\rho)$ is decreasing in ρ, meaning that the agent is more likely to choose $\overline{z}$ when the prior probability of state $\overline{\theta}$ is larger. Notice that $\epsilon^*(1) = q$ and $\epsilon^*(0) = 1 - q$, so that the agent's actions coincide with the full-information optimum when the prior is degenerate.

The decision rule embedded in $\epsilon^*(\rho)$ maximizes the probability that an agent survives a single period. However, evolution will select for the decision rule that conditions the agent's behavior on his posterior ρ (only, since it is a sufficient statistic for the problem facing the agent) so as to choose[9]

[9] The evolutionarily most successful strategy in this setting will (as nearly as possible) maximize the probability of surviving for T periods, for all sufficiently large values of T. The probability of surviving for T periods is given by $\sum_{k=0}^{T} \overline{p}(\epsilon^*)^k \underline{p}(\epsilon^*)^{T-k} \text{prob}_T(k)$, where $\text{prob}_T(k)$ is the probability that, over the course of T periods, exactly k of these periods are characterized by state $\overline{\theta}$ and $T-k$ by state $\underline{\theta}$. As T gets large, this maximization problem converges to (7), in the sense that the maximizer of (7) comes arbitrarily close to maximizing the T-period survival probability for all large T, and hence evolution will select for the maximizer of (7). (The key to establishing this equivalence is to note that over long periods of time, the proportion of $\overline{\theta}$ states is very nearly $\rho(\overline{\theta}|\xi)$. The probability of survival until (large) period T is then very nearly $\overline{p}(\epsilon^*)^{\rho(\overline{\theta}|\xi)}(1 - \overline{p}(\epsilon^*))^{\rho(\underline{\theta}|\xi)}$, which is equivalent to (7).) The equivalence of (7) and (6) then follows from noting that (6) is necessary and sufficient (excepting deviations on sets of measure zero) for maximizing (7). See Robson [38]. Results of this type are standard in evolutionary biology (Charlesworth [11, pp. 47–50], Tuljapurkar [40]).

$$\bar z \quad (\;) \quad \Big(\rho(\bar\theta|\xi) \ln h(\bar z, \epsilon, \bar\theta) + (1 - \rho(\bar\theta|\xi)) \ln h(\bar z, \epsilon, \underline\theta) >$$

$$\rho(\bar\theta|\xi) \ln h(\underline z, \epsilon, \bar\theta) + (1 - \rho(\bar\theta|\xi)) \ln h(\underline z, \epsilon, \underline\theta) \Big) \equiv \ln\tfrac{1}{2}. \quad (6)$$

Equivalently, let $p(\epsilon^*, \bar\theta)$ be the expected probability that an agent carrying a gene inducing decision cutoff $\epsilon^*(\rho)$ survives Nature's survival lottery in a period in which the state of the environment is $\bar\theta$, with the expectation taken over the value of ϵ. Let $p(\epsilon^*, \underline\theta)$ be similarly defined for state $\underline\theta$. Then (6) is equivalent to choosing the rule $\epsilon^*(\rho)$ to maximize:

$$\rho(\bar\theta|\xi) \ln p(\epsilon^*, \bar\theta) + (1 - \rho(\bar\theta|\xi)) \ln p(\epsilon^*, \underline\theta). \quad (7)$$

In contrast, the decision rule given by (5) maximizes the average survival probability $\rho(\bar\theta|\xi) p(\epsilon^*, \bar\theta) + (1 - \rho(\bar\theta|\xi)) p(\epsilon^*, \underline\theta)$. To see the difference, consider an environment in which state $\underline\theta$ is quite unlikely. Maximizing the probability of surviving a single period (as in (5)) may then call for a strategy that entails a zero survival probability in state $\underline\theta$ (in return for a higher survival probability in the much more likely state $\bar\theta$). But over a longer period of time, this strategy is a sure recipe for extinction. A superior strategy increases the survival probability in state $\underline\theta$ at the expense of state $\bar\theta$, as does the maximizer of (7).

When $\rho = \frac{1}{2}$, (5) will set $\epsilon^* = \frac{1}{2}$. In contrast, (7) calls for a value of ϵ^* that exceeds $\frac{1}{2}$. In particular, the functions $p(\epsilon^*, \bar\theta)$ and $p(\epsilon^*, \underline\theta)$ are given by:

$$p(\epsilon^*, \bar\theta) = \int_0^{\epsilon^*} \frac{1}{2} d\epsilon + \int_{\epsilon^*}^1 h(\bar z, \epsilon, \bar\theta) d\epsilon$$

$$p(\epsilon^*, \underline\theta) = \int_0^{\epsilon^*} \frac{1}{2} d\epsilon + \int_{\epsilon^*}^1 h(\bar z, \epsilon, \underline\theta) d\epsilon.$$

These functions are strictly concave and have slopes of equal absolute value but opposite sign when $\epsilon^* = \frac{1}{2}$.[10] To maximize the product $\frac{1}{2} \ln p(\epsilon^*, \bar\theta) + \frac{1}{2} \ln p(\epsilon^*, \underline\theta)$, one then increases ϵ^* above $\frac{1}{2}$, trading off a decrease in $p(\epsilon^*, \bar\theta)$ for a smaller increase in the value of $p(\epsilon^*, \underline\theta)$. Because $p(\epsilon^*, \underline\theta)$ is relatively small, this increases the sum $\frac{1}{2} \ln p(\epsilon^*, \bar\theta) + \frac{1}{2} \ln p(\epsilon^*, \underline\theta)$ and hence increases the long-term survival probability.

[10] The slope characterization follows immediately from the observation that $h(\bar z, \frac{1}{2}, \bar\theta) - \frac{1}{2} = \frac{1}{2} - h(\bar z, \frac{1}{2}, \underline\theta)$, as shown in Figure 1. It is this property which ensures that setting $\epsilon^* = \frac{1}{2}$ maximizes the single-period survival probability $\frac{1}{2}(p(\epsilon^*, \bar\theta) + p(\epsilon^*, \underline\theta))$.

As ρ increases, the optimal value ε^* decreases and $p(\varepsilon^*; \overline{\theta})$ increases while $p(\varepsilon^*; \underline{\theta})$ decreases. The investment in survival probability is thus shifted toward the more likely state.

3 Relative consumption effects

3.1 Observable-Consumption Optimum

We now turn to the case in which agents can infer information about θ from the noisy signal provided by observing n (> 0) agents from the preceding generation (as well as from the signal ξ). If the state affects the relative payoff of a high-consumption hunting strategy, for example, then an agent may be able to infer information about θ by watching whether his surviving neighbors appear to be mostly those who hunt or those who plant.

Nature's optimum incorporates the agent's information into his decision rule.[11] A decision rule for an agent is now a collection of functions $\{\varepsilon(n,\xi), \varepsilon(n-1,\xi), \ldots, \varepsilon(0,\xi)\}$, where $\varepsilon(k,\xi)$ is the cutoff value of ε for choosing $\overline{z}$ ($\overline{z}$ being chosen for larger values of ε) when the agent has observed k cases of $\overline{z}$ and $n-k$ cases of $\underline{z}$ in the previous generation and has observed signal ξ. Figure 2 illustrates such a strategy for the case in which $n = 1$. We let $\mathcal{E}$ denote a collection $\{\varepsilon(n,\xi), \ldots, \varepsilon(0,\xi)\}$.

It is clear that it can never be optimal for $\varepsilon(k,\xi)$ to lie outside the bounds $[q, 1-q]$. Our intuition is that higher values of ξ and more observations of $\overline{z}$ (rather than $\underline{z}$) will indicate that state $\overline{\theta}$ is more likely, making the agent more inclined to choose $\overline{z}$. Hence, we expect equilibrium decision rules to be at least weakly decreasing in ξ and to satisfy, for all ξ,

$$q \le \varepsilon(n,\xi) \le \cdots \le \varepsilon(0,\xi) \le 1-q.$$

We refer to a collection of functions $\{\varepsilon(n,\xi), \ldots, \varepsilon(0,\xi)\}$ that satisfies these inequalities as being admissible.

Let ψ_t be the proportion of strategy $\overline{z}$ among those agents who survived period $t-1$. Then ψ_t describes the distribution from which period-t agents draw their observations of z, with each period-t new-born observing $\overline{z}$ on each draw with probability ψ_t and observing $\underline{z}$ with probability $1-\psi_t$. Let $\gamma_{\mathcal{E}}(\psi_t, \theta_t)$ be the proportion of surviving period-t agents who chose $\overline{z}$, given that (i) these agents, as new-borns, drew observations from the distribution described by ψ_t, (ii) the period-t state of the environment relevant for

[11] Any decision rule or mix of rules that does not make use of this information, whether tailored to state $\overline{\theta}$, $\underline{\theta}$, or some combination of the two, is evolutionarily dominated by a single rule that responds optimally to the information.

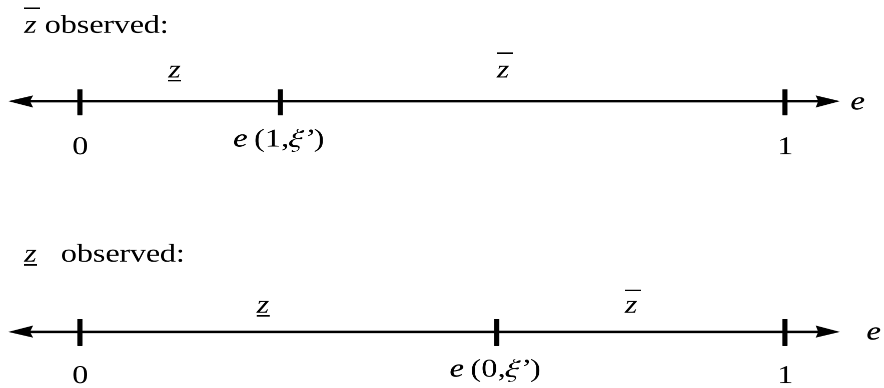

Figure 2: Admissible decision rule $(\epsilon(1;\xi);\epsilon(0;\xi))$ when $n = 1$, conditional on a value $\xi = \xi'$

Nature's survival lotteries is θ_t, and (iii) the agents' decision rule is given by E. Given that Nature first runs the period-t survival lotteries and then determines the period-(t+1) state, we can describe our system as a Markov process $(\psi_t;\theta_t)$ defined on the state space $[0;1]\times\{\underline{\theta};\bar{\theta}\}$. Letting $\Theta(\theta)$ denote the transition rule given by (3), $(\Phi_{\mathcal{E}};\Theta)$ denotes the transition rule for the process $(\psi_t;\theta_t)$:

$$\begin{aligned} \psi_{t+1} &= \Phi_{\mathcal{E}}(\psi_t;\theta_t) \\ \theta_{t+1} &= \Theta(\theta_t). \end{aligned}$$

Let $\mu_{\mathcal{E}}((\psi_t;\theta_t);(\psi_0;\theta_0))$ be the probability distribution induced by the transition rule $(\Phi_{\mathcal{E}};\Theta)$ in period t given initial condition $(\psi_0;\theta_0)$. Then define, for $k = 0;\ldots;n$ (suppressing the dependence of Φ on E, the dependence on the initial condition, and some time subscripts on θ)

$$\rho_{\mathcal{E}}(\bar{\theta}_t|k;\xi;t) = \frac{\int_{\psi_t=0}^{1} g(\xi;\bar{\theta})r(k;\psi_t)d\mu(\psi_t;\bar{\theta})}{\int_{\psi_t=0}^{1} g(\xi;\bar{\theta})r(k;\psi_t)d\mu(\psi_t;\bar{\theta}) + \int_{\psi_t=0}^{1} g(\xi;\underline{\theta})r(k;\psi_t)d\mu(\psi_t;\underline{\theta})} \quad (8)$$

where

$$r(k;\psi_t) = \binom{n}{k}\psi_t^k(1-\psi_t)^{n-k}.$$

The function $\rho_{\mathcal{E}}(\bar{\theta}_t|k;t)$ thus gives the probability that the state in time t is $\bar{\theta}$, given a time-t observation of signal ξ and k values of $\bar{\theta}$.[12] This

[12] The numerator of the right side is the joint probability that the state was $\bar{\theta}$ and the

probability reflects two considerations—the extent to which an observation of ξ or $\bar z$ indicates that the previous-period state was relatively favorable for ξ or strategy $\bar z$ (i.e., was $\bar\theta$), and the probability that the state may have changed since the previous period.

Our first observation establishes conditions under which these probabilities have well-behaved limits:

Lemma 1 There exists a value $q^* \in (0;\frac{1}{2})$ such that for any $q \in (q^*;\frac{1}{2})$ and any admissible E, there exist probabilities $\rho_{\mathcal{E}}(\bar\theta|k;\xi)$ $(k = 0;\ldots;n)$ satisfying, for all initial conditions,

$$\lim_{t\to\infty} \rho_{\mathcal{E}}(\bar\theta_t|k;\xi;t) = \rho_{\mathcal{E}}(\bar\theta|k;\xi):$$

In addition, the $\rho_{\mathcal{E}}(\bar\theta|k;\xi)$ are strictly increasing in ξ and satisfy $\rho_{\mathcal{E}}(\bar\theta|k+1;\xi) > \rho_{\mathcal{E}}(\bar\theta|k;\xi)$.

