Rational beliefs when the truth is not an option[*]

Filippo Massari[†] and Jonathan Newton[‡]

October 23, 2023

Abstract

In misspecified environments, should an economic agent act rationally towards optimizing some goal? If so, what should that goal be? Recent work has focused on the goal of bidirectional consistency of beliefs and actions, in effect finding a Nash equilibrium of an imaginary game in which one player chooses actions and another player chooses beliefs. In general, such outcomes maximize neither log-likelihood nor objective payoffs over the combined space of beliefs and actions. We suggest a goal and associated learning algorithm to maximize these latter quantities. When parameters are suitably chosen, this goal function selects models favored by evolutionary forces.

Keywords: misspecified learning, evolutionary models.

JEL Codes: C7, D8.

1. Introduction

Consider a decision maker (DM) whose action choice induces a distribution over outcomes. The DM receives a payoff that depends on the outcome. The DM

[*] Newton's work was supported by KAKENHI B funding from the Japan Society for the Promotion of Science (Grant 21H00695). We thank Murali Agastya, Tilman Börgers, Ken Binmore, Yi-Chun Chen, Omer Edhan, Takako Fujiwara-Greve, Sujoy Mukerji, Xiao Luo, Heinrich Nax, Lionel Page, Larry Samuelson, as well as seminar audiences at Keio University, City University of London, Bar Ilan University, University of Sydney, National University of Singapore, and the LEG2021 conference.

[†] Dipartimento di Scienze Economiche, Universita di Bologna, Piazza Scaravilli 2, Bologna, Italy e-mail: filippo.massari5@unibo.it; website: fmassari.com.

[‡] Institute of Economic Research, University of Kyoto, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan. e-mail: newton@kier.kyoto-u.ac.jp; website: jonathannewton.net.

has a set of possible models, with each model specifying a subjective distribution over outcomes for each possible action. If none of the models in this set specifies the true distributions, the DM's learning problem is said to be misspecified.[1]

What is a suitable solution concept for such problems? Esponda and Pouzo (2016) suggest Berk-Nash equilibrium (BNE), in which the DM chooses a model and a (possibly mixed) strategy such that

(C1) Actions in the support of the strategy are best responses to the model, and

(C2) The model maximizes expected log-likelihood given the strategy.

This can be interpreted as a Nash equilibrium of a game that takes place inside the mind of the DM. There are two players in this game: one who chooses a strategy to maximize expected payoff and another who chooses a model to maximize expected log-likehood. Each player takes the other player's choice as given.

Such an approach breaks the decision problem into two components and sets a goal of attaining consistency between them. The solutions that attain this consistency will not, in general, maximize log-likelihood or realized payoffs, even when we restrict attention to outcomes that satisfy (C1). Furthermore, there may exist outcomes that Pareto dominate the solution according to these criteria.

In this paper, we propose a goal function for a DM who cares about log-likelihood and realized payoffs. We show how such a goal can be attained using a generalization of Bayesian updating (Grünwald et al., 2017). Strikingly, our DM can obtain higher payoffs than solutions defined by consistency of (C1) and (C2), whilst simultaneously attaining greater accuracy (in a Bayesian sense). Pragmatically, the adoption of consistency-based solutions may be irrational in the sense that

> *"...a mode of behavior is irrational for a given decision maker, if, when the decision maker behaves in this mode and is then exposed to the analysis of her behavior, she feels embarrassed"*
>
> – Gilboa (2009, pp.139).

Considering fitness as a function of payoffs, we then describe how evolution acts on the model set, showing how selection differs according to whether agents

[1] Alternatively, one can consider a DM whose prior over models does not include the true (or equivalent) model in its support.

face individual or aggregate payoff uncertainty. In the former case, models favored by evolution maximize our goal function when it places all weight on payoffs rather than accuracy. In the latter case, payoffs are first subjected to a logarithmic transformation and there may exist fitness benefits from mixing between models.

Other recent work has examined the robustness of solutions for misspecified environments (Fudenberg and Lanzani, 2022; He and Libgober, 2021).[2] These models operate within the framework of BNE. In comparison, the current paper steps outside of this framework to examine alternative goals and learning rules. As such, it relates to work that considers the intersection of evolutionary incentives and learning algorithms (see, e.g. Edhan et al., 2017) and to a more functional, goals-based approach to understanding human behavior (Page, 2021) that seeks to understand beliefs through the lens of fitness maximization (Johnson and Fowler, 2011; Jouini et al., 2013; Frenkel et al., 2018; Heller, 2014).

In accepting (C1), the current paper follows not only BNE, but also other concepts that assume best responses to beliefs that are not necessarily correct, for example self-confirming equilibrium (Fudenberg and Levine, 1993), rationalizable conjectural equilibrium (Rubinstein and Wolinsky, 1994), mutually acceptable courses of action (Greenberg et al., 2009) and wishful thinking (Caplin and Leahy, 2019). Where these papers differ is in the restrictions placed upon beliefs.

The paper is organized as follows. Section 1.1 introduces key ideas via a simple illustrative example. Section 2 decribes the general model and introduces our goal function and solution concept. Section 3 presents results relating our concept to existing concepts. Section 4 continues the analysis through a series of examples. Section 5 gives learning and evolutionary foundations for our goal function. Section 6 extends our model to a multi-player environment. Section 7 concludes.

[2] At a further degree of separation, there is also non-evolutionarily based work such as Murooka and Yamamoto (2023); Esponda et al. (2021); Fudenberg et al. (2021); Frick et al. (2023).

1.1 Illustrative example

1.1.1 Coin tosses

Consider a coin which can land either heads (H) or tails (T) and has a true probability $p(H) = 0.7$ of landing heads. Every period, the DM earns a dollar if she correctly guesses the outcome of the coin toss. The DM considers two models,

$$\text{Blue model:} \quad p_{\text{blue}}(H) = 0.45, \qquad\qquad \text{Red model:} \quad p_{\text{red}}(H) = 0.9.$$

Condition (C1) specifies that a DM who follows each model should best respond,

$$\text{Blue model} \;\Rightarrow \text{bet on } T, \qquad\qquad \text{Red model} \;\Rightarrow \text{bet on } H.$$

Expected log-likelihoods given the true probability $p(H) = 0.7$ are

$$LL(\text{blue}) = -0.738, \qquad\qquad LL(\text{red}) = -0.765.$$

Given these log-likelihoods, (C2) implies adoption of the blue model, the model that would be learned by standard Bayesian updating (Berk, 1966). Combining, the unique BNE is for the DM to follow the blue model and bet on tails.

Noting that the average realized payoff of 0.3 from following the blue model is less than the average realized payoff of 0.7 from following the red model, we propose a DM who learns a model to maximize some weighted sum of payoffs and likelihood. If she puts all weight on likelihood, she will learn the blue model and bet on tails. This is identical to BNE (Proposition 2). If she puts all weight on payoff, she will learn the red model and bet on heads. This is identical to playing a Nash equilibrium from the set of actions that can be justified by some model in the model set (Proposition 3). For intermediate weightings, if payoff is weighted lower than some threshold, the blue model will be chosen, and if payoff is weighted higher than the threshold, the red model will be chosen.

As a comparison, consider adding the true model $p(H) = 0.7$ to the model set so that the problem becomes well-specified. Any weighting that puts non-zero weight on likelihood will then select the true model. Furthermore, any weighting will select models with the same best response as the true model (Proposition 4). Of particular note, if all weight is placed on payoff, the DM is indifferent between the red model and the true model.

1.1.2 Improvement in both dimensions

Consider a slight perturbation of the above problem. The true model is the same, but we adjust red and blue so that they assign slightly different probabilities depending on the DM's bet.

$$\text{Blue:} \quad p_{\text{blue}}(H|\text{ bet on } T) = 0.45, \qquad p_{\text{blue}}(H|\text{ bet on } H) = 0.47,$$

$$\text{Red:} \quad p_{\text{red}}(H|\text{ bet on } T) = 0.9, \qquad p_{\text{red}}(H|\text{ bet on } H) = 0.88.$$

As before, (C1) dictates that a DM who follows the blue model should bet on tails, whereas a DM who follows the red model should bet on heads. Calculating expected log-likelihood for each model-action pair, we obtain

$$LL(\text{blue}, H) > LL(\text{red}, H) > LL(\text{blue}, T) > LL(\text{red}, T).$$

Not only is the payoff from (red, H) greater than the payoff from (blue, T), but log-likelihood is also greater. That is, of the two model-action pairs that satisfy (C1), (red, H) is Pareto superior to (blue, T), exhibiting more accurate beliefs and higher realized payoffs.

In comparison, the unique BNE remains (blue, T). The reason is that *fixing* either action (H or T), log-likelihood is maximized by the blue model. As before, (C2) rules out the red model. There is a trade-off between the form of rationality embodied in (C2) and the rationality of a DM who values accuracy and payoffs.

We show in Section 4 that similar analysis applies to the canonical misspecified monopolist example from the literature (Nyarko, 1991; Esponda and Pouzo, 2016; Fudenberg and Lanzani, 2022). Furthermore, goal functions to optimize accuracy and payoffs can be learned using a generalization of Bayesian updating (Proposition 6). When weighted towards payoffs, these goal functions select models favored by evolutionary forces (Proposition 7).

2. Model

A decision maker (DM) faces the following problem. A state ω is randomly chosen from a set of states Ω according to probability distribution p. The DM chooses an action x from a finite set of actions $x \in \mathbb{X}$. A function $f : \mathbb{X} \times \Omega \to \mathbb{Y}$ then determines a consequence $y = f(x, \omega)$ from a set of possible consequences $\mathbb{Y}$. The DM's payoff is then given by payoff function $\pi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$. Let $Q(\cdot|x)$ denote the true distribution over consequences conditional on action x being chosen.

Θ is the parameter set. Each $\theta \in \Theta$ indexes a possible model that the DM can learn. Specifically, each θ is associated with a subjective distribution over consequences $Q_\theta(\cdot|x)$ conditional on each action x being chosen. If there is no θ such that $Q_\theta(\cdot|x) = Q(\cdot|x)$ for all x, we say that the problem is *misspecified*. Otherwise, the problem is *well-specified*. The set of *best responses* induced by θ is given by

$$X^*(\theta) = \underset{x \in \mathbb{X}}{\operatorname{argmax}} E_{Q_\theta(\cdot|x)} \pi(x, Y).$$

In our search for solutions, we restrict attention to model-action pairs such that the action is a best response to the model. That is, we assume the condition we refer to as (C1) in the introduction. This is uncontroversial to the extent that the functional implication of a model is that it leads the DM to select some set of actions. Referring to such actions as 'best responses' is consistent with an interpretation of actions as revealed (subjective) preferences. Formally, we consider

$$\Lambda = \{(\theta, x) : \theta \in \Theta, x \in X^*(\theta)\}.$$

Note that every $\theta \in \Theta$ appears in at least one element of Λ, but the same is not true of $x \in \mathbb{X}$. If x is not a best response for any model, then it will not be part of any element of Λ. Conversely, the same actions can occur in multiple elements of Λ. If x is a best response to some $\theta \in \Theta$, we say that x is *justifiable*. Given a model-action pair $(\theta, x) \in \Lambda$, the objective (expected) payoff is

$$\Pi(\theta, x) = E_{Q(\cdot|x)} \pi(x, Y).$$

Note that $\Pi(\theta, x)$ is not directly affected by θ. Another quantity that we wish to consider is the (expected) log-likelihood of pairs $(\theta, x) \in \Lambda$,

$$LL(\theta,x) = E_{Q(\cdot|x)} \log Q_\theta(Y|x).$$

We consider a DM who wishes to maximize a goal function that combines objective payoff and log-likelihood. The exact specification does not matter, but for the sake of expositional clarity, we take

$$G(\theta,x) = \alpha\Pi(\theta,x) + (1-\alpha)LL(\theta,x),$$

where $\alpha \in [0,1]$ determines the relative weighting of payoff and log-likelihood in the goal function.[3,4,5]

Again for expositional simplicity, assume that for all justifiable $x \in \mathbb{X}$, $\max_{\theta: x \in X^*(\theta)} LL(\cdot,x)$

[3] The negative of $LL(\theta,x)$ is known as the *cross entropy* of Q and Q_θ, and can be written

$$-LL(\theta,x) = -E_{Q(\cdot|x)} \log Q_\theta(Y|x) = \underbrace{E_{Q(\cdot|x)} \log \frac{Q(Y|x)}{Q_\theta(Y|x)}}_{=:D(Q(\cdot|x)||Q_\theta(\cdot|x))} \underbrace{-E_{Q(\cdot|x)} \log Q(Y|x)}_{=:H(Q(\cdot|x))}.$$

Kullback-Leibler divergence $D(Q||Q_\theta)$ measures the difference between Q_θ and Q. *Shannon entropy* $H(Q)$ measures the level of uncertainty in the true distribution Q. Substitution gives

$$G(\theta,x) = \alpha\Pi(\theta,x) - (1-\alpha)D(Q(\cdot|x)||Q_\theta(\cdot|x)) - (1-\alpha)H(Q(\cdot|x)).$$

Similar to $\Pi(\theta,x)$, the Shannon entropy term $H(Q(\cdot|x))$ is not directly affected by θ. Consequently, for given α, our specification is equivalent to decreasing the payoff from x by the Shannon entropy and using Kullback-Leibler divergence alone as a measure of accuracy.

[4] For fixed x, log-likelihood and Kullback-Leibler divergence give identical orderings of models in terms of accuracy (see Footnote 3). Across x, this equivalence fails, so the informational differences between log-likelihood and Kullback-Leibler divergence become salient. In particular, $LL(\theta,x)$ can be estimated using a simple empirical mean of $\log Q_\theta(\cdot|x)$, whereas estimating $D(Q(\cdot|x)||Q_\theta(\cdot|x))$ additionally requires an estimate of the true distribution $Q(\cdot|x)$. If such information can be used, it calls into question the necessity of a model of misspecification in the first place. Nevertheless, any reader who strongly prefers Kullback-Leibler divergence to log-likelihood should feel free to substitute the former for the latter in our goal function.

[5] Our goal function is similar to the *wishful thinking* goal function of Caplin and Leahy (2019),

$$E_{Q_\theta(\cdot|x)}\pi(x,Y) - \rho D(Q(\cdot|x)||Q_\theta(\cdot|x)).$$

with the crucial difference that where we have $\Pi(\theta,x) = E_{Q(\cdot|x)}\pi(x,Y)$, the cited paper has $E_{Q_\theta(\cdot|x)}\pi(x,Y)$. The wishful thinking DM desires a high *subjective* payoff and accurate beliefs, whereas our goal function favours a high *objective* payoff and accurate beliefs. This makes sense. The cited model is, after all, a model of wishful thinking, whereas our focus is on normative considerations and evolutionary foundations. Indeed, the relationship between objective payoffs and evolution is considered in Section 5.

exists. We suggest the solution set

$$\Lambda^* = \arg \max_{(\theta,x)\in\Lambda} G(\theta,x). \tag{1}$$

Let $G^* = \max_{(\theta,x)\in\Lambda} G(\theta,x)$. If $\alpha = 1$, then the DM's goal is to maximize payoff.[6] If $\alpha = 0$, then the DM's goal is to maximize log-likelihood. For given values of α, we will occasionally denote the solution set by Λ^*_α.

3. Initial analysis

To get a feeling for the concepts defined in the model section, we shall examine some edge cases. As a point of comparison, it suits to formally define the alternative approach of Nyarko (1991) and Esponda and Pouzo (2016), in which the DM chooses a model while keeping x fixed. Specifically, consider the solution set

$$\Lambda^{'} = \left\{ (\theta',x) \in \Lambda : \theta' \in \arg\max_{\theta\in\Theta} LL(\theta,x) \right\}. \tag{2}$$

Considering the definitions of Λ and Λ', we see that Λ' is effectively the set of Nash equilibria of a game between one player who chooses x to be a best response to θ and another player who chooses θ to maximize $LL(\theta,x)$ given x. Elements of Λ' are pure *Berk-Nash equilibria* (BNE). Comparing to Λ^*, we see that Λ^* involves holistic choice of model and action by the DM. In contrast, Λ' involves separate modules that optimize for model and action respectively, each taking the other module's choice as given.

The example of Section 1.1 shows that solutions in Λ' can be Pareto-dominated by solutions in Λ^*. The definitions of Λ^* and Λ' imply that the opposite is not possible. A solution in Λ^* will never be Pareto-dominated by a solution in Λ'.

Proposition 1. *Let* $(\theta^*,x^*) \in \Lambda^*_\alpha$ *and* $(\theta',x') \in \Lambda'$.

(i) $\alpha \in (0,1)$ *and* $LL(\theta^*,x^*) < LL(\theta',x') \implies \Pi(\theta^*,x^*) > \Pi(\theta',x')$,

[6] For $\alpha = 1$, actions played in elements of Λ^* are those that would be learned by a belief-free robust learning algorithm over $\{x : \exists\theta \text{ such that } (\theta,x) \in \Lambda\}$.

(ii) $\alpha \in (0,1)$ *and* $\Pi(\theta^*, x^*) < \Pi(\theta', x') \implies LL(\theta^*, x^*) > LL(\theta', x')$,

(iii) $\alpha = 0 \implies LL(\theta^*, x^*) \geq LL(\theta', x')$,

(iv) $\alpha = 1 \implies \Pi(\theta^*, x^*) \geq \Pi(\theta', x')$.