The restriction that $q > q^*$ ensures that the population can never get too heavily concentrated on a single consumption strategy, either $\bar z$ or $\underline{z}$. This in turn ensures that changes in the state of the environment are reflected relatively quickly in the observed distribution of consumption strategies, and hence that the latter is informative. To see how this could fail, consider the extreme case of $q = 0$. In this environment, there is no value of ε for which $\bar z$ is a dominant strategy. As a result, it is possible that virtually the entire population chooses $\underline{z}$. A change from state $\underline\theta$ to $\bar\theta$ will then not produce a noticeable change in the distribution of consumption strategies for an extraordinarily long time, causing this distribution to be relatively uninformative.[13]

The inequality $\rho_{\mathcal{E}}(\bar\theta|k+1;\xi) > \rho_{\mathcal{E}}(\bar\theta|k;\xi)$ indicates that observations of high consumption enhance the posterior probability that the state of the environment is $\bar\theta$. This is the foundation of relative consumption effects.

The obvious approach to proving Lemma 1 is to show that the Markov process $(\gamma_{\mathcal{E}};\Theta)$ has an ergodic measure $\pi^*_{\mathcal{E}}(\psi;\theta;\xi)$ which can be inserted in (8) to calculate the probabilities $\rho_{\mathcal{E}}(\bar\theta|k;\xi)$. However, the Markov process

agent drew ξ and k observations of $\bar z$, while the denominator is the probability of drawing ξ and k observations of $\bar z$.

[13] If $q = 0$, our model would also be sensitive to the specification of the Markov process governing the state θ. If this process were asymmetric, then the possibility would arise that the population could be absorbed in a state in which all agents always choose either $\bar z$ or $\underline{z}$. A similar possibility arises in Ellison and Fudenberg [23].

$(\psi_{\mathcal{E}}, \Theta)$ is not irreducible, foiling the standard route to ergodicity.[14] The proof constructs a candidate measure μ^* and establishes an appropriate convergence result.

Definition 1 The functions $E = \{\epsilon(n, \xi), \ldots, \epsilon(0, \xi)\}$ are an equilibrium if $\epsilon(k, \xi)$ solves, for each value of k and ξ (cf. (7)),

$$\max_{\epsilon} \text{prob}(\bar{\theta} \mid k, \xi) \ln p(\epsilon, \bar{\theta}) + (1 - \text{prob}(\bar{\theta} \mid k, \xi)) \ln p(\epsilon, \underline{\theta}) \tag{9}$$

where

$$\text{prob}(\bar{\theta} \mid k, \xi) = \rho_{\mathcal{E}}(\bar{\theta} \mid k, \xi). \tag{10}$$

In defining an equilibrium, we use the limiting probabilities $\rho_{\mathcal{E}}(\bar{\theta} \mid k, \xi)$ to evaluate the payoff of a strategy. This reflects two timing assumptions. First, the state of the environment changes rapidly relative to the frequency with which Nature can respond via mutation and selection, forcing Nature to rely on agents' observations of others' consumption to infer information about the environment. Second, the process governing the state of the environment persists for a sufficiently long time that (i) Nature can adapt her agents to this process, in particular tuning their information-updating rules to this process, and (ii) the limiting probabilities $\rho_{\mathcal{E}}(\bar{\theta} \mid k, \xi)$ are useful approximations of the information-updating problem facing the agents.

Proposition 1 Let $q \in (q^*, \frac{1}{2})$. Than an equilibrium with admissible functions $\{\epsilon^*(n, \xi), \ldots, \epsilon^*(0, \xi)\}$ exists. In any such equilibrium, $\epsilon^*(k, \xi)$ is strictly decreasing in ξ and satisfies $\epsilon^*(k+1, \xi) < \epsilon^*(k, \xi)$.

Figure 3 depicts the resulting equilibrium. Agents are more likely to choose high consumption, i.e., choose $\bar{z}$ for a wider range of ϵ, when k and ξ are large. Observations of high consumption, by increasing the expectation that the environment is in a state favorable to high consumption, increase an agent's propensity to choose high consumption. Relative consumption thus becomes useful as a response to environmental information that neither the agents nor Nature can observe.[15]

---

[14] In particular, ψ_t is drawn from the uncountable set $[0, 1]$, while fixing a sequence of values of θ and iterating the function $\Psi_{\mathcal{E}}(\psi, \theta)$ can generate only a countable set of values of ψ (though different initial conditions lead to different sets of such values). If we set $\tau = \frac{1}{2}$, sacrificing the (essential, for our purposes) persistence of the process Θ, then the period-t state would be independent of its predecessor and we could appeal to Futia's results on random contractions [25, Defs. 5.1, 6.2] to establish ergodicity.

[15] Given our restriction to admissible $\mathcal{E}$ (e.g., to cases in which $\epsilon(k+1, \xi) \leq \epsilon(k, \xi)$), showing that consumption observations are informative entails only precluding equality. We could strengthen our results to show that there is no equilibrium in which $\epsilon(k+1, \xi) > \epsilon(k, \xi)$, but the argument is tedious.

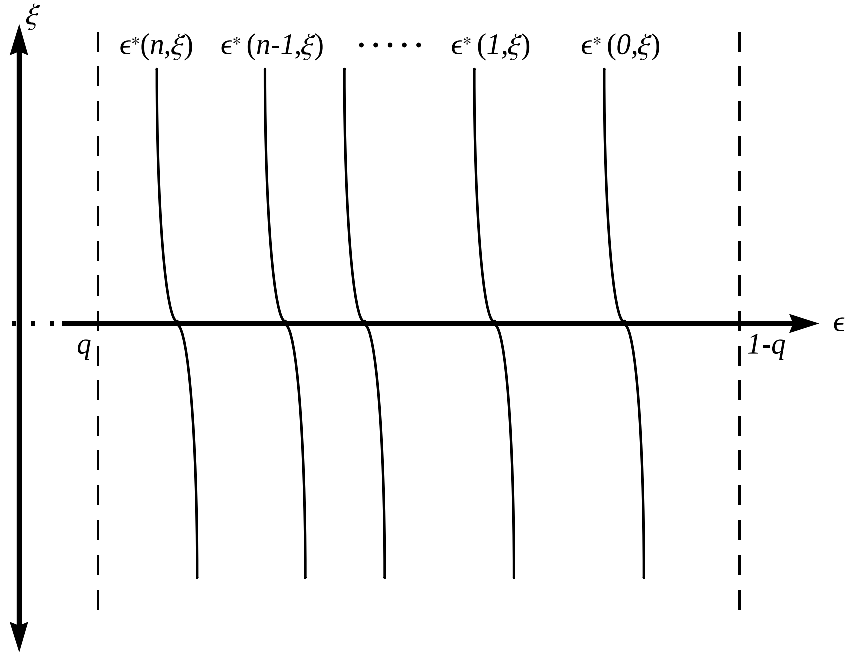

Figure 3: Equilibrium consumption strategies

It is important to note that an agent's survival in our model depends only on the agent's own consumption. The route to genetic success is to choose optimal consumption levels, regardless of others' choices. The consumption levels of others are relevant (only) because they serve as valuable indicators of the state of the environment.

3.2 Imperfect Information Processing

The agents in our model optimally react to the uncertainty in their environment by being more likely to choose high consumption strategies when they observe others who have chosen high consumption. Nature can ensure such behavior by equipping her agents with an understanding of Bayes' rule and their environment and with a utility function over decision cutoffs ϵ^* given by

$$u(\epsilon^*;\theta) = \ln p(\epsilon^*;\theta).$$

An agent's expected-utility maximization problem, conditional on an observation $(k;\xi)$, is then given by the counterpart of (7):

$$\max_{\epsilon^*} \rho_{\mathcal{E}}(\theta|k;\xi)\ln p(\epsilon^*;\bar{\theta}) + (1-\rho_{\mathcal{E}})(\theta|k;\xi)\ln p(\epsilon^*;\underline{\theta}). \qquad (11)$$

Information is then used optimally, but is not produced optimally. Instead, agents make decisions concerning the production of information that ignore the value of this information to others.[16] As a result, policy interventions designed to increase the production of information can yield Pareto superior outcomes. However, an appropriate policy could consist solely of providing information. If the resulting information was a sufficient statistic for observation of others' consumption, then relative consumption effects would disappear.

Unfortunately for Nature, she faces formidable constraints on her ability to design agents who process information perfectly. Cognitive resources are costly (Clark [13, Chapter 4]). Nature cannot ensure we have sufficient resources to flawlessly process the information contained in our environment, instead being forced to work with agents whose information processing is noisy. To capture this, we expand the model by assuming that, when calculating the posterior probability of state $\bar{\theta}$, there is probability $\lambda \in (0,1)$ that the agent mistakenly conditions on an uninformative signal ζ rather than ξ. Hence, the agent's posterior belief, now denoted $\hat{\rho}_{\mathcal{E}}(\theta|k;\xi;\zeta)$, is given by

$$\hat{\rho}_{\mathcal{E}}(\theta|k;\xi;\zeta) = \begin{cases} \rho_{\mathcal{E}}(\theta|k;\xi) & \text{with probability } 1-\lambda \\ \rho_{\mathcal{E}}(\theta|k;\zeta) & \text{with probability } \lambda \end{cases};$$

where the function ρ is given by (8). The agent calculates the posterior correctly with probability $1-\lambda \in (0,1)$. A miscalculating agent not only fails to recognize when his information is based on the signal ζ, but fails to recognize the possibility of conditioning on such an uninformative signal (hence continuing to apply (8)). This is the essence of the agent's imperfect information processing.

We let $f(\zeta)$ denote the density of ζ. We assume this density is positive on $\mathbb{R}$ and does not depend upon the state θ, ensuring that ζ is uninformative. However, we ensure that ζ is not too idiosyncratically distributed by assuming

$$s'' > s' \implies \begin{cases} \frac{g(s'',\bar{\theta})}{f(s'')} > \frac{g(s',\bar{\theta})}{f(s)} \\ \frac{f(s'')}{g(\xi'',\underline{\theta})} > \frac{f(s)}{g(\xi',\underline{\theta})} \end{cases}; \tag{12}$$

so that large signals (for example) are more likely to come from the informative distribution (given state $\bar{\theta}$) than the uninformative distribution. The

[16] An efficient solution would spread the values of $\{\epsilon(n,\xi),\ldots,\epsilon(0,\xi)\}$ further apart, trading second-order losses in individual optimization for the production of more information. In the absence of group selection, the restriction of Nature's design mechanism to individual survival puts such a solution beyond her grasp.

value of η and whether $\rho_{\mathcal{E}}(\theta|k;\xi)$ or $\rho_{\mathcal{E}}(\theta|k;\eta)$ is the applicable posterior are drawn independently of one another and the other random variables in the model.

Inserting the posterior $\hat{\rho}_{\mathcal{E}}(\theta|k;\xi;\eta)$ into (11), the agent's maximization problem is now

$$\max_{\epsilon^*} \hat{\rho}_{\mathcal{E}}(\theta|k;\xi;\eta)\ln p(\epsilon^*;\bar{\theta}) + (1-\hat{\rho}_{\mathcal{E}}(\theta|k;\xi;\eta))\ln p(\epsilon^*;\underline{\theta}). \qquad (13)$$

The noise captured by η pushes the agent's values of $\epsilon(k;\xi)$, away from their optimal values, so that information is no longer used optimally. Nature can always improve these choices by enhancing the agent's information-processing ability, which corresponds to decreasing λ in this context, but eventually finds the opportunity cost of further improvements prohibitive.[17] An effective policy response to the resulting relative consumption effects now calls for the provision of information as well as the provision of assistance in processing information, with appropriately provided and processed information again banishing relative consumption effects and their attendant inefficiencies.

Given that she must work with the noisy beliefs $\hat{\rho}$ rather than ρ, however, Nature will no longer find it optimal to let the utility function be given by $u(\epsilon^*;\theta) = \ln p(\epsilon^*;\theta)$ (as in (13)). Instead, Nature can improve the agent's performance by building the observation k into the utility function. Intuitively, she uses "distortions" of the utility function to correct for the noise in the agent's information processing.