Define the set of pure *justifiable Nash equilibria* (JNE), the set of Nash equilibria when the DM is is restricted to only choose actions that are justifiable according to some model in Θ.

$$(3) \qquad \mathscr{J} = \underset{\substack{x \in \mathbb{X}, \\ x \text{ is justifiable}}}{\operatorname{argmax}} \; E_{Q(\cdot|x)} \pi(x, Y).$$

In analyzing Λ^*, the first case we consider is $\alpha = 0$. It turns out that if the true distribution and subjective distributions over consequences are independent of x, then $\Lambda^*_{\alpha=0}$ and Λ' are equivalent. That is, a DM who is only concerned with maximizing log-likelihood will play a pure BNE. To formalize, define conditions

(I1) $Q(\cdot|x)$ is constant in x.

(I2) For all $\theta \in \Theta$, $Q_\theta(\cdot|x)$ is constant in x.

Condition (I1) corresponds to the true distribution over consequences being independent of the actions of the DM. Condition (I2) corresponds to the DM modelling consequences as being independent of actions. Under these conditions, log-likelihood $LL(\theta, x)$ is independent of x. It follows that it does not matter whether maximization of $LL(\theta, x)$ happens over pairs (θ, x) as required for $\Lambda^*_{\alpha=0}$, or over θ while keeping x fixed as required for Λ'.

Proposition 2. *If (I1) and (I2) hold, then* $\Lambda^*_{\alpha=0} = \Lambda'$.

In general, when consequences are not independent of actions, sets $\Lambda^*_{\alpha=0}$ and Λ' differ. In particular, solutions in $\Lambda^*_{\alpha=0}$ may be strictly more accurate than solutions in Λ'. The reason is that Λ' contains models that maximize log-likelihood for fixed actions, whereas $\Lambda^*_{\alpha=0}$ maximizes log-likelihood across all model-action pairs. Examples where these sets are not equal will be given in Section 4.[7]

Having considered $\alpha = 0$, we turn to the opposed case $\alpha = 1$. In this case, the DM's goal is to maximize objective payoffs $\Pi(\theta,x)$ over pairs $(\theta,x) \in \Lambda$. As θ does not directly affect $\Pi(\theta,x)$, the only role of Θ is to restrict the possible values of x in $(\theta,x) \in \Lambda$ to those that are justifiable by some $\theta \in \Theta$. Consequently, when $\alpha = 1$, the solution set $\Lambda^*_{\alpha=1}$ corresponds to the set of JNE.

Proposition 3. $(\theta^*,x^*) \in \Lambda^*_{\alpha=1}$ *if and only if* $x^* \in \mathscr{J}$.

Note that Proposition 3 does not require independence assumptions. Further, if every action is justifiable, then $\Lambda^*_{\alpha=1}$ corresponds to the pure Nash equilibria of the game with actions $\mathbb{X}$ and payoffs $E_{Q(\cdot|x)}\pi(x,Y)$. This is true regardless of whether the problem is well-specified. That is, the actions played in $\Lambda^*_{\alpha=1}$ depend on the set of justifiable actions, but not on how those actions are justified. If a particular action leads to the highest payoffs, it will be played in $\Lambda^*_{\alpha=1}$ regardless of whether it is justified by correct beliefs or by completely erroneous beliefs.

Taking Propositions 2 and 3 together we see that, under independence assumptions (I1) and (I2), solution set Λ^* varies between a Berk-Nash equilibrium and a justifiable Nash equilibrium as α varies from 0 to 1. If, in addition to independence, we assume that the problem is well-specified, then these solutions converge.

Proposition 4. *Let the problem be well-specified so that there exists* $\theta^\dagger \in \Theta$ *such that for all x,* $Q_{\theta^\dagger}(\cdot|x) = Q(\cdot|x)$. *If (I1) and (I2) hold, then for all* $x \in X^*(\theta^\dagger)$, *for all* α, *we have* $(\theta^\dagger,x) \in \Lambda^*_\alpha$. *Conversely, if* $(\theta^*,x^*) \in \Lambda^*_\alpha$, *then* $x^* \in X^*(\theta^\dagger)$, *and if* $(\theta^*,x^*) \in \Lambda^*_{\alpha<1}$, *then* $Q_{\theta^*}(\cdot|x) = Q_{\theta^\dagger}(\cdot|x)$ *for all* $x \in \mathbb{X}$.

That is, in well-specified problems in which consequences are independent of the DM's actions, Λ^*, Λ' and $\mathscr{J}$ all give the same predictions in terms of actions, although the models that support these actions may differ.

[7] Note that our description of BNE as a Nash equilibrium internal to the DM can be extended to $\Lambda^*_{\alpha=0}$ in the following way. Consider a *sequential* game in which one player who wishes to maximize $LL(\cdot,\cdot)$ chooses $\theta \in \Theta$, following which the other player best responds to θ. Solutions $(\theta^*,x^*) \in \Lambda^*_{\alpha=0}$ will be subgame perfect equilibria of this game. If best responses are unique, they will be the only subgame perfect equilibria. This analogy ceases to hold for $\alpha > 0$.

4. Examples

4.1 Coin tosses

Here we discuss the illustrative example from the introduction. A decision maker guesses the outcome of a coin toss, $\mathbb{X} = \{H, T\}$. The outcome of the coin toss is independent of the decision maker's action and is given by $y = f(x, \omega) = \omega$, where $\omega = H$ with probability 0.7 and $\omega = T$ with probability 0.3. Hence we have $Q(H|x) =: Q(H) = p(H) = 0.7$ for all x. Payoffs are given by $\pi(x,y) = 1$ if $x = y$ and $\pi(x,y) = 0$ if $x \neq y$. The parameter set is $\Theta = \{\theta^1, \theta^2\}$ and we let $Q_{\theta^1}(H|x) =: Q_{\theta^1}(H) = 0.45$ and $Q_{\theta^2}(H|x) =: Q_{\theta^2}(H) = 0.9$ for all x. Note that T is the unique best response to beliefs Q_{θ^1}, whereas H is the unique best response to beliefs Q_{θ^2} or to the true model Q. Therefore, we have $\Lambda = \{(\theta^1, T), (\theta^2, H)\}$ with objective payoffs $\Pi(\theta^1, T) = 0.3$ and $\Pi(\theta^2, H) = 0.7$. However, θ^1 is more accurate than θ^2 in the sense that $LL(\theta^1, T) > LL(\theta^2, H)$. Combining,

$$G(\theta^1, T) \geq G(\theta^2, H) \quad \iff \quad \alpha \leq \frac{\log \frac{1331}{1024}}{4 + \log \frac{1331}{1024}}.$$

So for low values of α, the DM prioritizes accuracy over payoffs and the solution set $\Lambda^* = \{(\theta^1, T)\}$. For high values of α, the DM values payoffs enough that the solution set is $\Lambda^* = \{(\theta^2, H)\}$. Comparing to the propositions in the previous section, we see this is consistent with $\alpha = 0$ giving Berk-Nash equilibria and $\alpha = 1$ giving justifiable Nash equilibria. That is, consistent with Proposition 2, we have $\Lambda^*_{\alpha=0} = \Lambda' = \{(\theta^1, T)\}$. Consistent with Proposition 3, we have $\Lambda^*_{\alpha=1} = \{(\theta^2, H)\}$ and $\mathscr{J} = \{H\}$.

Consider adding the true model $\theta^\dagger$, $Q_{\theta^\dagger}(H|x) =: Q_{\theta^\dagger}(H) = 0.7$, to the parameter set so that $\Theta = \{\theta^1, \theta^2, \theta^\dagger\}$. Consistent with Proposition 4, we obtain $\Lambda^*_\alpha = \{(\theta^\dagger, H)\}$ for $\alpha < 1$. Proposition 2 then implies that $\Lambda' = \{(\theta^\dagger, H)\}$.

For $\alpha = 1$, the goal function places no weight on accuracy and all weight on payoffs, so we obtain $\Lambda^*_{\alpha=1} = \{(\theta^\dagger, H), (\theta^2, H)\}$. That is, because $\theta^\dagger$ and θ^2 induce identical actions, the DM who cares only about payoffs is indifferent between them. Proposition 3 then implies that $\mathscr{J} = \{H\}$.

4.2 Arrow-Debreu securities

We extend the example of the preceding subsection so that the DM chooses shares $x \in \mathbb{X} = \{0, 0.01, \ldots, 0.99, 1\}^n$, $\sum_{j=1}^n x_j = 1$, of a unit of Arrow-Debreu security to invest in outcomes $H_1, \ldots, H_n$. Similar to before, $y = f(x, \omega) = \omega$, where $\omega = H_j$ with probability p_{H_j}. The DM's action does not affect outcome probabilities. Hence we have $Q(H_j|x) = Q(H_j) = p(H_j)$ for all x, j. The DM is aware that his action does not affect outcome probabilities and has Bernoulli beliefs parametrized by $\Theta \subseteq \mathbb{X}$, so that $\forall \theta \in \Theta, Q_\theta(H_j) = \theta_j$. Payoffs are given by $\pi(x, H_j) = u(x_j)$ for all j, where u is a utility function.

Proposition 4 tells us that if the problem is well-specified, that is if there exists $\theta^\dagger \in \Theta$ such that $Q_{\theta^\dagger}(H_j) = \theta_j^\dagger = p(H_j) = Q(H_j)$ for all H_j, then for $0 < \alpha < 1$, we have $\Lambda_\alpha^* = \{(\theta^\dagger, x) : x \in X^*(\theta^\dagger)\}$.

If, however, the problem is misspecified, then different values of α will, in general, give different solution sets. An interesting exception is log utility.

Proposition 5. *If $u(\cdot) = \log(\cdot)$, then Λ^* is the same for all values of α.*

Proof of Proposition 5.
Substituting into the definitions of $\Pi(\theta, x)$ and $LL(\theta, x)$,

$$\Pi(\theta, x) = E_{Q(\cdot)}\, \pi(x, Y) = \sum_{j=1}^n p(H_j)\, u(x_j), \tag{4}$$

$$LL(\theta, x) = E_{Q(\cdot)} \log Q_\theta(Y) = \sum_{j=1}^n p(H_j) \log \theta_j. \tag{5}$$

For all $\theta \in \Theta$, best responses $x \in X^*(\theta)$ maximize

$$E_{Q_\theta(\cdot)}\, \pi(x, Y) = \sum_{j=1}^n Q_\theta(H_j)\, u(x_j) = \sum_{j=1}^n \theta_j \log x_j, \tag{6}$$

such that $\sum_{j=1}^n x_j = 1$. Solving this, we obtain $x_j = \theta_j$ for all j.

Substituting $x_j = \theta_j$ and $u(\cdot) = \log(\cdot)$ into the right hand side of (4), we obtain an expression identical to the right hand side of (5). So for all α,

$$(7) \qquad G(\theta,x) = \alpha\,\Pi(\theta,x) + (1-\alpha)\,LL(\theta,x) = \sum_{j=1}^{n} p(H_j)\,\log\theta_j,$$

which is independent of α. Therefore, $\Lambda^* = \text{argmax}_{(\theta,x)\in\Lambda}\, G(\theta,x)$ is also independent of α. □

This result is a consequence of the celebrated Kelly criterion which states that, under log utility, an optimizing investor with model θ will invest a share in H_j that equals the subjective probability $Q_\theta(H_j)$ of H_j. As a consequence of this investment behavior, the realized payoff when H_j occurs equals the log-likelihood of H_j. That is, under these conditions, payoffs equal log-likehoods and therefore the tension between maximizing payoff and maximizing log-likelihood is resolved without recourse to any weighting parameter α.

4.3 Well-specified model without independence

Consider the following simple example of a well-specified model in which (I1) and (I2) do not hold. There are two states $\Omega = \{0,1\}$, $p(0) = p(1) = 1/2$, two actions $\mathbb{X} = \{a,b\}$ and two possible consequences $\Upsilon = \{0,1\}$. Action a leads to consequence 1 with certainty, $f(a,0) = f(a,1) = 1$. Action b leads to a consequence equal to the state, $f(b,\omega) = \omega$. There is only one model in the parameter set, $\Theta = \{\theta^\dagger\}$, and this is the true model $Q_{\theta^\dagger}(\cdot|x) = Q(\cdot|x)$.

Let payoffs be independent of actions and consequences, $\pi(\cdot,\cdot) \equiv 0$, so that $\Pi(\cdot,\cdot) \equiv 0$. Calculating log-likelihoods,

$$LL(\theta^\dagger,a) = E_{Q(\cdot|a)} \log Q(Y|a) = \log \underbrace{Q(1|a)}_{=1} = 0,$$

$$LL(\theta^\dagger,b) = E_{Q(\cdot|b)} \log Q(Y|b) = \frac{1}{2}\log \underbrace{Q(0|b)}_{=1/2} + \frac{1}{2} log \underbrace{Q(1|b)}_{=1/2} = \log\frac{1}{2}.$$

Consequently, for any $\alpha < 1$, we have that $\Lambda^*_\alpha = \{(\theta^\dagger,a)\}$. By definition, $\Lambda' = \{(\theta^\dagger,a),(\theta^\dagger,b)\}$, so $\Lambda^*_{\alpha=0} \neq \Lambda'$, in contrast to Proposition 2. In addition, note that best responses to $\theta^\dagger$ are $X^*(\theta^\dagger) = \{a,b\}$, yet $(\theta^\dagger,b) \notin \Lambda^*_{\alpha<1}$, in contrast to Proposition 4.

4.4 Misspecified model without independence

Here we consider an example from Nyarko (1991) as adapted by Esponda and Pouzo (2016). A monopolist chooses a price $x \in \mathbb{X} = \{2, 10\}$ that generates demand $d = \phi_0(x) + \omega$, where $\omega \sim N(0, 1)$. It is assumed that $\phi_0(2) = 34$ and $\phi_0(10) = 2$. The outcome $y = f(x, \omega) = d$ and payoff is $\pi(y) = xy$.

The monopolist describes demand using a parametric model $d = f_\theta(x, \omega) = a - bx + \omega$, where $\theta = (a, b) \in \Theta$ is a parameter vector and $\omega \sim N(0, 1)$. The set of possible models is given by $\Theta = [33, 40] \times [3, 3.5]$.

Let $\theta_0 \in \mathbb{R}^2$ provide a perfect fit for the demand so that $\phi_0(x) = \phi_{\theta_0}(x)$ for all $x \in \mathbb{X}$. This gives $\theta_0 = (a_0, b_0) = (42, 4) \notin \Theta$ and therefore the monopolist has a misspecified model. Note that $Q(\cdot|x)$ is normal with mean $\phi_0(x)$ and unit variance. Similarly, $Q_\theta(\cdot|x)$ is normal with mean $\phi_\theta(x) = a - bx$ and unit variance.

4.4.1 Maximizing the goal function

By substituting the true model parameters into the payoff function, we obtain expected objective payoffs from (θ, x),

$$\begin{aligned} \Pi(\theta, x) &= E_{Q(\cdot|x)}[\pi(x, Y)] = E_p[x(42 - 4x + \omega)] \\ &= x(42 - 4x) + xE_p[\omega] = \begin{cases} 68 & \text{if } x = 2 \\ 20 & \text{if } x = 10 \end{cases} \end{aligned} \tag{8}$$

If $\alpha = 1$, then this is the end of the story. As $x = 2$ leads to a higher payoff than $x = 10$, the solution set is $x = 2$ combined with any beliefs to which $x = 2$ is a best response. That is,

$$\Lambda^* = \left\{(\theta, 2) : \theta = (a, b), \frac{a}{b} \leq 12\right\}.$$

If $\alpha < 1$, then for either of the x that could be chosen, we find the $(\theta, x) \in \Lambda$ that maximizes $LL(\theta, x)$. For $x = 2$, we obtain $\theta = (40, {}^{10}\!/\!_{3})$. In words, of all parameters in Θ that induce 2 as a best response, log-likelihood is maximized by $\theta = (40, {}^{10}\!/\!_{3})$. For $x = 10$, we obtain $\theta = (36, 3)$. That is, of all parameters in Θ that induce 10 as a best response, log-likelihood is maximized by $\theta = (36, 3)$. Calculating,

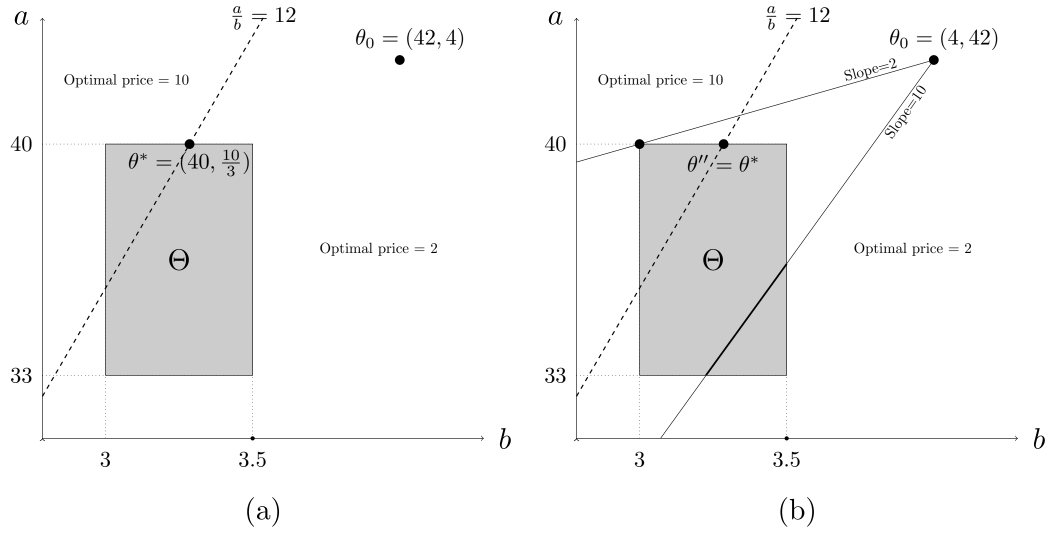

Figure 1: **The misspecified monopolist.** Set Θ does not contain the true parameters θ_0. Line $a/b = 12$ is the boundary between the regions in which $x = 10$ and $x = 2$ are best responses. **Panel (a).** For $\alpha = 1$, the solution set Λ^* is the entire region for which $x = 2$ is a best response. For $\alpha < 1$, $\Lambda^* = \{(\theta^*, 2)\}$. **Panel (b).** Maximizing $LL(\theta, 2)$ over Θ, we obtain $\theta = (40, 3)$, at which $x = 2$ is not a best response. Maximizing $LL(\theta, 10)$ over Θ, we obtain the the thick black line, on which $x = 10$ is not a best response. Therefore, $\Lambda' = \varnothing$. Fixing $\sigma(2) = 35/36$ and $\sigma(10) = 1/36$, expected log-likelihood is maximized at $\theta'' = \theta^*$, therefore $\Lambda'' = \{\theta'', (35/36, 1/36)\}$.