Let $\phi(\bar{\theta}|\epsilon^*)$ be the probability of state $\bar{\theta}$ that causes ϵ^* to solve the maximization problem (11) (or equivalently, (13)). Then:

Proposition 2 *Nature can achieve her optimum, given that the agents' information updating is given by $\hat{\rho}$, with a utility function $u(\epsilon^*;k;\theta)$ for which, for each $k = 0,\ldots,n$, there is a value $\Upsilon(k)$ such that*

$$\left(\phi(\bar{\theta}|\epsilon^*)\frac{du(\epsilon^*;k;\bar{\theta})}{d\epsilon^*} + (1-\phi(\bar{\theta}|\epsilon^*))\frac{du(\epsilon^*;k;\underline{\theta})}{d\epsilon^*}\right.$$

[17] Among other difficulties, a perfect ability to process information runs the risk of being too cumbersome to be useful. We lack the time to analyze every piece of information. Nature again responds by building reactions to common information-processing and decision problems into our behavior. If Nature's agents habitually calculate that it is optimal to hedge consumption levels in the direction of those observed in their peers, then Nature will dispense with the calculation and embed relative considerations into the utility function. LeDoux [31] discusses the incentives for Nature to arm us with a mix of "hard-wired" and cognitive responses to our environment, arguing that many of our seemingly hard-wired reactions are engineered to ensure that we do not tarry to process information before acting.

$$> (<) \left(\phi(\bar{\theta}|\epsilon^*) \frac{d \ln p(\epsilon^*; \bar{\theta})}{d\epsilon^*} + (1 - \phi(\bar{\theta}|\epsilon^*)) \frac{d \ln p(\epsilon^*; \underline{\theta})}{d\epsilon^*} \right)$$

$$\iff \epsilon^* < (>) \hat{\epsilon}(k). \tag{14}$$

Recognizing that both ξ and ζ play a role in shaping the agent's beliefs, Nature regards the agent's posterior beliefs as being less informative than does the agent. She responds by adjusting the utility function to reduce the sensitivity of the agent's actions to his beliefs.[18] She does so by designing the utility function to discourage the agent from straying too far from an action $\hat{\epsilon}(k)$ (for each observation k), increasing (decreasing) the derivative of expected utility (relative to the utility function $\ln p(\epsilon^*; k)$) for smaller (larger) actions. Hence, Nature requires that the agent observe more persuasive information before straying too far from a target consumption plan, where the latter depends upon observations of others' consumption. The key observation is that Nature's manipulation of the utility function, and hence the agent's utility, depends upon the value of k. Relative consumption effects now take the form of incorporating others' consumption levels directly into the utility function.

Nature's problem would be particularly simple if ξ were uninformative. In this case, the agents have no useful information beyond that which is conveyed by the number of $\hat{\epsilon}$ observations k, upon which Nature can condition utilities. She could then abandon all thoughts of allowing her agents to process information, having them maximize a utility function that depends upon k and that induces the optimal strategy $(\epsilon^*(n); \ldots; \epsilon^*(0))$. Nature is pushing her agents toward this outcome when designing the utility function u of Proposition 2, and will do so more vigorously the less informative is ξ. Similarly, Nature would not encounter difficulties if she could simply condition utilities directly on ξ (but not ζ), again essentially achieving perfect information processing. Nature's difficulties arise when she would like her agents to condition their actions on information such as ξ that she cannot "observe", but which her agents cannot process perfectly.[19]

---

[18] Nature would prefer to make the agents *beliefs* less sensitive to ζ, and will do so to the extent possible. But as long as she cannot purge ζ from the agent's beliefs, there are gains from manipulating the utility function. Why doesn't Nature instead make the agent less responsive to his beliefs by giving him a stronger prior or otherwise tinkering with the function ρ? Because this makes the agent less responsive to k as well as ζ.

[19] What if the agent can also be mistaken concerning the information conveyed by k? Then Nature's design problem is more difficult and the details of the solution will differ, but the utility function will again be manipulated to compensate for shortcomings in information processing.

We now have an expected utility function of the form

$$\max_{\epsilon^*} \rho_{\mathcal{E}}(\theta | k, \xi, \zeta) u(\epsilon^*, k, \bar{\theta}) + (1 - \rho_{\mathcal{E}}(\theta | k, \xi, \zeta)) u(\epsilon^*, k, \underline{\theta}), \qquad (15)$$

or, equivalently, we can write this in terms of a utility function defined over consumption levels $z \in \{\underline{z}, \bar{z}\}$ rather than over consumption cutoffs ϵ^* (just as (6) and (7) are equivalent):

$$\max_{z} \rho_{\mathcal{E}}(\theta | k, \xi, \zeta) u(z, \epsilon, k, \bar{\theta}) + (1 - \rho_{\mathcal{E}}(\theta | k, \xi, \zeta)) u(z, \epsilon, k, \underline{\theta}). \qquad (16)$$

Relative consumption effects now arise not only through the processing of information but also directly through utility externalities. An appropriate policy response may then call for more than simply providing and processing information. In particular, removing uncertainty from the economy will not banish relative consumption effects and their attendant inefficiencies.

Notice that there is no question of Nature's designing us to solve some problems of inordinate complexity. The ability to read these words reflects a triumph of biological engineering. Our argument requires only that Nature cannot ensure that we can solve every complex problem we encounter, and that she will accordingly adopt information-processing shortcuts whenever she can. "In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned." (Clark [13, p. 64].) We have developed a particularly simple model of the limitations Nature faces in designing her agents and the resulting information-processing shortcuts. The details of the shortcuts will depend upon the specific form of the limitations, but the general principle remains: information processing complexities can prompt relative consumption effects to spill over into utility externalities.

3.3 Example

We can illustrate the implications of (16). To focus attention on the persistence of relative consumption effects in the absence of uncertainty, we assume that the distribution of the state θ and the signals ξ and ζ, and hence the agents' beliefs captured in ρ, are degenerate. To simplify the derivations, we take a step toward realism in assuming that the consumption choice z is drawn from $\mathbb{R}_+$ (rather than $\{\underline{z}, \bar{z}\}$). The information received by the agent is then a vector of n consumption levels. We denote this observation by $(\hat{z}_1, \ldots, \hat{z}_n) \equiv \hat{z}$ and let $m(\hat{z})$ be the resulting mean observed consumption level.

Let the utility function corresponding to (15) be given by

$$u(z;\varepsilon;\hat{z}|\alpha) = z - \tfrac{1}{2}\alpha(z - m(\hat{z}))^2 - \frac{z^2}{2\varepsilon}. \tag{17}$$

We interpret $z^2/2\varepsilon$ as the disutility associated with securing consumption z and interpret ε (distributed uniformly among the continuum of agents) as an index of productivity. As suggested by Proposition 2, relative consumption effects (captured by the term $-\frac{1}{2}\alpha(z - m(\hat{z}))^2$) take the form of discouraging the agent from choosing a consumption level too different from a reference level that depends upon observations of others' consumption. Higher values of the parameter $\alpha \geq 0$ indicate that relative consumption effects are more important.

Let F be a distribution function with $F(z)$ identifying the proportion of the agents in the economy choosing a consumption level less than or equal to z. Let ϕ^n_F be the measure over n-tuples $\{\hat{z}_1, \ldots, \hat{z}_n\}$ induced by the distribution F. An equilibrium is then a distribution of consumption levels with the property that, if each agent is allowed to draw n observations from the previous generation's distribution and then choose an optimal consumption level, we reproduce the distribution.[20]

Definition 2 An information-based equilibrium is a distribution $F^*(z)$ and a specification of utility-maximizing choices $z^*(\varepsilon;\hat{z})$ such that

$$F^*(z) = (\lambda \times \phi^n_{F^*})(\{(\varepsilon;\hat{z}) : z^*(\varepsilon;\hat{z}) \leq z\}). \tag{18}$$

where λ is Lebesgue measure.

Establishing the existence of such an equilibrium is a straightforward fixed-point argument. Solving the first-order condition for utility maximization, we obtain the optimal choice function

$$z^*(\varepsilon;\hat{z}|\alpha) = \varepsilon\frac{1 + \alpha m(\hat{z})}{1 + \alpha\varepsilon}. \tag{19}$$

If $\alpha > 0$, then relative consumption effects push the allocation away from the "no distortion" solution of $z^*(\varepsilon; m(\hat{z})|0) = \varepsilon$, with larger values of α prompting large distortions. The optimal choice function $z^*(\varepsilon; m(\hat{z})|\alpha)$ is increasing in both ε and $m(\hat{z})$, so that larger characteristics and larger observations of

[20] We could equivalently interpret this as a rational-expectations equilibrium in which agents sample their own generation.

others' consumption lead to larger consumption. The equilibrium is inefficient if $\alpha > 0$, with gains to be had from pushing the consumption of each agent for whom $\epsilon > 0$ closer to the mean consumption.

Information-based relative consumption effects thus induce some agents to increase consumption and others to decrease consumption, compared to the base case of no such effects. The former will be those who have observed other agents with particularly high consumption levels, while the latter will have observed small consumption levels.

We can calculate the support of the equilibrium distribution of consumption levels by first noticing that

$$z^*(0, m(\hat{z})|\alpha) = 0;$$

so that agents with the lowest characteristic choose their no-distortion consumption level of zero, regardless of what they observe. Next, notice that

$$z^*(\epsilon, z^*(\epsilon, m(\hat{z})|0)|\alpha) = z^*(\epsilon, m(\hat{z})|0) = \epsilon. \qquad (20)$$

Hence, an individual who observes a mean consumption equal to the consumption the agent would have been chosen in the absence of relative consumption effects acts is if there are no relative consumption effects. It follows from (20) that the least upper bound on the range of equilibrium values of $z^*(\epsilon, m(\hat{z})|\alpha)$ is unity.[21] As α increases from zero, the range of realized values thus remains fixed at $[0, 1]$, while the distortion created by relative consumption effects grows stronger.

Now consider the effects of a policy intervention that replaces ϵ in the utility function given by (17) with $a + (1 - a)\epsilon$ for some $a \in [0, 1]$.[22] Figure 4 depicts this change. We can think of the policy intervention as a program

[21] To verify this, let $\tilde{z}(\alpha)$ be the upper bound. It suffices to consider $\epsilon = 1$ (since $z^*(\epsilon, m(\hat{z})|\alpha)$ is increasing in ϵ). If the range of $z^*(\epsilon, m(\hat{z})|\alpha)$ is finite, it suffices to observe that $z^*(1, \tilde{z}(\alpha)|\alpha) = \tilde{z}(\alpha)$ must hold, which can be satisfied only if $\tilde{z}(\alpha) = 1$. To see that an unbounded range of values of $z^*(1, m(\hat{z})|\alpha)$ is impossible, let $z^*(1, m(\hat{z})|\alpha) = h(m(\hat{z}))$. Then note that for $m(\hat{z}) > 1$, we have $m(\hat{z}) < h(m(\hat{z}))$ and $dh(m(\hat{z}))/dm(\hat{z}) < 1$. A positive measure of observations in the interval $[1 + \delta, 1 + \delta + \eta]$ would then require an identical measure in each of the infinite sequence of ascending, disjoint (for sufficiently small η) intervals $[h^{-1}(1 + \delta), h^{-1}(1 + \delta + \eta)]$, $[h^{-1}(h^{-1}(1 + \delta)), h^{-1}(h^{-1}(1 + \delta + \eta))]$, $[h^{-1}(h^{-1}(h^{-1}(1 + \delta))), h^{-1}(h^{-1}(h^{-1}(1 + \delta + \eta)))]$, and so on, an impossibility.

[22] A growing literature addresses the policy implications of relative consumption effects. For example, Corneo [15] shows that in the presence of rank-based relative income effects, redistributive taxation may be most useful in societies with relatively egalitarian distributions of income. Ljundqvist and Uhlig [32] show that if current utilities depend upon past consumption, then countercyclical fiscal policy can be effective in dampening the relative effects.

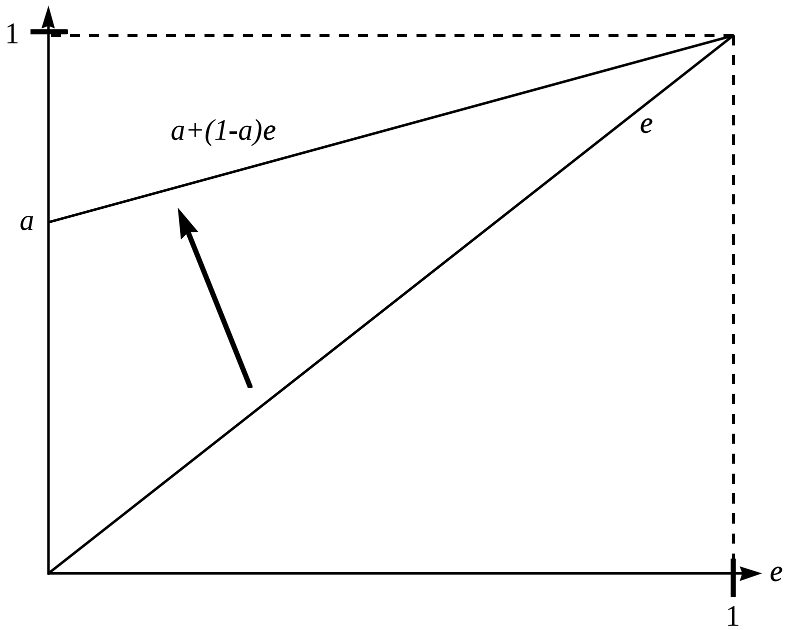

Figure 4: Productivity enhancement

enhancing productivities in the population, concentrating on those at the bottom end of the productivity scale, perhaps by increasing access to educational resources. As $a \to 1$, the productivity of every agent in the economy approaches the maximum productivity of unity. We began this example with the case of $a = 0$.

Utility maximization now yields an optimal choice function of (cf. (19)):

$$z^*(\epsilon; m(\hat{z})|\alpha; a) = \frac{(a + (1-a)\epsilon)(1 + \alpha m(\hat{z}))}{1 + \alpha(a + (1-a)\epsilon)}.$$

Once again, for any value of α, an agent of characteristic ϵ who observes $m(\hat{z}) = z^*(\epsilon; m(\hat{z})|0; a) = a + (1-a)\epsilon$ will choose $z = a + (1-a)\epsilon$, so that an observation of a mean consumption level equal to the choice that would have been chosen in the absence of relative consumption effects leaves the optimal choice unaltered.

An argument analogous to the case of $a = 0$ allows us to conclude that, for any value of $\alpha \geq 0$, the range of the equilibrium consumption levels is now $[a; 1]$ (compared to $[0; 1]$ in the original case of $a = 0$). As $a \to 1$, equilibrium consumption levels $z^*(\epsilon; m(\hat{z})|\alpha; a)$ thus converge on a mass point at unity (for all α):

$$z^*(\epsilon; m(\hat{z})|\alpha; 1) = 1 \qquad (21)$$

$$u(z^*(\varepsilon;m(\varepsilon)|\alpha,1);\varepsilon;m(\varepsilon)|\alpha,1) = \tfrac{1}{2}. \tag{22}$$

Hence, relative consumption effects and their attendant inefficiency disappear as productivity variation is eliminated from the population. Utilities converge to a value of $\frac{1}{2}$, yielding an efficient outcome in which each agent receives the maximum utility level in the population before the policy intervention.

4 A Comparison with Rank-Based Effects

We contrast the previous example with one in which relative consumption effects arise out of rank considerations.[23] We again consider a continuum of agents with characteristics ε distributed uniformly on $[0,1]$ and who must choose consumption or effort levels $z \in \mathbb{R}_+$. Let the utility function now be given by

$$u(z;\varepsilon;r(z)) = r(z) - \frac{z^2}{2\varepsilon}\frac{(1+\alpha\varepsilon)^3}{(1+\alpha m)^2}; \tag{23}$$

where α matches its value from Section 3.3 and m is the mean consumption level for the information-based equilibrium of Section 3.3 when $a = 0$ (both for reasons that will soon be apparent), and $r(z) : \mathbb{R}_+ \to [0,1]$ identifies the rank of an agent who chooses consumption level z. We interpret rank as being valuable because there are prizes allocated on the basis or rank. The variable z may thus be valued for its own sake as well as for its ability to secure rank, where (23) captures the extreme case in which only rank is important.

Let $z^*(\varepsilon;r)$ be a utility-maximizing consumption choice given characteristic ε and rank function r.

Definition 3 A rank-based equilibrium is a strictly increasing function $r^*(z) : \mathbb{R}_+ \to [0,1]$ and specification of utility-maximizing consumption levels $z^*(\varepsilon;r^*)$ such that

$$r^*(z) = \lambda(\{\varepsilon : z^*(\varepsilon;r^*) \leq z\});$$

where λ is Lebesgue measure.

A rank-based equilibrium is thus a distribution of consumption levels that is reproduced when each agent makes an optimal consumption choice, given

[23] See Hopkins and Kornienko [27] for a similar analysis of rank-based effects.

the equilibrium choices of others and hence the translation of consumption levels to ranks.

In equilibrium, we must have $r(z(\epsilon)) = \epsilon$, and hence the efficient solution maximizes

$$u(z;\epsilon;\epsilon) = \epsilon - \frac{z^2}{2\epsilon}\frac{(1+\alpha\epsilon)^3}{(1+\alpha m)^2};$$

giving

$$z^*(\epsilon;\epsilon) = 0:$$

Because only rank matters for allocation, efficient effort levels are zero.

The first-order condition for the maximization of (23) is

$$\frac{dr(z)}{dz} - \frac{z}{\epsilon}\frac{(1+\alpha\epsilon)^3}{(1+\alpha m)^2} = 0:$$

Using the equilibrium condition $r(z(\epsilon)) = \epsilon$, we have

$$\frac{dz}{d\epsilon} = \frac{1}{dr/dz} = \frac{\epsilon}{z}\frac{(1+\alpha m)^2}{(1+\alpha\epsilon)^3}; \qquad (24)$$

yielding a differential equation that we can solve for equilibrium consumption levels:

$$z^*(\epsilon; r^*) = \epsilon\frac{1+\alpha m}{1+\alpha\epsilon}:$$

Hence, this rank-based equilibrium gives rise to consumption levels identical to the mean consumption levels of the information-based equilibrium of Example 1 (cf. (19)). Average behavior in the two economies is thus indistinguishable.[24]

Rank concerns prompt agents to choose higher equilibrium effort levels than does the efficient solution.[25] It would be individually valuable to do so, but the equilibrium effect of all agents doing so is to decrease utility. In equilibrium, the rank of agent ϵ is given by ϵ. The efficient solution calls for agent ϵ to simply assume that his rank is fixed at ϵ, and then choose an optimal consumption plan contingent on this rank. The realized equilibrium produces the same distribution of ranks, but only after every agent has increased consumption in a collectively vain effort to increase rank.

[24] Brock and Durlauf [10] and Durlauf [18] discuss the issues that arise when studying behavior in the presence of relative consumption effects.

[25] If we replaced $r(z)$ with $(1-\gamma)z + \gamma r(z)$ in (23), so that effort level z leads both to direct consumption of $(1-\gamma)z$ and to rank-based consumption of $\gamma r(z)$, then efficient effort levels would be positive, but equilibrium effort levels would still exceed efficient levels, with the discrepancy increasing in γ.

As before, let the characteristic ϵ be replaced by $a + (1 - a)\epsilon$, so that an increase in a corresponds to an enhancement of productivities that is concentrated among low-productivity agents. As $a \to 1$, every agent's productivity increases to the maximum level of unity. The efficient level of effort remains unaltered at zero. Taking the first-order condition for maximizing utility and again using the equilibrium relationship that $r(z(\epsilon)) = \epsilon$, we can solve for (cf. (24))

$$\frac{dz}{d\epsilon} = \frac{a + (1 - a)\epsilon}{z} \frac{(1 + \alpha m)^2}{(1 + \alpha(a + (1 - a)\epsilon))^3} = \frac{1}{z} \frac{(1 + \alpha m)^2}{(1 + \alpha)^3};$$

where the second equality holds for the limiting case of $a = 1$. Confining attention to this case and using the boundary condition that $z(0) = 0$, this differential equation can be solved for

$$\begin{aligned} z^*(\epsilon; r^*|1) &= \left(2\epsilon \frac{(1 + \alpha m)^2}{(1 + \alpha)^3}\right)^{\frac{1}{2}} > \epsilon \frac{1 + \alpha m}{1 + \alpha \epsilon} \\ u(z^*(\epsilon; r^*|1); \epsilon; r(z^*(\epsilon; r^*|1))|1) &= 0 : \end{aligned}$$

Raising the productivity of every agent in the economy to the maximum tends to tighten the distribution of consumption, enhancing the payoff to seeking status and unleashing a countervailing tendency for every agent to compete more vigorously by increasing effort. Rank-based inefficiencies are thus exacerbated, with the entire population exerting more effort while being reduced to the zero utility level garnered by only the least fortunate agent before the policy intervention.[26]

Inequality policies can thus have quite different effects depending upon the nature of the relative consumption effects that appear in utility functions. If these effects are Nature's reaction to an inference problem, then smoothing productivities can alleviate the distortions of relative income effects. If the effects arise because rank or status is an important allocation device, then smoothing productivities can be ruinous. By increasing social mobility, inequality smoothing in the latter case ensures that more resources are wasted in the zero-sum quest for status. The inefficiency of relative consumption effects might be reduced in this case by entrenching consumption differences more solidly, making social mobility more difficult. Perhaps paradoxically, everyone in our deliberately extreme example might prefer social rigidity to social mobility, including those on the bottom.

[26] For similar reasons, equilibria in rent-seeking games or wars of attrition are typically most wasteful when contestants are evenly matched.

5 Conclusion

The message of this paper is that concerns about rank and status are not the only possible origins of relative consumption effects. Instead, relative consumption effects may be an important information-processing tool for making better decisions in a fluctuating environment. In addition, the relative consumption effects arising out of these two motivations can have quite different properties. Smoothing the distribution of productivities can push equilibrium consumption patterns toward the base case of no relative consumption effects when the latter reflect information concerns, suggesting that equalization policies may be surprisingly effective in boosting utilities. Smoothing the productivity distribution can push consumption patterns away from the base case of no relative consumption effects when the latter arise out of status considerations, suggesting that attempts to tighten the distribution of status may counterproductively redouble the quest for high status. Relative-consumption inefficiencies are alleviated in the former case, but exacerbated in the latter.

Is there any evidence for information-based relative consumption effects, or for the limitations in our ability to process information that would prompt Nature to build relative consumption effects into our utility functions? Psychologists report experimental evidence suggesting that people are poor Bayesians (Kahneman and Tversky [29]). Psychologists also report an inclination to conform to the behavior of others (Aronson [1, Chapter 2], Cialdini [12, ch. 4]), even in situations in which one would be extremely hard-pressed to identify an information-based reason for doing so.[27]

Information-based relative consumption effects induce pressure for people to conform to the consumption decisions of others. Low productivity agents will strive to increase consumption, while the high productivity agents will attenuate their consumption in order to not be too conspicuously different. The latter finding contrasts with the popular view of relative consumption effects as creating incessant incentives to consume more. Attempts to "keep up with the Joneses" appear to be ubiquitous, but no one conspicuously strives to "keep down with the Joneses." Does this suggest that status

[27] Among the most striking is the classic work of Asch [2, 3], in which an apparent desire to conform prompted experimental subjects to make obviously incorrect choices when matching the lengths of lines, while denying that they were influenced by the choices of others. A committed Bayesian could still rationalize this behavior with a pure information-processing story, centered around subjects' uncertainty as to the consequences of the various choices they faced. But a pure information processing model is not useful if it can be preserved only with the help of complicated updating models that depend too finely on the circumstances of the choice.

is more important than information as a source of relative consumption effects?

Two responses are in order. First, it is likely that the observations which motivate relative consumption effects are stratified, with people influenced more by the consumption of others who appear to be "like them" than people whose lifestyles are quite different. We may be unfazed by comparisons with internet billionaires, but may be much more conscious of how our consumption compares with that of our colleagues. Similarly, people with high consumption levels may then compare themselves primarily to people in similar circumstances, freeing the internet billionaire from desperate attempts to reduce consumption and opening the possibility of an upward bias in consumption even at the high end of the consumption scale.[28]

In addition, the concept of likeness on which such stratification is based is both endogenous and liable to manipulation. Much like Sherlock Holmes' dog that didn't bark in the night, it is easier to observe what someone has than what they don't have, making high consumption levels inherently more obvious than low levels. More importantly, the advent of modern advertising and mass communication may especially contribute to this asymmetry.[29] Relative consumption effects may then push primarily upwards.

Second, and more importantly, information-based relative consumption effects do not imply that we should observe people anxious to reduce consumption levels to those of their peers (however broadly or narrowly defined). Instead, individual consumption patterns cannot distinguish information-based and status-based effects. The utility function given by (17), incorporating information-based effects, is concave in consumption z and has the comparative-static implication

$$\frac{dz^*(\epsilon;\hat{z}|\alpha)}{d\hat{z}_i} > 0;$$

so that higher values of observed consumption induce higher consumption. But the latter is what one means by "keeping up with the Joneses," and these are precisely the properties one expects from a utility function with status-based relative consumption effects. Information-based relative consumption effects imply not that we must observe people trying to keep down

[28] Notice that Nature has incentives to encourage such stratification, since information about the environment is more relevant when it comes from observing those in similar circumstances. At the same time, Nature must beware of building too much taste for stratification into our comparisons, since excessively narrow comparison groups may ensure that virtually no information gets transferred.

[29] Aronson [1, Chapter 3] discusses the persuasive ability of modern communications.

with the Joneses, but rather that people whose characteristics lead to high consumption levels should strive less vigorously to keep ahead of the Joneses than they would to catch up if the Joneses were ahead. This ability to coast while ahead is reflected in the long-standing notion of the idle rich. "...we must make allowances for the rich men's failings, and recollect that we, too, were very likely indolent and voluptuous, had we ... the daily temptation of a large income."[30] The salient observation is not that wealth may induce decreased effort, but rather that such effects should continue to characterize the relatively wealthy even as absolute wealth levels change.[31]

An ability to observe the qualitative features of utility functions is thus insufficient to distinguish the two models. Suppose instead we could observe utility functions exactly. In the case of (17), we would observe

$$z(1 + 2\alpha m(\hat{z})) - z^2 \left(\alpha + \frac{1}{2\epsilon}\right) - \alpha m(\hat{z})^2. \qquad (25)$$

One could rearrange this function as in (17) and conclude that information-based effects are at work. But one could also interpret this as a version of (23), in which the rank of consumption level z is given by $z + 2\alpha z m(\hat{z}) - \alpha m(\hat{z})^2$ and conclude that status-based effects are at work. Again, individual-level data may be insufficient to distinguish the two.

Sharp differences in information-based and status-based relative consumption effects appear in the restrictions they place on the collection of consumption levels and utility functions in the economy (rather than on any particular utility function). These are most obviously reflected in the contrasting comparative static implications sketched in Sections 3.3 and 4.

Distinguishing information-based and status-based relative consumption effects thus promises to be a formidable task. But given the sensitivity of economic policy outcomes to the nature of relative consumption effects, it is clearly an important task. Further theoretical modeling is likely to be useful in this respect, as are carefully designed experiments.