$$LL((40, 10/3), 2) > LL((36, 3), 10).$$

Combining $\Pi(\theta, x)$ and $LL(\theta, x)$, we see that $G(\theta, x)$ is uniquely maximized at $(\theta^*, x^*) = ((40, 10/3), 2)$ and therefore solution set Λ^* is a singleton.

$$\Lambda^* = \left\{ \left(\left(40, \frac{10}{3} \right), 2 \right) \right\}.$$

4.4.2 Comparison to the internal consistency approach

Consider $\alpha < 1$. The argument of Nyarko (1991) is that if a Bayesian monopolist consistently chooses action 2, then he will learn a model to which the only best response is action 10. Specifically, $\theta = (40, 3)$ maximizes $LL(\cdot, 2)$ over Θ and the unique best response to $\theta = (40, 3)$ is to play $x = 10$. Conversely, if he consistently chooses action 10, then he will learn a model to which the unique best response is action 2 (see Figure 1). In conclusion, $\Lambda' = \varnothing$.

Esponda and Pouzo (2016) solve the non-existence problem by allowing the monopolist to mix between actions 2 and 10. Let strategy $\sigma \in \Delta(\mathbb{X})$ be a probability distribution over actions. Define BNE in mixed strategies as

$$\Lambda'' = \Big\{ (\theta'', \sigma) : \sigma(x) > 0 \Rightarrow (\theta'', x) \in \Lambda \qquad \text{and} \quad \theta'' \in \arg\max_{\theta \in \Theta} \sum_{x \in \mathbb{X}} \sigma(x)\, LL(\theta, x) \Big\}. \tag{9}$$

The idea is that if the DM mixes between 2 and 10 in the right proportions, then he can learn a model to which both actions are a best response. Specifically, he must learn $\theta = (a, b)$ such that $a/b = 12$. For such a model, specifically $\theta'' = (40, \frac{10}{3})$, to maximize expected log-likelihood given mixing proportions, we require $\sigma = (\sigma(2), \sigma(10)) = (35/36, 1/36)$. Hence,

$$\Lambda'' = \left\{ \left(\left(40, \frac{10}{3} \right), \left(\frac{35}{36}, \frac{1}{36} \right) \right) \right\}.$$

At first glance, the internal consistency approach underpinning Λ' and Λ'' appears to model a DM who cares about payoffs (solutions must be in Λ) as well as about accuracy (solutions maximize log-likelihood given x or σ). However, as noted in Section 3, such solutions are effectively finding Nash equilibria of a game between two modules within the decision maker, and as such involve a disconnect between (i) the idea of payoff maximization that underpins Nash-style concepts, and (ii) the choice of beliefs, which ignores payoffs and maximizes log-likelihood.

The monopolist example of this section makes an even stronger point. Even if the DM cares about accuracy as well as payoffs, the solutions in Λ' or Λ'' may do a bad job with respect to both. Indeed, such solutions may fail to be even locally Pareto efficient. Consider the solution in Λ'' and consider a marginal increase in $\sigma(x)$, holding θ fixed. By (8), expected objective payoff will increase. Furthermore, calculations reveal that $LL(\theta, x)$ will increase.[8] That is, by increasing $\sigma(x)$, both

[8] Also, the Kullback-Leibler divergence of Q_θ from the true distribution will decrease.

payoffs and log-likelihood are increased. If we continue to increase $\sigma(2)$, these improvements continue until we reach $\sigma(2) = 1$, the unique solution in Λ^*.[9,10]

What has been sacrificed to obtain these improvements? The answer is "rationality" in the specific sense that for fixed $\sigma(x) > 35/36$, the model $\theta = (40, 10/3)$ does not maximize log-likelihood. Conversely, in the sense of the quote we give in the introduction, the solution in Λ'' could be considered irrational, as it is predicated upon randomization between two actions and learning from outcomes, but never recognizing that the expected payoff from one action is higher than the expected payoff from the other action.

5. Learning

Consider a DM who is part of a (possibly infinite) population. In every period $t \geq 1$, for every $(\theta, x) \in \Lambda$, there are some members of the population who follow the model θ and play x. Every member of the population experiences the same sequence of states $\{\omega_t\}_t$. Assume a uniform bound on the absolute value of log-likelihoods. The DM learns from the realized payoffs and realized log-likelihoods of players in the population. For each $(\theta, x) \in \Lambda$, the DM's *generalized likelihood* (see Grünwald et al., 2017) after t periods is

[9] In general, it is clear that we cannot increase the expected value of our goal function $G(\cdot,\cdot)$ by mixing between actions or model-action pairs. Indeed, for pairs (θ^1, x^1) and (θ^2, x^2) to give the same value of $G(\cdot,\cdot)$, any payoff difference between the pairs must be *exactly* offset by a log-likelihood difference in the other direction. Of course, if our goal were to maximize some concave function of $G(\cdot,\cdot)$, then mixing might give an advantage, and we discuss something similar following Proposition 8 in Section 5.

[10] Generalizing, for $(\theta'', \sigma) \in \Lambda''$, if σ has n actions in its support, then any distribution over these n actions is a best response to θ''. Let V be the $n-1$ dimensional simplex of such distributions. Note that (expected) objective payoffs and log-likelihood are linear on V. Choose a set S of $n-1$ unit basis vectors for V. Each $s \in S$ is associated with a change in objective payoffs and log-likelihood, a 2-dimensional *change* vector that we denote $c(s)$. Let $C = \{c(s) \in \mathbb{R}^2 : s \in S\}$. If there exists a strictly positive linear combination of vectors in C, then (θ'', σ) is not Pareto optimal. In the monopolist example, $n = 2$ and the (1-dimensional) probability of playing $x = 2$ is a basis associated with a strictly positive change vector. Therefore, the BNE is not Pareto optimal. If the change vector had instead different signs on objective payoffs and log-likelihood, then the BNE would have been Pareto optimal within $\{(\theta'', \hat{\sigma}) : \hat{\sigma} \in V\}$. For $n \geq 3$, things are simpler as $|S| > 1$, so it will usually be the case that $|C| > 1$. If any pair of vectors in C are linearly independent, then we can choose a strictly positive linear combination of these vectors, so (θ'', σ) is not Pareto optimal. That is, for $n \geq 3$, Pareto optimality of (θ'', σ) requires knife-edge conditions to hold.

(10) $$gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t) = \prod_{\tau=1}^{t}\left(e^{\alpha\pi(x,f(x,\omega_\tau))+(1-\alpha)\log Q_\theta(f(x,\omega_\tau)|x)}\right).$$

The evolution of generalized Bayes' rule mimics Bayes' rule: the generalized posterior weight of each model is proportional to its generalized likelihood. Given prior distribution μ_0 on Λ, the *generalized Bayesian posterior* distribution $g\mu_t$ is

(11) $$g\mu_t(A) = \frac{\int_A gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)\,d\mu_0(\theta,x)}{\int_\Lambda gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)\,d\mu_0(\theta,x)}, \qquad A \subseteq \Lambda.$$

That is, the DM obtains information on the realized payoffs and log-likelihoods that arise from playing different actions. As the sequence of states is realized over time, he uses this information to update his generalized Bayesian posterior over model-action pairs in Λ. Applying the Strong Law of Large Numbers, we are able to show that the DM gradually places more and more weight on the model-action pairs (θ,x) that optimize his preferred weighting of payoffs and likelihood as given by $G(\theta,x) = \alpha\Pi(\theta,x) + (1-\alpha)LL(\theta,x)$.