---

[30] William Madepeace Thackery, "George the Third," in *The Four Georges and The English Humorists* (London, Collins' Clear-Type Press 1910 (originally 1855)).

[31] Two centuries before Thackery, Robert Burton would comment that "Idleness is an appendix to nobility" (*Anatomy of Melancholy* (East Lansing, MI, Michigan State University Press 1965 (originally 1621), part 1, section 2, member 2, subsection 6), while nearly a century later Andrew Carnegie would view the disincentive effects of a fortune as sufficiently deleterious as to comment that "I would as soon leave my son a curse as the almighty dollar" (quoted in Burton J. Hendrick, *Life of Andrew Carnegie* (Garden City, NJ, Doubleday, Doran and Company, 1932), volume 1, chapter 17).

6 Appendix: Proofs

6.1 Proof of Lemma 1

6.1.1 Survival Proportions

We first calculate $\Psi_{\mathcal{E}}(\psi_t, \bar{\theta})$. Fix an admissible $E = \{\varepsilon(n,\xi), \ldots, \varepsilon(0,\xi)\}$, let the state be $\bar{\theta}$, and fix the proportion ψ_t. Then proportion $r(k, \psi_t)$ of the agents in period t will observe k values of $\bar{z}$, where

$$r(k, \psi_t) = \binom{n}{k} \psi_t^k (1 - \psi_t)^{n-k}.$$

These agents will choose $\varepsilon(k, \xi)$ as their decision cutoff. Hence, proportion

$$S(\varepsilon(k,\xi), \bar{z}, \bar{\theta}) = \int_{-\infty}^{\infty} \int_{\epsilon(k,\xi)}^{1} \left(\tfrac{1}{2} + b(\varepsilon - q)\right) g(\xi, \bar{\theta})\, d\varepsilon\, d\xi$$

of these agents will choose strategy $\bar{z}$ and survive, while proportion

$$S(\varepsilon(k,\xi), \underline{z}, \bar{\theta}) = \int_{-\infty}^{\infty} \int_{0}^{\epsilon(k)} \tfrac{1}{2}\, g(\xi, \underline{\theta})\, d\varepsilon\, d\xi$$

will choose $\underline{z}$ and survive. We then have

$$\Psi_{\mathcal{E}}(\psi_t, \bar{\theta}) = \frac{\sum_{k=0}^{n} r(k, \psi_t) S(\varepsilon(k,\xi), \bar{z}, \bar{\theta})}{\sum_{k=0}^{n} r(k, \psi_t) S(\varepsilon(k,\xi), \bar{z}, \bar{\theta}) + \sum_{k=0}^{n} r(k, \psi_t) S(\varepsilon(k,\xi), \underline{z}, \bar{\theta})}.$$

Evaluating this equation at $\psi_1 = 0$ and $\psi_t = 1$ gives

$$\begin{aligned} \Psi_{\mathcal{E}}(0, \bar{\theta}) &> 0 \\ \Psi_{\mathcal{E}}(1, \bar{\theta}) &< 1, \end{aligned}$$

and hence that probabilities remain within the unit interval. We next establish that $\Psi_{\mathcal{E}}(\psi_t, \bar{\theta})$ is increasing in ψ_t. To do so, we note that

$$\frac{S(\varepsilon(k), \bar{z}, \bar{\theta})}{S(\varepsilon(k), \bar{z}, \bar{\theta}) + S(\varepsilon(k), \underline{z}, \bar{\theta})}$$

is increasing in k, since more agents choose $\bar{z}$ as k increases, increasing $S(\varepsilon(k), \bar{z}, \bar{\theta})$ and decreasing $S(\varepsilon(k), \underline{z}, \bar{\theta})$. The result that $\Psi_{\mathcal{E}}(\psi_t, \bar{\theta})$ is increasing in ψ_t then follows from noting that the distribution $r(k, \psi_t)$ increases, in the sense of first-order stochastic dominance, as ψ_t increases. We illustrate $\Psi_{\mathcal{E}}(\psi_t, \bar{\theta})$ in Figure 5.

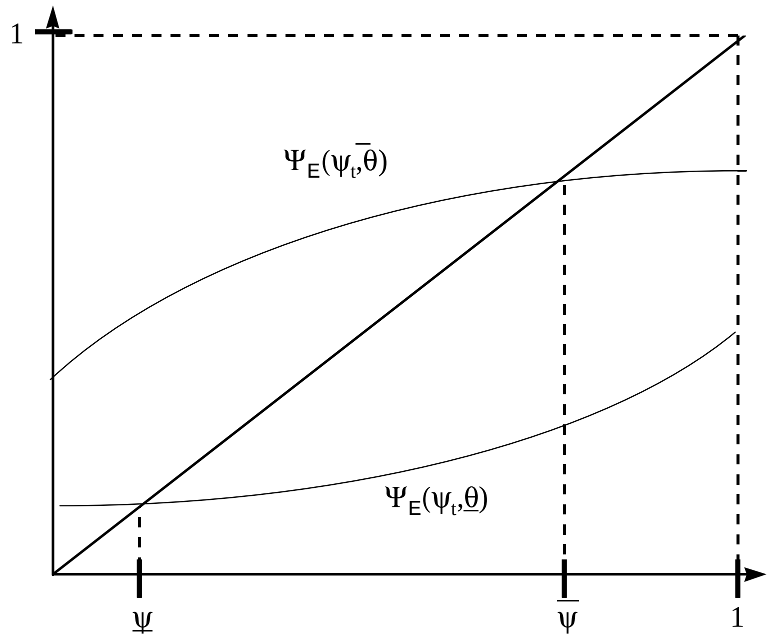

Figure 5: Functions $\Psi_{\mathcal{E}}(\psi_t;\overline{\theta})$ and $\Psi_{\mathcal{E}}(\psi_t;\underline{\theta})$

Let $\{q_i\}_{i=1}^{\infty}$ be a sequence of values of q and let E(q) be admissible given q. Then as $q \to \frac{1}{2}$, the sequence of functions $\Psi_{\mathcal{E}(q)}(\psi_t;\overline{\theta})$ converges uniformly (over admissible E and $\psi_t \in [0,1]$) to a constant function. Hence, there exists a value $q^* \in [0,\frac{1}{2})$ such that for all $q \in [q^*,\frac{1}{2}]$, the slope of $\Psi_{\mathcal{E}}(\psi_t;\overline{\theta})$ does not exceed $\frac{1}{2}$.

We now let the state be $\underline{\theta}$ and perform a similar examination of the function $\Psi_{\mathcal{E}}(\psi_t;\underline{\theta})$. Here, we have

$$\Psi_{\mathcal{E}}(\psi_t;\underline{\theta}) = \frac{\sum_{k=0}^{n} r(k;\psi_t) S(\epsilon(k;\xi);\overline{z};\underline{\theta})}{\sum_{k=0}^{n} r(k;\psi_t) S(\epsilon(k;\xi);\overline{z};\underline{\theta}) + \sum_{k=0}^{n} r(k;\psi_t) S(\epsilon(k;\xi);\underline{z};\underline{\theta})};$$

where

$$S(\epsilon(k;\xi);\overline{z};\underline{\theta}) = \int_{-\infty}^{\infty}\int_{\epsilon(k)}^{1} \left(\tfrac{1}{2} + b(\epsilon - (1-q))\right) g(\xi;\underline{\theta})\, d\epsilon\, d\xi$$

$$S(\epsilon(k;\xi);\underline{z};\underline{\theta}) = \int_{-\infty}^{\infty}\int_{0}^{\epsilon(k)} \tfrac{1}{2} g(\xi;\underline{\theta})\, d\epsilon\, d\xi.$$

Once again, $\Psi_{\mathcal{E}}(\psi_t;\underline{\theta})$ is strictly contained within the unit interval. The fraction

$$\frac{S(\epsilon(k;\xi);\overline{z};\underline{\theta})}{S(\epsilon(k;\xi);\overline{z};\underline{\theta}) + S(\epsilon(k;\xi);\underline{z};\underline{\theta})}$$

is increasing in k, as a smaller value of $\varepsilon(k)$ induces more agents to choose $\bar{z}$ (increasing $S(\varepsilon(k,\xi),\bar{z},\underline{\theta})$ and decreasing $S(\varepsilon(k,\xi),\underline{z},\underline{\theta})$), and hence $\Psi_{\mathcal{E}}(\psi_t,\underline{\theta})$ is increasing in ψ_t. An argument analogous to that for the case of $\Psi_{\mathcal{E}}(\psi_t,\bar{\theta})$ establishes that there exists q^* such that for all $q \in (q^*,\frac{1}{2}]$, both $\Psi_{\mathcal{E}}(\psi_t,\underline{\theta})$ and $\Psi_{\mathcal{E}}(\psi_t,\bar{\theta})$ have slopes less than $\frac{1}{2}$. We hereafter assume that $q \in (q^*,\frac{1}{2})$. Finally, we note that, for all k,

$$\frac{S(\varepsilon(k,\xi),\bar{z},\bar{\theta})}{S(\varepsilon(k,\xi),\bar{z},\bar{\theta})+S(\varepsilon(k,\xi),\underline{z},\bar{\theta})} > \frac{S(\varepsilon(k,\xi),\bar{z},\underline{\theta})}{S(\varepsilon(k,\xi),\bar{z},\underline{\theta})+S(\varepsilon(k,\xi),\underline{z},\underline{\theta})}$$

and hence

$$\Psi_{\mathcal{E}}(\psi_t,\bar{\theta}) > \Psi_{\mathcal{E}}(\psi_t,\underline{\theta}).$$

Figure 5 illustrates both functions.

6.1.2 An Algebra of Sets

Let Υ denote the process generated by the transition rule $(\Psi_{\mathcal{E}},\Theta)$. Let $\mu((\psi_t,\theta_t);(\psi_0,\theta_0))$ be the probability measure over the time-t state (ψ_t,θ_t) given initial condition (ψ_0,θ_0). We construct a measure μ^* and show μ converges to μ^* as t gets large.

Because q has been chosen so that $q > q^*$, and hence the slopes of $\Psi_{\mathcal{E}}(\psi_t,\bar{\theta})$ and $\Psi_{\mathcal{E}}(\psi_t,\underline{\theta})$ fall short of $\frac{1}{2}$, it must be that

$$\Psi_{\mathcal{E}}(\underline{\psi},\bar{\theta}) > \Psi_{\mathcal{E}}(\bar{\psi},\underline{\theta}).$$

Figure 6 illustrates this property, while footnote 32 describes how it is used.

We now define a pair of countably infinite collections of sets,

$$\{\overline{A}_k\}_{k=1}^{\infty}; \qquad \{\underline{A}_k\}_{k=1}^{\infty}; \tag{26}$$

where each of the sets $\overline{A}_k$ or $\underline{A}_k$ is a subset of the state space $[\underline{\psi},\bar{\psi}] \times \{\underline{\theta},\bar{\theta}\}$. We will adopt the convention that for any k, $\underline{A}_k \subset [\underline{\psi},\bar{\psi}] \times \{\underline{\theta}\}$ and $\overline{A}_k \subset [\underline{\psi},\bar{\psi}] \times \{\bar{\theta}\}$, and hereafter suppress the notation for the values of θ when specifying a set $\overline{A}_k$ or $\underline{A}_k$. We also drop the time subscripts and drop the subscript E on the function Ψ.

The collections $\{\overline{A}_k\}_{k=1}^{\infty}$ and $\{\underline{A}_k\}_{k=1}^{\infty}$ are defined recursively. First, we define two infinite collections of the level-one sets by letting

$$\begin{aligned}\overline{A}_{11} &= (\underline{\psi}, \Psi(\underline{\psi},\bar{\theta}))\\ \overline{A}_{1i+1} &= \Psi(\overline{A}_{1i},\bar{\theta}).\end{aligned}$$

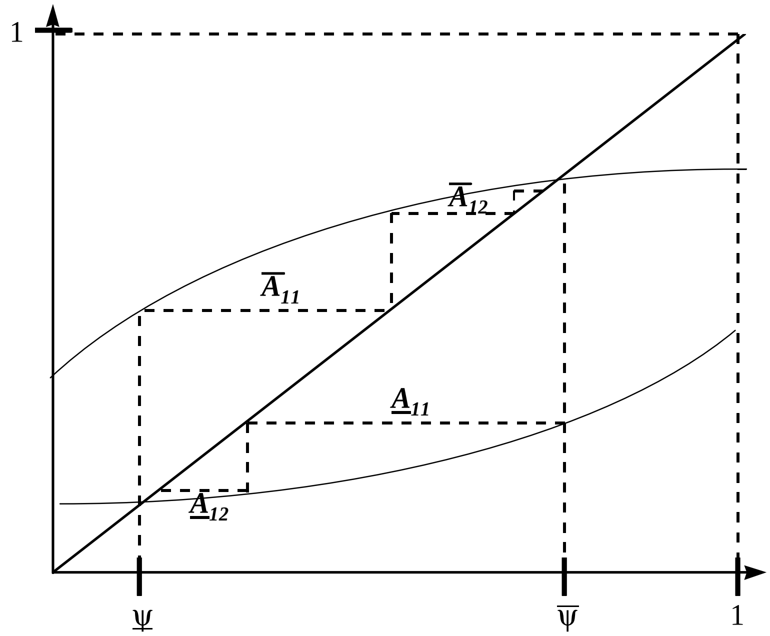

Figure 6: Functions $ª_{\mathcal{E}}(Ã_t;\bar{z})$ and $ª_{\mathcal{E}}(Ã_t;\underline{z})$, with initial first-level sets

and

$$\underline{A}_{11} = (ª(\bar{Ã};\underline{µ});\bar{Ã})$$
$$\underline{A}_{1i+1} = ª(\underline{A}_{1i};\underline{µ}):$$

Figure 6 illustrates the initial steps in this construction. Notice that the collections $\{\bar{A}\}_{1i}^{\infty}$ and $\{\underline{A}\}_{1i}^{\infty}$ are disjoint and that the union of each collection equals $[\underline{Ã};\bar{Ã}]$, except for a countable collection of endpoints that have been omitted. As we define each level of sets, we shall similarly exclude a countable collection of endpoints. Let M denote the union of these collections of excluded sets and note that M will itself be countable. It will be clear from the construction that the sets M and $[0;1] \setminus M$ are both forward invariant under the Markov process ¨ (i.e., ¨ never leaves M if it begins in M, and that no path originating outside M can enter M).

We now define the level-two sets. For each first-level set $\underline{A}_{1i}$ or $\bar{A}_{1i}$, we define a countably infinite collection of second-level sets, given by

$$\bar{A}_{21} = \{ª(\underline{A}_{1i};\underline{µ})\} \times \{\bar{µ}\}$$
$$\bar{A}_{2j+1} = ª(\bar{A}_{2j};\bar{µ})$$

$$\underline{A}_{21} = \{ª(\bar{A}_{1i};\bar{µ})\} \times \{\underline{µ}\}$$

$$\underline{A}_{2j+1} = \Psi(\underline{A}_{2j}, \underline{\theta}).$$

We continue in this fashion, defining, for each level-$(n-1)$ set $\underline{A}_{(n-1)i}$ or $\overline{A}_{(n-1)i}$, a countably infinite collection of level-n sets, given by

$$\begin{aligned}
\overline{A}_{n1} &= \{\Psi(\underline{A}_{(n-1)i}, \underline{\theta})\} \times \{\overline{\theta}\} \\
\overline{A}_{nj+1} &= \Psi(\overline{A}_{nj}, \overline{\theta})
\end{aligned}$$

$$\begin{aligned}
\underline{A}_{n1} &= \{\Psi(\overline{A}_{(n-1)i}, \overline{\theta})\} \times \{\underline{\theta}\} \\
\underline{A}_{nj+1} &= \Psi(\underline{A}_{nj}, \underline{\theta}).
\end{aligned}$$

We thus have a countably infinite collection of levels $\{1, 2, \ldots\}$, with the sets defined at each level $n > 1$ containing a countably infinite collection of sets for each of the sets occurring at a previous level, of which there are countably many. The entire collection of such sets is thus countably infinite, and the sets can be placed in an order captured by the index k of (26).

Let Λ_i denote the collection of level-i sets, and let Λ denote the collection of all such sets along with the empty set, or $\Lambda = \cup_{i=1}^{\infty} \Lambda_i \cup \{\emptyset\}$. Let $\overline{\Lambda}$ be the collection of countable unions of disjoint sets in Λ. Then we note that $\overline{\Lambda}$ is an algebra on $\cup \Lambda = [0, 1] \times \{\underline{\theta}, \overline{\theta}\} \setminus \mathcal{M}$.[32]

6.1.3 Candidate Limiting Measure μ^*

We now define a sequence of measures $\{\mu_k\}_{k=1}^{\infty}$.

To define the first-level measure μ_1, consider a hypothetical Markov process Υ_1, in which the states are taken to be the first-level sets, so that the state space is Λ_1 (where each set is taken to be a state, rather than the union of the sets taken to be the state space). Let the transition probability between sets A and A', under process Υ_1, be the probability that the process

[32] An algebra on a set A is a collection of subsets of A that includes A and the empty set and that is closed under the taking of complements and finite unions. Our assumption that the slopes of $\Psi(\psi_t, \overline{\theta})$ and $\Psi(\psi_t, \underline{\theta})$ are bounded below one-half, implied by $q > q^*$, makes its first appearance here. By ensuring that $\Psi_{\mathcal{E}}(\underline{\psi}, \overline{\theta}) > \Psi_{\mathcal{E}}(\overline{\psi}, \underline{\theta})$, this assumption ensures that each level-i set is contained in a unique level-j set for every $j < i$ (rather than intersecting more than one level-j set) and that each level-j set is a countable union of level-i sets for $j < i$. This allows us to conclude that $\overline{\Lambda}$ is an algebra, with the complement of any finite union of elements in $\overline{\Lambda}$ being a countable union of elements in $\overline{\Lambda}$. In addition, it simplifies subsequent calculations. It allows us to recursively calculate the measures μ_i in Section 6.1.3 and verify in each case that the resulting measures are consistent, in that the level-i measure of a set A is the sum of the level-k measures of its subsets, for any $k > i$. It also simplifies the calculation of the stationary distribution in Section 6.1.6.

Υ moves from A to A'. Our sets are constructed so that this probability is independent of the particular state in A in which the process Υ finds itself, and hence this transition probability is well defined. The Markov process Υ_1 has a countably infinite state space, and is an aperiodic, irreducible, positive recurrent process. It is accordingly ergodic, with a unique stationary distribution, which we denote by μ_1. We can view μ_1 as a measure on the countable set Λ_1, endowed with the discrete σ-algebra. We can also view μ_1 as a measure on the space $[0,1] \times \{\underline{\theta}, \bar{\theta}\}$, with the σ-algebra generated by the members of Λ_1, each interpreted as a subset of $[0,1] \times \{\underline{\theta}, \bar{\theta}\}$.

At the first level, it is straightforward to calculate the stationary distribution μ_1. The measure μ_1 is given by

$$
\begin{aligned}
\mu_1(\bar{A}_{11}) &= \tfrac{1}{2}\tau \\
\mu_1(\bar{A}_{1i+1}) &= (1-\tau)\mu_1(\bar{A}_{1i})
\end{aligned}
$$

$$
\begin{aligned}
\mu_1(\underline{A}_{11}) &= \tfrac{1}{2}\tau \\
\mu_1(\underline{A}_{1i+1}) &= (1-\tau)\mu_1(\underline{A}_{1i}).
\end{aligned}
$$

Similarly, we can define a Markov process Υ_2 whose state space is the collection of second-level sets Λ_2. This Markov process is again aperiodic, irreducible, and positive recurrent. It is accordingly ergodic, with a unique stationary distribution which we call μ_2, which we can interpret as a measure on the discrete σ-algebra on the elements of Λ_2, or as a measure on $[0,1] \times \{\underline{\theta}, \bar{\theta}\}$ with the σ-algebra generated by the subsets of $[0,1] \times \{\underline{\theta}, \bar{\theta}\}$ contained in Λ_2. In addition, every level-one set is a countable union of level-two sets (minus a countable collection of endpoints), with a measure μ_1 that is the sum of the μ_2-measures of the level-two sets in contains. Hence, the σ-algebra generated by Λ_2 matches that generated by $\Lambda_1 \cup \Lambda_2$, and μ_1 and μ_2 agree on any set from this σ-algebra for which μ_1 is defined.

Continuing in this way, we define the entire collection

$$\{\mu_k\}_{k=1}^{\infty},$$

where each μ_k is defined on the σ-algebra generated by $\cup_{i=1}^{k} \Lambda_i$.

For any $A \in \Lambda$ and μ_k, either A is not contained in Λ_i for any $i \leq k$, in which case $\mu_k(A)$ is undefined, or A is contained in some Λ_i with $i \leq k$, in which case $\mu_k(A)$ is defined and is equal to $\mu_{k'}(A)$ for any $k' \geq i$. We can then define μ^* on Λ by taking $\mu^*(A) = \mu_k(A)$ for some k for which $\mu_k(A)$ is defined.

We can think of μ^* as being the limit of the sequence $\{\mu_k\}_{k=1}^{\infty}$. However, instead of a sequence of changing measures on a fixed domain, we have a sequence of measures on an ever-expanding sequence of domains, with the measures being constant along the sequence, in the sense that once a set enters the domain, its measure remains fixed along the remainder of the sequence.

6.1.4 μ^* is a measure

Intuitively, our next step is to show that μ^* is a measure. More precisely, we embed $\mathcal{A}$ in a σ-algebra and then show that μ^* extends uniquely to a measure on that σ-algebra. Recall that $\overline{\mathcal{A}}$ is the collection of countable unions of disjoint sets in $\mathcal{A}$, and is an algebra on $\cup\mathcal{A}$. We then proceed in five steps.

First, we extend μ^* from $\mathcal{A}$ to $\overline{\mathcal{A}}$. For each set A in $\overline{\mathcal{A}}$, we take a minimal representation of A to be the smallest collection of sets in $\mathcal{A}$ whose union is A. Because of the recursive structure of $\overline{\mathcal{A}}$, minimal representations are uniquely defined and consist of disjoint sets. Then $\mu^*(A')$ is defined for any A' in the minimal representation of A, and we take $\mu^*(A)$ to be the sum of the measures $\mu^*(A')$ of the sets in the minimal representation. It is then straightforward that μ^* is finite on $\overline{\mathcal{A}}$ (being bounded above by unity) and is finitely additive.

Second, we note that μ^* is countably subadditive on $\overline{\mathcal{A}}$. Suppose not, so we can find a case in which

$$\cup_{i=1}^{\infty} A_i = A; \qquad \sum_{i=1}^{\infty} \mu^*(A_i) > \mu^*(A);$$

where the union is disjoint. Because $\mu^*(A)$ is finite, there must be a finite n such that

$$\sum_{i=1}^{n} \mu^*(A_i) > \mu^*(A):$$

Because μ^* is finitely additive, we then have

$$\mu^*(\cup_{i=1}^{n} A_i) > \mu^*(A); \qquad \cup_{i=1}^{n} (A_i) \subset A;$$

with strict set inclusion in the second case, which is precluded by the construction of μ^*.

Third, we argue that μ^* is continuous from above at the empty set (Ash and Doléans-Dade [4, p. 10]), which in turn implies that it is countably additive ([4, Theorem 1.2.8])). To verify this, let $\{B_n\}_{n=0}^{\infty}$ be a sequence of sets in $\overline{\mathcal{A}}$ with

$$B_n \to \emptyset;:$$

We need to show that

$$\mu^*(B_n) \to 0 .$$

We can take the sequence B_n to be decreasing without sacrificing any generality. (If necessary, redefine each B_n to be the union $\cup_{k=n}^{\infty} B_k$.) Given that $\overline{\Lambda}$ consists of countable unions of sets in Λ, there must be a sequence of disjoint sets $\{C_m\}_{m=0}^{\infty}$ with each C_m in $\overline{\Lambda}$ and, for each B_n, a sequence of numbers k_{ni}, $i = 1, \ldots, \infty$, such that

$$B_n = \cup_{i=1}^{\infty} C_{k_{ni}},$$

for each set B_n. Because μ^* is countably subadditive, we have

$$\sum_{m=1}^{\infty} \mu^*(C_m) \leq 1,$$

and hence,

$$\lim_{k\to\infty} \sum_{m=k}^{\infty} \mu^*(C_m) = 0 .$$

But for every k, there is an h(k) such that

$$B_n = \cup_{i=1}^{\infty} C_{k_{ni}} \subset \cup_{m=h(k)}^{\infty} C_m ,$$

and the fact that $B_n \to \emptyset$ implies that $h(k) \to \infty$, giving $\mu^*(B_n) \to 0$, as required.

Fifth, we have now shown that μ^* is a countably additive, positive set function on the algebra $\overline{\Lambda}$. The Carathéodory Hahn Extension Theorem (Wheeden and Zygmund [42, p. 206]) then ensures that μ^* extends uniquely to a measure that is defined on the σ-algebra $\sigma(\overline{\Lambda})$ generated by $\overline{\Lambda}$ and that agrees with μ^* on Λ. Hence, μ^* is a measure (on $\sigma(\overline{\Lambda})$).

6.1.5 Ergodicity

We now establish the required limiting property. We construct a pair of sequences of Markov processes $\{\underline{\Upsilon}_i\}_{i=1}^{\infty}$ and $\{\overline{\Upsilon}_i\}_{i=1}^{\infty}$ from Υ by replacing (ψ_t, θ_t) in each sample path generated by Υ with (ψ, θ_t), where ψ is given by

$$\begin{array}{lll} \min\{\psi \in \overline{A}_{ij}\} & \text{if} & (\psi_t, \theta_t) \in \overline{A}_{ij} \\ \max\{\psi \in \overline{A}_{ij}\} & \text{if} & (\psi_t, \theta_t) \in \underline{A}_{ij} \end{array}$$

in the case of $\{\underline{\Upsilon}_i\}_{i=1}^{\infty}$ and

$$\begin{array}{lll} \max\{\psi \in \overline{A}_{ij}\} & \text{if} & (\psi_t; \theta_t) \in \overline{A}_{ij} \\ \min\{\psi \in \overline{A}_{ij}\} & \text{if} & (\psi_t; \theta_t) \in \underline{A}_{ij} \end{array}$$

in the case of $\{\overline{\Upsilon}_i\}_{i=1}^{\infty}$. The processes $\underline{\Upsilon}_i$ and $\overline{\Upsilon}_i$ are the counterparts of the Markov process Υ_i, constructed in Section 6.1.3, obtained by having each of $\underline{\Upsilon}_i$ and $\overline{\Upsilon}_i$ assign a value of ψ to every element in the state space of Υ.