Proposition 6. *Let* $\Lambda^*_\varepsilon = \{(\theta,x) \in \Lambda : G(\theta,x) \geq G^* - \varepsilon\}$*. Assume that* $\mu_0(\Lambda^*_\varepsilon) > 0$ *for all* $\varepsilon > 0$*. Then, for all* $\varepsilon > 0$*, we have* $g\mu_t(\Lambda^*_\varepsilon) \to 1$ *p-a.s. as* $t \to \infty$*.*

Remark 1. The procedure learns over model-action pairs in Λ. If every model in Θ has a unique best response, this is identical to learning over Θ. If a model $\theta \in \Theta$ has multiple best responses $x', x'', \ldots$, then (θ,x'), $(\theta,x''), \ldots \in \Lambda$ have distinct generalized likelihoods.

Remark 2. The generalized likelihood of (θ,x) is determined by observing a sequence of realized payoffs and realized log-likelihoods. It does not matter whether such sequences arise from each member of the population following $\lambda \in \Lambda$ deterministically or randomizing over elements of Λ.

Remark 3. Our procedure uses information related to all actions $x \in X$. As such, it is suited to information-rich environments in which a diversity of behavior can be observed either directly, or indirectly as information disperses through the population. In contrast, BNE are justified by fixing a candidate x', after which the updating procedure does not use information on outcome distributions for alternative actions. Of course, choosing suitable candidates for x' is itself a distinct problem and assessing all x to find all suitable candidates requires no less information than our procedure for $\alpha = 0$. A converse issue, using too little information, arises in the case of mixed BNE. This is clear from the monopolist example of Section 4.4.2, where we saw that in mixing between two actions, in order to maintain BNE, the monopolist has to ignore evidence that one action leads to higher log-likelihood than the other.

Remark 4. If $\alpha = 1$, then from (10) we obtain

$$\log gQ_{(\theta,x)}(\omega_1,\dots,\omega_t) = \sum_{\tau=1}^{t} \pi(x, f(x,\omega_\tau)). \tag{12}$$

Observe that (12) is similar to the reinforcement algorithm of Erev and Roth (1998). Consider finite Λ and strictly positive π. Relax the assumption that we update every $gQ_{(\theta,x)}$ every period and instead begin by updating each $(\theta,x) \in \Lambda$ once, then in each subsequent period update each $(\theta,x) \in \Lambda$ with probability proportional to $\log gQ_{(\theta,x)}$. Theorem 2 of Beggs (2005) implies that this process converges, $g\mu_t(\Lambda^*_{\alpha=1}) \to 1$ p-a.s. as $t \to \infty$. Finally, note that this argument remains true even if we consider $\alpha < 1$, providing that $\alpha\pi + (1-\alpha)\log Q_\theta$ is suitably normalized to be strictly positive at all outcomes.

5.1 Evolution

If realized payoffs are understood as evolutionary fitness (assuming $\pi > 0$), we can analyze which models will have an evolutionary advantage. It is to be expected that agents who follow models that lead to high payoffs will outperform agents who follow models that lead to lower payoffs.

Consider a population of unit mass. Let the share of the population following each model at time t be given by ς_t, a probability measure on Λ. For expositional simplicity, assume ς_0 has finite support $\bar{\Lambda} \subseteq \Lambda$. For every $t \geq 1$, from period $t-1$ to t, the share of the population playing $\lambda \in \bar{\Lambda}$ changes proportionally to the mean realized payoff of agents that play λ in period $t-1$, with a normalization so that the total mass of the population remains one.

5.1.1 Independent states

If every agent has an independent realization of the state ω_t, the mean realized payoff of agents that play $\lambda = (\theta, x)$ at time t will equal the expected objective payoff from λ. Of all the models that are followed at $t = 0$, denote the set of models with the highest expected objective payoffs by

$$\bar{\Lambda}^* = \operatorname*{argmax}_{(\theta,x)\in\bar{\Lambda}} E_{Q(\cdot|x)}\, \pi(x,Y).$$

These models are favored by evolution.

Proposition 7. *If agents have independent realizations of the state, then $\varsigma_t(\bar{\Lambda}^*) \to 1$ as $t \to \infty$.*

Proposition 7 holds because, when states are independent, the models with the highest expected objective payoffs consistently lead to the highest mean realized payoffs. If $\bar{\Lambda} = \Lambda$, this can be restated in terms of the solution concepts considered earlier in the paper. Evolution leads agents to adopt the solution set $\Lambda^*_{\alpha=1}$ and, by Proposition 3, play $x^* \in \mathscr{J}$. Evolution favors justifiable Nash equilibria.

5.1.2 Shared states

Now consider the situation in which every agent faces the same realization of the state ω_t. In this case, as agents' payoffs are correlated, maximizing expected fitness for a single period is no longer the same as maximizing long run fitness. To maximize long run fitness we should instead maximize the expected growth rate (Lewontin and Cohen, 1969; Robson, 1996). Denote the set of models with the highest expected growth rate by

$$\bar{\Lambda}^{**} = \operatorname*{argmax}_{(\theta,x)\in\bar{\Lambda}} E_{Q(\cdot|x)} \log \pi(x,Y).$$

These models are favored by evolution.

Proposition 8. *If agents have the same realization of the state, then $\varsigma_t(\bar{\Lambda}^{**}) \to 1$ p-a.s. as $t \to \infty$.*

That is, in the long run the models with the highest expected growth rates dominate, even though they may occasionally suffer adversity in the form of a bad realization of ω_t. If $\bar{\Lambda} = \Lambda$, this can be restated in terms of the solution concepts considered earlier in the paper. If we replace original payoffs π by $\log \pi$, then evolution leads agents to adopt the solution set $\Lambda^*_{\alpha=1}$ of the transformed problem. Evolution thus favors justifiable Nash equilibria of the transformed problem.

There is a further subtlety that cannot be explained with reference to a simple transformation. This is that the concavity of the log function means that even higher long run growth may sometimes be achieved via mixing between models. That is, there exist situations in which mixing between models may evolutionarily outperform any single model. That this is not true for independent states but is true for shared states is immediately clear from a comparison of $\bar{\Lambda}^*$ and $\bar{\Lambda}^{**}$. An analogy of this observation can be found in the literature on the evolution of risk preferences. Specifically, the forces that encourage mixing in our setting are analogous to the forces that benefit a genotype that can generate heterogeneous risk preferences in phenotype (Heller and Nehama, 2022). Preferences are, after all, part of a model of the world.

6. Equilibrium with more than one player

Here, we extend the analysis to situations with a set of players M. Adjust the consequence function so that it depends on the profile of actions, $f : \mathbb{X}^M \times \Omega \to \Upsilon$. For player $i \in M$, let Θ^i be the parameter set, π^i the payoff function, $X^{*i}(\theta^i)$ the best responses to $\theta^i \in \Theta^i$, Λ^i the set of model-action pairs such that the action is a best response to the model. For simplicity, assume in this section that Θ^i are finite.

The true distribution over outcomes will depend on the actions taken by all of the players. We use notation that accommodates the possibility of mixing over elements of Λ^i. For any given player i, let $\lambda^i \in \Delta(\Lambda^i)$, $i \in M$ be a probability distribution over model-action pairs, and let λ be the vector of λ^i, $i \in M$. Denote the true conditional distribution faced by player i, keeping λ^j, $j \neq i$ fixed, by $Q^i_\lambda(\cdot|x)$. Given this true distribution, objective payoffs Π^i_λ, log-likelihoods LL^i_λ, and

$$G^i_\lambda = \alpha^i \Pi^i_\lambda + (1-\alpha^i) LL^i_\lambda$$

are defined as before. Define the solution set

$$\mathbf{\Lambda}^* = \left\{ \lambda : \operatorname{supp} \lambda^i \subseteq \Lambda^{*i}_\lambda \right\}.$$

Solutions in $\mathbf{\Lambda}^*$ involve each player i choosing model-action pairs to maximize their goal function, keeping fixed the distribution over the actions of the other players. Pure solutions involve each player choosing a single model-action pair in Λ^i. It is possible that a pure solution may not exist, but mixing guarantees existence.[11]

Proposition 9. $\mathbf{\Lambda}^* \neq \varnothing$

One interpretation of this approach is that it reduces the problem to a game with player set M, action sets Λ^i and payoff functions $G^i_\lambda(\cdot)$. Each pure (mixed) solution corresponds to a pure (mixed) Nash equilibrium of the reduced game together with supporting models. The supporting intuition is that, under an appropriate learning procedure, the role of model misspecification is to reduce the choice of model-action pairs available to a decision maker. The decision maker then assesses the success he obtains from playing a model-action pair according to (i) the payoff from playing the action, and (ii) how well the model fits observed outcomes.

An important special case is when player i only cares about payoffs, $\alpha^i = 1$. Fixing λ_j, $j \neq i$, Proposition 3 tells us that actions played by i with positive probability under $\mathbf{\Lambda}^*$ will maximize $\mathscr{J}^i_\lambda$. If this is true for all players, then behavior under $\mathbf{\Lambda}^*$ corresponds to the Nash equilibria of the game in which players are restricted to justifiable actions and payoffs are given by Π^i_λ.