Letting $\underline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i)$ and $\overline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i)$ be the counterparts of $\rho_{\mathcal{E}}(\bar{\theta}|k;\xi;t)$ for the processes $\underline{\Upsilon}_i$ and $\overline{\Upsilon}_i$, we have

$$\underline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i) \leq \rho_{\mathcal{E}}(\bar{\theta}|k;\xi;t) \leq \overline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i):$$

We obtain this result by noting that the process $\underline{\Upsilon}_i$ chooses the minimum (maximum) possible value of ψ (consistent with staying within the element of Λ_i occupied by the process Υ_i) whenever the state is $\bar{\theta}$ ($\underline{\theta}$), and hence minimizes the extent to which an observation of $\bar{z}$ signals state $\bar{\theta}$. The process $\overline{\Upsilon}_i$ does just the opposite, and hence maximizes the extent to which $\bar{z}$ indicates that the state if $\bar{\theta}$.

In addition, the processes $\underline{\Upsilon}_i$ and $\overline{\Upsilon}_i$ inherit the ergodicity of Υ_i. Hence, $\underline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i)$ and $\overline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i)$ converge (as $t \to \infty$) to limits

$$\underline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;i) \leq \overline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;i):$$

Furthermore, the former sequence is increasing in i and the latter sequence decreasing in i, with the former approaching the latter as $i \to \infty$. The probability $\rho_{\mathcal{E}}(\bar{\theta}|k;\xi;t)$, being bounded by $\underline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i)$ and $\overline{\rho}_{\mathcal{E}}(\bar{\theta}|k;\xi;t;i)$, must accordingly converge to this common limit. This allows us to establish the desired conclusion, namely that there exist $\rho_{\mathcal{E}}(\bar{\theta}|k;\xi)$, $k = 0;:::;n$, such that:

$$\lim_{t\to\infty} \rho_{\mathcal{E}}(\bar{\theta}|k;\xi;t) = \rho_{\mathcal{E}}(\bar{\theta}|k;\xi):$$

This establishes the result for any initial condition not contained in the set M. To extend the result to initial conditions in M, we note that because $\Psi_{\mathcal{E}}(\psi_t;\bar{\theta})$ and $\Psi_{\mathcal{E}}(\psi_t;\underline{\theta})$ have slopes bounded below $\frac{1}{2}$, we have, for any fixed sequence of realizations of values of θ and any two initial conditions ψ and ψ', that

$$|\psi_{t+1} - \psi'_{t+1}| \leq \frac{1}{2}|\psi_t - \psi'_t|:$$

Hence, for any realization of the sequence of random variables θ_t, the sample paths of realized values of ψ_t generated by any two initial conditions converge at an exponential rate.[33] Furthermore, the sample paths of realized values of ψ_t have the monotonicity property that $\psi_0 < \psi_0' \Rightarrow \psi_t < \psi_t'$ for all $t > 0$. This in turn ensures that the limiting probability $\rho_{\mathcal{E}}(\bar\theta|k;\xi)$ given an initial state in M lies between the limiting probabilities for some pair of larger and smaller initial states not contained in M, and hence must converge to the common limit attained by the latter pair.

6.1.6 $\rho_{\mathcal{E}}(\bar\theta|k;\xi)$ is increasing in ξ with $\rho_{\mathcal{E}}(\bar\theta|k+1;\xi) > \rho_{\mathcal{E}}(\bar\theta|k;\xi)$

That $\rho_{\mathcal{E}}(\bar\theta|k;\xi)$ is increasing in ξ follows from (8) and the monotone likelihood ratio property. Next, to establish $\rho_{\mathcal{E}}(\bar\theta|k+1;\xi) > \rho_{\mathcal{E}}(\bar\theta|k;\xi)$, it suffices (from (8)) to show that

$$\int_{\psi=0}^{1} \psi \, d\mu^*(\psi;\bar\theta) > \int_{\psi=0}^{1} \psi \, d\mu^*(\psi;\underline{\theta}). \tag{27}$$

First, we recall that

$$\underline{\psi} < \Psi_{\mathcal{E}}(\bar\psi;\underline{\theta}) \le \Psi_{\mathcal{E}}(\underline{\psi};\bar\theta) < \bar\psi;$$

which we use to construct the intervals in the following expression. Next, notice that the construction of the measure μ^* ensures, for any $A \in \sigma(\bar\Lambda)$

$$\begin{aligned}
A \subset [\underline{\psi}; \Psi_{\mathcal{E}}(\bar\psi;\underline{\theta})] &\Rightarrow \mu^*(A;\bar\theta) = \tau\mu^*(A;\underline{\theta}) \\
A \subset [\Psi_{\mathcal{E}}(\bar\psi;\underline{\theta}); \Psi_{\mathcal{E}}(\underline{\psi};\bar\theta)] &\Rightarrow \mu^*(A;\bar\theta) = \mu^*(A;\underline{\theta}) = 0 \\
A \subset [\Psi_{\mathcal{E}}(\underline{\psi};\bar\theta);\bar\psi] &\Rightarrow \tau\mu^*(A;\bar\theta) = \mu^*(A;\underline{\theta}).
\end{aligned}$$

Since $\tau < \frac{1}{2}$, this implies (27). ∥

6.2 Proof of Proposition 1

Let Ξ be the set of admissible $\mathcal{E}$, i.e., the set of $n-$tuples of decreasing functions from $\mathbb{R}$ into $[q, 1-q]$, denoted by $\{\varepsilon(n;\xi), \ldots, \varepsilon(0;\xi)\}$, with the property that $\varepsilon(k+1;\xi) \le \varepsilon(k;\xi)$. We construct a function $H(\mathcal{E})$, defined on Ξ. For any admissible collection $\mathcal{E}$, calculate the probabilities $p_{\mathcal{E}}(\bar\theta|k)$ (whose existence is established in Lemma 1), insert these probabilities into the maximization problem given by (9)–(10), solve for equilibrium values $\varepsilon(k;\xi)$, and let these new values be $H(\mathcal{E})$.

We recall from Lemma 1 that, for any admissible collection, the posterior probability attached to $\bar\theta$ after an observation of $k+1$ values of $\bar z$ must exceed

[33] It would suffice for this last step that the slopes of $\Psi_{\mathcal{E}}(\psi_t, \overline{\theta})$ and $\Psi_{\mathcal{E}}(\psi_t, \underline{\theta})$ have an upper bound less than one.

that after an observation of k such values ($k = 0,\dots,n-1$), conditional on ξ, and that this posterior is increasing in ξ. This ensures that the mapping we have constructed will always choose new cutoffs that are decreasing in ξ and satisfy $\varepsilon(k+1;\xi) \leq \varepsilon(k;\xi)$, and hence the map is from the set Ξ into itself.

We next note that, for fixed probabilities $p_{\mathcal{E}}(\bar{\theta}|k;\xi)$, the objective contained in (9)–(10) are strictly concave, and hence the maximizer $\varepsilon(k;\xi)$ is unique, ensuring that our mapping is a function.

The set Ξ is a convex subset of a locally convex linear topological space (the metric space of $n-$-tuples of decreasing, bounded (by $[q;1-q]$) functions, with the metric induced by the L^1 norm:

$$||E,E'|| = \sum_{k=1}^{n} \int_{\xi=-\infty}^{\infty} |\varepsilon(k;\xi) - \varepsilon'(k;\xi)| d\eta$$

where η is the measure induced by the density $\frac{1}{2}g(\varepsilon;\theta)+\frac{1}{2}g(\varepsilon;\bar{\theta})$, which is the density of ξ conditional on θ being drawn from the stationary distribution.

The set Ξ is also compact. In particular, a subset of a metric space is compact if it is closed and sequentially compact. The set Ξ is clearly closed. To establish sequential compactness, fix a value of k and let $\{\varepsilon_i\}_{i=1}^{\infty}$ be a sequence of decreasing functions $\varepsilon_i(k;\xi)$. We show that there exists a converging subsequence. By successively taking subsequences, we can find a subsequence $\{\varepsilon_j\}_{j=1}^{\infty}$ that converges at every rational value of ξ, and let ε^*, defined only on the rationals, be the (pointwise) limit of this subsequence. Then ε^* must be decreasing. Extend ε^* to the reals in $\mathbb{R}$ by letting (for $\xi^* \in \mathbb{R}$) $\varepsilon^*(k;\xi^*) = \lim_{n\to\infty} \varepsilon^*(k;\xi_n(\xi^*))$, where $\xi_n(\xi^*)$ is a sequence of rationals approaching ξ^* from below. Then $\varepsilon^*(k;\xi)$ is decreasing.

We next note that the sequence of functions $\varepsilon_j(k;\xi)$ must converge to $F^*(k;\xi)$ almost everywhere. In particular, $F_j(k;\xi)$ must converge to $F^*(k;\xi)$ at any value ξ at which $F^*(k;\xi)$ is continuous. Since $F^*(k;\xi)$ is decreasing, it is continuous at all but countably infinitely many values, ensuring that the sequence $F_j(k;\xi)$ converges to $F^*(k;\xi)$ almost everywhere. Because each F_j and F^* are bounded above by the η-integrable function $\varepsilon(\xi) = 1-q$), Lebesgue's dominated convergence theorem (Billingsley [8, Theorem 16.4]) ensures that $||E,E'|| \to 0$.

It then remains to verify the continuity of the function H. To do so, we must show that the probabilities $\rho_{\mathcal{E}}(\bar{\theta}|k;\xi)$ are continuous in E, for which it suffices to note that the functions $\Psi_{\mathcal{E}}(\psi_t;\bar{\theta})$ and $\Psi_{\mathcal{E}}(\psi_t;\theta)$ move continuously (in the sup norm) in E.

This gives a continuous mapping from a convex subset of a locally convex linear topological space into itself. The Schauder-Tychonoff theorem

(Dunford and Schwartz [17, p. 456]) ensures that this mapping has a fixed point $\mathcal{E}^*$, and hence that an equilibrium exists.

It follows from the full-support assumptions and monotone-likelihood-ratio-property assumptions on $g(\xi;\underline{\theta})$ and $g(\xi;\overline{\theta})$ that $\varepsilon^*(k;\xi)$ is strictly decreasing in ξ. Now suppose that $\varepsilon^*(k+1;\xi) = \varepsilon^*(k;\xi)$. In order for these values to be optimal, it must be that the $k+1$st observation of $\overline{z}$ does not change the likelihood that the state if $\overline{\theta}$. But this can occur only if no additional observations of $\overline{z}$ are informative, i.e., only if $p_{\mathcal{E}^*}(\overline{\theta}|k+1;\xi) = p_{\mathcal{E}^*}(\overline{\theta}|k;\xi)$ for $k = 0,\ldots,n-1$. But then identical proportions of the agents choose $\overline{z}$ and $\underline{z}$, regardless of their observations, ensuring that their exist numbers $\overline{p}$ and $\underline{p} < \overline{p}$ such that proportion of surviving agents choosing $\overline{z}$ is given by $\overline{p}$ when the state is $\overline{\theta}$ and $\underline{p}$ when the state is $\underline{\theta}$. But then an observation of $\overline{z}$ yields a higher posterior probability of state $\overline{\theta}$, or $p_{\mathcal{E}^*}(\overline{\theta}|k+1;\xi) > p_{\mathcal{E}^*}(\overline{\theta}|k;\xi)$, a contradiction. ∎

6.3 Proof of Proposition 2

First, fix an admissible $\mathcal{E}$ and hence information updating rules $\rho_{\mathcal{E}}(\overline{\theta}|k;\xi)$ and $\hat{\rho}_{\mathcal{E}}(\overline{\theta}|k;\xi;\eta)$. Consider an agent who has observed k values of $\overline{z}$ and a signal $s \in \mathbb{R}$, where s may be a realization of either ξ or η. Let $\gamma_{\mathcal{E}}(k)$ be the posterior probability that the state is $\overline{\theta}$, given k observations of $\overline{z}$ (but ignoring any other signals). Then the agent's posterior $\hat{\rho}(\overline{\theta}|k;\xi;\eta)$ is given by (8), which we can rewrite as:

$$\rho_{\mathcal{E}}(\overline{\theta}|k;s) = \frac{g(s;\overline{\theta})\gamma_{\mathcal{E}}(k)}{g(s;\overline{\theta})\gamma_{\mathcal{E}}(k) + g(s;\underline{\theta})(1-\gamma_{\mathcal{E}}(k))}.$$

Nature, recognizing the possibility that the signal s is an uninformative realization of the random variable η, would instead assign posterior

$$\rho^N_{\mathcal{E}}(\overline{\theta}|k;s) = \frac{[(1-\lambda)g(s;\overline{\theta}) + \lambda f(s)]\gamma_{\mathcal{E}}(k)}{[(1-\lambda)g(s;\overline{\theta}) + \lambda f(s)]\gamma_{\mathcal{E}}(k) + [(1-\lambda)g(s;\underline{\theta}) + \lambda f(s)](1-\gamma_{\mathcal{E}}(k))}.$$

Dividing the numerator and denominator of this expression by $[(1-\lambda)g(s;\overline{\theta}) + \lambda f(s)]$, we obtain an expression to which (4) and (12) can be applied to conclude that[34]

$$0 < \frac{d\rho^N_{\mathcal{E}}(\overline{\theta}|k;s)}{ds} < \frac{d\rho_{\mathcal{E}}(\overline{\theta}|k;s)}{ds}. \tag{28}$$

[34]In particular, (4) and (12) imply that the term $[(1-\lambda)g(s,\underline{\theta})+\lambda f(s)]/[(1-\lambda)g(s,\overline{\theta})+\lambda f(s)]$ is decreasing in s, but at a slower rate than $g(s,\underline{\theta})/g(s,\overline{\theta})$, which suffices for (28).

Hence, higher signals increase the likelihood that the state is $\bar\theta$, but Nature reacts less sharply to the agent's information than does the agent, reflecting Nature's realization that the signal may be uninformative.

Condition (12) ensures that there are unique signals $\underline{s}$ and $\bar s > \underline{s}$ such that

$$g(\underline{s};\underline\theta) = f(\underline{s}); \qquad g(\bar s;\bar\theta) = f(\bar s):$$

This in turn ensures that there exists a value $\hat s \in [\underline{s};\bar s]$ such that

$$\rho_{\mathcal{E}}(\bar\theta|k;\hat s) = \rho^N_{\mathcal{E}}(\bar\theta|k;\hat s):$$

Let $\varepsilon^N_{\mathcal{E}}(k;s)$ be the value of ε^* that Nature would like to induce, given that the agent has observed k values of $\hat z$ and signal s. Let $\varepsilon_{\mathcal{E}}(k;s)$ be the value of ε^* that maximizes (13), and hence the value the agent will choose give utility function $\ln p(\varepsilon^*;\theta)$. Nature would chose $\varepsilon^N_{\mathcal{E}}(k;s)$ to solve

$$\max_{\varepsilon^*} \rho^N_{\mathcal{E}}(\bar\theta|k;s)\ln p(\varepsilon^*;\bar\theta) + (1-\rho^N_{\mathcal{E}}(\bar\theta|k;s))\ln p(\varepsilon^*;\underline\theta): \qquad (29)$$

This is Nature's counterpart of (7), given Nature's expectation conditional on k observations of $\bar z$ and the agent's signal s. The objective given in (29) is strictly concave in ε^* and is continuous in s (because $\rho^N_{\mathcal{E}}(\bar\theta|k;s)$ is) and hence the maximizer $\varepsilon^N_{\mathcal{E}}(k;s)$ is continuous and strictly decreasing in s.

Condition (28) implies that if Nature allows her agents to solve (13), then they will choose values of ε^* that are too small (from her point of view) whenever $s > \hat s$ and too large when $s < \hat s$. Let the value of ε^* chosen by the agent (and preferred by Nature) when $s = \hat s$ be denoted $\hat\varepsilon(k)$ $(= \varepsilon_{\mathcal{E}}(k;s) = \varepsilon^N_{\mathcal{E}}(k;s))$.

Notice next that, on the interval $[q;1-q]$, the derivatives $d\ln p(\varepsilon^*;\bar\theta)/d\varepsilon^*$ and $d\ln p(\varepsilon^*;\underline\theta)/d\varepsilon^*$ are both strictly decreasing, with

$$\frac{d\ln p(q;\underline\theta)}{d\varepsilon^*} = 0 \qquad \frac{d\ln p(1-q;\bar\theta)}{d\varepsilon^*} = 0:$$

Now for each signal s, define $\nu(\varepsilon^N_{\mathcal{E}}(k;s);k)$ to satisfy

$$\rho_{\mathcal{E}}(\bar\theta|k;s)\left(\frac{d(\ln p(\varepsilon^N(k;s);\bar\theta))}{d\varepsilon^*} + \nu(\varepsilon^N_{\mathcal{E}}(k;s);k)\right)$$

$$+(1-\rho_{\mathcal{E}}(\bar\theta|k;s))\left(\frac{d(\ln p(\varepsilon^N(k;s);\underline\theta))}{d\varepsilon^*} + \nu(\varepsilon^N_{\mathcal{E}}(k;s);k)\right) = 0: \qquad (30)$$

Because the left side of (30) is linear in ν, each $\nu(\varepsilon^N_{\mathcal{E}}(k;s);k)$ is well defined. Furthermore, because $\varepsilon^N_{\mathcal{E}}(k;s)$ is decreasing in s, we can view ν as a function º (²*; k) defined on the set of all values of ²* that are Nature's optimal choice $²^N_{\mathcal{E}}$(k; s) for k and some signal s. It follows from (30) that this function must be continuous in ²*. Then define a function u′(²*; k; µ) by

$$u'(²^*; k; \bar{µ}) = \frac{d \ln p(²^*; µ)}{d²^*} + º(²^*; k)$$
$$u'(²^*; k; µ) = \frac{d \ln p(²^*; \bar{µ})}{d²^*} + º(²^*; k):$$

We will then let the utility function u(²*; k; µ) be the integral of u′(²*; k; µ) with respect to ²*. (Integrability follows from the continuity of º(²*; k) in ²*.)

The agent's first-order necessary conditions for utility maximization, given utility function u(²*; k; µ), are given by (30) and are solved by Nature's optimum $²^N_{\mathcal{E}}$(k; s). The utility function u(²*; k; µ) thus achieves Nature's optimum if the latter uniquely solves these conditions, which we show below.

Next, let $½^N(\bar{µ}|²^*)$ be the value of the posterior ½ at which Nature would optimally induce choice ²* (and hence that causes the agent to choose ²*, given utility function u(²*; k; µ)). Recall that $Á(\bar{µ}|²^*)$ is the posterior that would cause the agent to choose ²*, given utility function ln p(²*; µ). Comparing Nature's optimization (29) with the agent's (11), condition (28) and the fact that ²̂(k) = $²_{\mathcal{E}}$(k; s) = $²^N_{\mathcal{E}}$(k; s) imply:

$$½^N(\bar{µ}|²^*) > (<)\ Á(\bar{µ}|²^*) \quad () \quad ²^* < (>)\ \hat{²}(k):$$

From (30), this ensures that

$$º(²^*) > (<)\ 0 \quad () \quad ²^* < (>)\ \hat{²}(k);$$

and hence that (14) holds.

For the agent's utility-maximization problem to implement Nature's optimum given utility function u(²*; k; µ), we need that the agent's first-order condition have a unique solution, for each posterior probability of $\bar{µ}$, within the range of values ²* that can arise as solutions to Nature's optimization problem. (Nature can ensure other values of ²* are not chosen by attaching large utility penalties to such values.) Suppose this is not the case, so that there are values ²′ and ²″ that both solve the first-order condition for some posterior ½′, the former being Nature's optimum for this posterior. But then ²″ must be Nature's optimum for some posterior ½″, ensuring that ²″ solves the first-order condition for posteriors ½′ and ½″, a contradiction to the fact that, conditional on ²*, the first order condition is linear in the posterior ½.

It remains to show that an equilibrium E exists, given that Nature response to E with an optimal utility function $u(\epsilon^*; k; \mu)$. Given the continuity of Nature's optimum in the signal s, this argument is a straightforward adaptation of Lemma 1 and Proposition 1. ||

References

[1] Elliott Aronson. The Social Animal. W. H. Freeman and Company, New York, 1995.

[2] Solomon E. Asch. Effects of group pressure on the modification and distortion of judgements. In Groups, Leadership and Men. Carnegie Press, Pittsburgh, 1951.

[3] Solomon E. Asch. Studies of independence and conformity: A minority of one against a unanimous majority. Psychological Monographs, 70(416), 1956.

[4] Robert B. Ash and Catehrine A. Doléans-Dade. Probability and Measure Theory. Academic Press, New York, 2000.

[5] Abhijit Banerjee. A simple model of herd behavior. Quarterly Journal of Economics, 107:797–817, 1992.

[6] Abhijit Banerjee and Drew Fudenberg. Word-of-mouth learning. Mimeo, MIT and Harvard University, 1995.

[7] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fasion, custom, and cultural exchange as information cascades. Journal of Political Economy, 100 992–1026, 1992.

[8] Patrick Billingsley. Probability and Measure. John Wiley and Sons, New York, 1986.

[9] David G. Blanchflower and Andrew Oswald. Well-being over time in britain and the USA. Mimeo, University of Warwick, 1999.

[10] William A. Brock and Steven N. Durlauf. Interactions-based models. SSRI Working Paper 9910R, University of Wisconsin, 2000.

[11] Brian Charlesworth. Evolution in Age-Structured Populations. Cambridge University Press, Cambridge, 1994.

[12] Robert B. Cialdini. Influence: Science and Practice. Scott, Foresman and Company, Boston, 1988.

[13] Andy Clark. Microcognition. MIT Press, Cambridge, Massachusetts, 1993.

[14] Hal L. Cole, George J. Mailath, and Andrew Postlewaite. social norms, savings behavior, and growth. Journal of Political Economy, 100 1092{1125, 1992.

[15] Giacomo Corneo. The efficient side of progressive income taxation. Working Paper 364, CESifo, Munich, Germany, 2000.

[16] W. Michael Cox and Richard Alm. Myths of Rich and Poor. Basic Books, New York, 1999.

[17] Nelson Dunford and Jacok T. Schwartz. Linear Operators Part I: General Theory. John Wiley and Sons, New York, 1988. Wiley Classics Library Edition.

[18] Steven N. Durlauf. A framework for the study of individual behavior and social interactions. SSRI Working Paper 2016, University of Wisconsin, 2000.

[19] J. S. Dusenberry. Income, Saving, and the Theory of Consumer Behavior. Harvard University Press, Cambridge, Massachussetts, 1949.

[20] R. A. Easterlin. Does economic growth improve the human lot? Some empirical evidence. In P. David and M. Reder, editors, Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz. Academic Press, New York, 1974.

[21] R. A. Easterlin. Will raising the incomes of all increase the happiness of all? Journal of Economic Behavior and Organization, 27 35{47, 1994.

[22] Glenn Ellison and Drew Fudenberg. Rules of thumb for social learning. Journal of Political Economy, 101 612{643, 1993.

[23] Glenn Ellison and Drew Fudenberg. Word-of-mouth communication and social learning. Quarterly Journal of Economics, 110 95{126, 1995.

[24] Robert H. Frank. Luxury Fever. Free Press, New York, 1999.

[25] Carl A. Futia. Invariant distributions and the limiting behavior of Markovian economic models. Econometrica, 50 377{408, 1982.

[26] Fred Hirsch. Social Limits to Growth. Harvard University Press, Cambridge, 1976.

[27] Ed Hopkins and Tatiana Kornienko. Consumer choice as a game of status. Mimeo, University of Edinburgh, 2000.

[28] R. Inglehart. The diminishing utility of economic grorwth: From maximizing security toward maximizing subjective well-being. Critical Review, 10 509{531, 1996.

[29] Daniel Kahneman and Amos Tversky. Judgement under uncertainty: Heuristics and biases. In Daniel Kahneman, Paul Slovic, and Amos Tversky, editors, On the Psychology of Prediction, pages 48{68. Cambridge University Press, Cambridge, 1982.

[30] Lawrence Katz and David Autor. Changes in the wage structure and earnings inequality. In Orley Ashenfelter and David Card, editors, Handbook of Labor Economics, Volume 3A, pages 1463{1555. North Holland, New York, 1999.

[31] Joseph LeDoux. The Emotional Brain. Simon and Schuster, New York, 1996.

[32] Lars Ljungqvist and Harald Uhlig. Tax policy and aggregate demand management under catching up with the joneses. american Economic Review, 90 356{366, 2000.

[33] Paul R. Milgrom. Good news and bad news: Representation theorems and applications. Bell Journal of Economics, 12 380{391, 1981.

[34] David Neumark and Andrew Postlewaite. Relative income concerns and the rise in married women's employment. Journal of Public Economics, 70 157{183, 1998.

[35] Andrew J. Oswald. Happiness and economic performance. Economic Journal, 107 1815{1831, 1997.

[36] Robert A. Pollak. interdependent preferences. American Economic Reivew, 66 309{320, 1976.

[37] Andrew Postlewaite. The social basis of interdependent preferences. European Economic Review, 42:779{800, 1998.

[38] Arthur J. Robson. A biological basis for expected and non-expected utility. Journal of Economic Theory, 68 397{424, 1996.

[39] Larry Samuelson and Jeroen Swinkels. Information and the evolution of the utility function. Mimeo, University of Wisconsin and Washington University in St. Louis, 2000.

[40] S. Tuljapurkar. Population Dynamics in Variable Environments. Springer-Verlag, Berlin, 1990.

[41] R. Veenhoven. Is happiness a trait? Social Indicators Research, 32:101–160, 1994.

[42] Richard L. Wheeden and Antoni Zygmund. Measure and Integral. Marcel Dekker Inc., New York, 1977.