Regarding learnability and evolutionary selection, analogies of the results in Section 5 can be constructed for a multi-player setting. Given the clear similarities

[11] The type of mixing used here is consistent with the "mass action" interpretation of mixing from John Nash's PhD thesis (Nash, 1950a). Under this interpretation, a mixture between (θ^{i1}, x^1) and (θ^{i2}, x^2) would indicate that player i is drawn from some population and that such a draw renders some chance of player i being of type $i1$, for whom x^1 is a best response, and some chance of player i being of type $i2$, for whom x^2 is a best response. Note that this is not the same as a mixing player holding beliefs that are a convex combination of $Q^i_{\theta^{i1}}$ and $Q^i_{\theta^{i2}}$, in which case it is possible that neither x^1 nor x^2 is a best response.

between standard Nash equilibrium and equilibrium concepts in misspecified settings, the reader will not be surprised that insights gained from studying models of learning (Newton, 2018), evolutionary stability of NE (see, e.g. Dekel et al., 2007; Heifetz et al., 2007; Ok and Vega-Redondo, 2001) and the effects of assortativity in matching (Alger and Weibull, 2013; Bergstrom, 1995; Eshel and Cavalli-Sforza, 1982; Newton, 2017) also apply to misspecified settings.

7. Conclusion

We conclude by anticipating some criticisms. First, our suggested approach describes a DM who is individually rational to the extent that his choice variables, θ and x, are jointly determined with regard to the maximization of a goal function. Rationality of an individual is usually contrasted with collective rationality of a group of individuals. However, it is also possible to contrast the rationality of an individual with the rationality of multiple selves within that individual. The internal consistency approach uses such multiple selves, treating the DM as a pair of agent-modules, then demanding rational choice from both the agent-module that chooses actions and the agent-module that chooses models. From an evolutionary perspective, if the unit of replication is the DM himself, we might expect more holistic choice procedures to be favored over such a disaggregated approach.

Second, maximizing $G(\theta,x)$ requires learning about objective payoffs and log-likelihoods for model-action pairs. In contrast, the consistency approach requires learning about log-likelihoods given a strategy and, if best responses change as a result, hoping that the process converges. If we ignore the issue of convergence, it seems that the information requirements of the consistency approach are lower as long as $\alpha < 1$. However, again considering the problem from an evolutionary perspective, it is clear why DMs that gain from finding solutions in Λ^* would want to learn, by observing society and history, information related to the components of $G(\theta,x)$. It also seems reasonable for the ability to garner high payoffs and make accurate predictions to be favored by natural selection. In contrast, if we consider BNE Λ' and mixed BNE Λ'', it is less clear why evolution would favour a DM that learns only from his own choices and ignores information about payoffs and log-likelihoods that could be gathered from observing society. Is consistency really such a compelling goal? If it is true that "the human mind is programmed for survival, not for truth" (Gray, 2018), then why would consistency favour survival?

We do not pretend to have definitive answers to all of the questions above, but hope to have convinced the reader that the issues we have considered here are worthy of further consideration.

A. Appendix: Omitted proofs

Proof of Proposition 1.
Immediate from definitions of Λ^*_α and Λ'. □

Proof of Proposition 2.
Under (I1) and (I2), $LL(\theta,x)$ is independent of x. Noting that every $\theta \in \Theta$ appears in at least one element of Λ, this implies that

$$(13) \qquad (\theta^*,x^*) \in \arg\max_{(\theta,x)\in\Lambda} LL(\theta,x) \quad \Rightarrow \quad \theta^* \in \arg\max_{\theta\in\Theta} LL(\theta,x^*).$$

That is, $\Lambda^*_{\alpha=0} \subseteq \Lambda'$.

Conversely, let $(\theta',x') \in \Lambda'$. By definition of Λ',

$$(14) \qquad \theta' \in \arg\max_{\theta\in\Theta} LL(\theta,x'),$$

which as $LL(\theta,x)$ is independent of x implies that $LL(\theta',x') \geq LL(\theta,x)$ for all $(\theta,x) \in \Lambda$. Therefore, $(\theta',x') \in \Lambda^*$. That is, $\Lambda' \subseteq \Lambda^*_{\alpha=0}$. □

Proof of Proposition 3.
By definition, x' is justifiable if and only if there exists nonempty $\bar{\Theta}(x') \subseteq \Theta$ such that, for all $\theta' \in \bar{\Theta}(x')$, $(\theta',x') \in \Lambda$.

For $\alpha = 1$,

$$(15) \qquad G(\theta,x) = \Pi(\theta,x) = E_{Q(\cdot|x)}\pi(x,Y) \quad \text{for all} \quad (\theta,x) \in \Lambda.$$

As $\Pi(\theta,x)$ does not directly depend on θ, it follows that x^* maximizes $E_{Q(\cdot|x)}\pi(x,Y)$ over justifiable x if and only if, for all $\theta^* \in \bar{\Theta}(x^*)$, (θ^*,x^*) maximizes $G(\theta,x)$. □

Proof of Proposition 4.
By definition of Λ^*, $(\theta,x) \in \Lambda^*$ solves $\max_{(\theta,x)\in\Lambda} G(\theta,x)$, with

$$G(\theta,x) = \alpha\Pi(\theta,x) + (1-\alpha)LL(\theta,x). \tag{16}$$

Note that $\Pi(\theta,x)$ does not directly depend on θ. Furthermore, if (I1) and (I2) hold, then $LL(\theta,x)$ is independent of x. Therefore, we choose x to maximize $\Pi(\theta,x)$ and θ to maximize $LL(\theta,x)$, before verifying that the resulting model-action pair is in Λ and therefore in Λ^*.

First, choose θ to maximize $LL(\theta,x)$. By definition of $\theta^\dagger$, for any given $x \in \mathbb{X}$,

$$\theta^\dagger \in \arg\max_{\theta\in\Theta} LL(\theta,x). \tag{17}$$

Next, choose x to maximize $\Pi(\theta,x)$,

$$x^\dagger \in \arg\max_{x\in\mathbb{X}} E_{Q(\cdot|x)}\pi(x,Y). \tag{18}$$

As $Q(\cdot|x) = Q_{\theta^\dagger}(\cdot|x)$, (18) implies

$$x^\dagger \in \arg\max_{x\in\mathbb{X}} E_{Q_{\theta^\dagger}(\cdot|x)}\pi(x,Y). \tag{19}$$

That is, $x^\dagger \in X^*(\theta^\dagger)$, and consequently $(\theta^\dagger,x^\dagger) \in \Lambda$.

For the converse part of the proposition, assume $(\theta^*,x^*) \in \Lambda^*$. If $Q_{\theta^*}(\cdot|x^*) \neq Q_{\theta^\dagger}(\cdot|x^*)$, then the DM is not maximizing $LL(\theta,x)$. Furthermore, by (I2), if $Q_{\theta^*}(\cdot|x^*) = Q_{\theta^\dagger}(\cdot|x^*)$, then $Q_{\theta^*}(\cdot|x) = Q_{\theta^\dagger}(\cdot|x)$ for all $x \in \mathbb{X}$. If $x^* \notin X^*(\theta^\dagger)$, then the DM is not maximizing $\Pi(\theta,x)$.

For $0 < \alpha < 1$, Λ^* maximizes a convex combination of $LL(\theta,x)$ and $\Pi(\theta,x)$. As we have seen, under the conditions of the proposition, this can be done independently for θ and x. Therefore, maximizing $G(\theta,x)$ requires that $Q_{\theta^*}(\cdot|x^*) = Q_{\theta^\dagger}(\cdot|x^*)$ and $x^* \in X^*(\theta^\dagger)$.

For $\alpha = 0$, Λ^* maximizes $LL(\theta, x)$, which requires $Q_{\theta^*}(\cdot|x^*) = Q_{\theta^\dagger}(\cdot|x^*)$. This implies that $x^* \in X^*(\theta^\dagger)$.

For $\alpha = 1$, Λ^* maximizes $\Pi(\theta, x)$ which requires $x^* \in X^*(\theta^\dagger)$. $\square$

Proof of Proposition 6.

The result follows from the strong law of large numbers (SLLN):

$$\begin{aligned}
g\mu_t(\Lambda^*_\varepsilon) = 1 - g\mu_t(\Lambda \setminus \Lambda^*_\varepsilon) \underbrace{=}_{\text{by}(11)} 1 - \frac{\int_{\Lambda\setminus\Lambda^*_\varepsilon} gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)\, d\mu_0(\theta,x)}{\int_\Lambda gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)\, d\mu_0(\theta,x)}\\
\geq 1 - \frac{\int_{\Lambda\setminus\Lambda^*_\varepsilon} gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)\, d\mu_0(\theta,x)}{\int_{\Lambda^*_{\varepsilon/2}} gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)\, d\mu_0(\theta,x)}\\
= 1 - \frac{\int_{\Lambda\setminus\Lambda^*_\varepsilon} e^{\ln gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)}\, d\mu_0(\theta,x)}{\int_{\Lambda^*_{\varepsilon/2}} e^{\ln gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t)}\, d\mu_0(\theta,x)}\\
\underbrace{=}_{\text{by}(10)} 1 - \frac{\int_{\Lambda\setminus\Lambda^*_\varepsilon} e^{t\sum_{\tau=1}^t \frac{1}{t}(\alpha\pi(x,f(x,\omega_\tau))+(1-\alpha)\log Q_\theta(f(x,\omega_\tau)|x))\, d\mu_0(\theta,x)}}{\int_{\Lambda^*_{\varepsilon/2}} e^{t\sum_{\tau=1}^t \frac{1}{t}(\alpha\pi(x,f(x,\omega_\tau))+(1-\alpha)\log Q_\theta(f(x,\omega_\tau)|x))\, d\mu_0(\theta,x)}}\\
\underbrace{\approx}_{\substack{p\text{-a.s. for } t \text{ large}\\ \text{by SLLN}}} 1 - \frac{\int_{\Lambda\setminus\Lambda^*_\varepsilon} e^{t\,G(\theta,x)}\, d\mu_0(\theta,x)}{\int_{\Lambda^*_{\varepsilon/2}} e^{t\,G(\theta,x)}\, d\mu_0(\theta,x)}\\
\geq 1 - \frac{e^{t\,(G^*-\varepsilon)}\,\mu_0(\Lambda\setminus\Lambda^*_\varepsilon)}{e^{t\,(G^*-\varepsilon/2)}\,\mu_0(\Lambda^*_{\varepsilon/2})} \xrightarrow{t\to\infty} 1.
\end{aligned}$$

$\square$

Proof of Proposition 7.

Consider λ, λ^* such that $\varsigma_0(\lambda), \varsigma_0(\lambda^*) > 0$. Let $\lambda \notin \bar{\Lambda}^*$, $\lambda^* \in \bar{\Lambda}^*$.

At time t every agent following λ has an independent state ω_t. Therefore, the mean realized payoff for agents following λ equals $E_{Q(\cdot|x)}\pi(x,Y)$. Similarly, the mean realized payoff for agents following λ^* equals $E_{Q(\cdot|x^*)}\pi(x,Y)$.

As $\lambda \notin \bar{\Lambda}^*$, $\lambda^* \in \bar{\Lambda}^*$, we have

(20) $$E_{Q(\cdot|x)}\pi(x,Y) < E_{Q(\cdot|x^*)}\pi(x^*,Y).$$

As $\varsigma_t(\cdot)$ grows proportionally to realized payoffs (subject to normalization),

(21) $$\frac{\varsigma_t(\lambda)}{\varsigma_t(\lambda^*)} = \frac{\varsigma_{t-1}(\lambda)}{\varsigma_{t-1}(\lambda^*)} \underbrace{\frac{E_{Q(\cdot|x)}\pi(x,Y)}{E_{Q(\cdot|x^*)}\pi(x^*,Y)}}_{<1 \text{ by } (20)} \quad \text{for} \quad t \geq 1.$$

Iterating (21), we obtain

(22) $$\frac{\varsigma_t(\lambda)}{\varsigma_t(\lambda^*)} = \frac{\varsigma_0(\lambda)}{\varsigma_0(\lambda^*)} \left(\frac{E_{Q(\cdot|x)}\pi(x,Y)}{E_{Q(\cdot|x^*)}\pi(x^*,Y)}\right)^t \xrightarrow{t\to\infty} 0.$$

As (22) holds for all $\lambda \notin \bar{\Lambda}^*$, we have $\varsigma_t(\bar{\Lambda}^*) \to 1$ as $t \to \infty$. □

Proof of Proposition 8.
Consider λ, λ^{**} such that $\varsigma_0(\lambda), \varsigma_0(\lambda^{**}) > 0$. Let $\lambda \notin \bar{\Lambda}^{**}$, $\lambda^{**} \in \bar{\Lambda}^{**}$, so that

(23) $$E_{Q(\cdot|x)} \log \pi(x,Y) < E_{Q(\cdot|x^{**})} \log \pi(x^{**},Y).$$

As $\varsigma_t(\cdot)$ grows proportionally to realized payoffs (subject to normalization),

(24) $$\frac{\varsigma_t(\lambda)}{\varsigma_t(\lambda^{**})} = \frac{\varsigma_{t-1}(\lambda)}{\varsigma_{t-1}(\lambda^{**})} \frac{\pi(x, f(x,\omega_{t-1}))}{\pi(x^{**}, f(x^{**},\omega_{t-1}))} \quad \text{for} \quad t \geq 1.$$

Iterating (24), we obtain

(25) $$\frac{\varsigma_t(\lambda)}{\varsigma_t(\lambda^{**})} = \frac{\varsigma_0(\lambda)}{\varsigma_0(\lambda^{**})} \prod_{\tau=0}^{t-1} \frac{\pi(x, f(x,\omega_\tau))}{\pi(x^{**}, f(x^{**},\omega_\tau))}.$$

Taking logs,

(26) $$\begin{aligned} &\log \varsigma_t(\lambda) - \log \varsigma_t(\lambda^{**}) - \log \varsigma_0(\lambda) + \log \varsigma_0(\lambda^{**}) \\ &= t \left(\frac{1}{t}\sum_{\tau=0}^{t-1} \log \pi(x, f(x,\omega_\tau)) - \frac{1}{t}\sum_{\tau=0}^{t-1} \log \pi(x^{**}, f(x^{**},\omega_\tau))\right) \\ &\xrightarrow[\text{by SLLN}]{t\to\infty} t \left(\underbrace{E_{Q(\cdot|x)} \log \pi(x,Y) - E_{Q(\cdot|x^{**})} \log \pi(x^{**},Y)}_{<0 \text{ by } (23)}\right) \end{aligned}$$

$$\xrightarrow{t\to\infty} -\infty.$$

Consider the left hand side of (26). As the second term is positive and the final two terms do not depend on t, it must be that the first term $\log \varsigma_t(\lambda) \to -\infty$ as $t \to \infty$. This implies that $\varsigma_t(\lambda) \to 0$ as $t \to \infty$. As this holds for all $\lambda \notin \bar{\Lambda}^{**}$, we have $\varsigma_t(\bar{\Lambda}^{**}) \to 1$ as $t \to \infty$. □

Proof of Proposition 9.
The game Γ with player set I, pure strategies $(\Lambda^i)_{i\in I}$ and payoffs given by G^i_λ is finite and thus has at least one, possibly mixed, Nash equilibrium by Nash's existence theorem (Nash, 1950b). Choose one such equilibrium and denote it by $\hat{\lambda} = (\hat{\lambda}^i)_i$.

We claim that $\hat{\lambda} \in \mathbf{\Lambda}^*$. If this is not the case, then by the definition of $\mathbf{\Lambda}^*$, for some $i \in I$, we have

$$\text{supp}\,\hat{\lambda}^i \not\subseteq \Lambda^{*i}_{\hat{\lambda}}. \tag{27}$$

That is, there exists $(\theta^i, x^i) \in \Lambda^i \setminus \Lambda^{*i}_{\hat{\lambda}}$ such that $\hat{\lambda}^i((\theta^i, x^i)) > 0$.

As $(\theta^i, x^i) \notin \Lambda^{*i}_{\hat{\lambda}}$, there exists $(\tilde{\theta}^i, \tilde{x}^i) \in \Lambda^i$ such that

$$G^i_{\hat{\lambda}}(\tilde{\theta}^i, \tilde{x}^i) > G^i_{\hat{\lambda}}(\theta^i, x^i). \tag{28}$$

Therefore, $\hat{\lambda}^i$ with $\hat{\lambda}^i((\theta^i, x^i)) > 0$ is not a best response to $\hat{\lambda}$, contradicting $\hat{\lambda}$ being a Nash equilibrium. Therefore, $\hat{\lambda} \in \mathbf{\Lambda}^*$, proving the proposition. □

References

Alger, I. and Weibull, J. W. (2013). Homo moralis–preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6):2269–2302.

Beggs, A. W. (2005). On the convergence of reinforcement learning. *Journal of economic theory*, 122(1):1–36.

Bergstrom, T. C. (1995). On the evolution of altruistic ethical rules for siblings. *American Economic Review*, pages 58–81.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58.

Caplin, A. and Leahy, J. V. (2019). Wishful thinking. Working Paper 25707, National Bureau of Economic Research.

Dekel, E., Ely, J. C., and Yilankaya, O. (2007). Evolution of preferences. *The Review of Economic Studies*, 74(3):685–704.

Edhan, O., Hellman, Z., and Sherill-Rofe, D. (2017). Sex with no regrets: How sexual reproduction uses a no regret learning algorithm for evolutionary advantage. *Journal of theoretical biology*, 426:67–81.

Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, 88(4):848–881.

Eshel, I. and Cavalli-Sforza, L. L. (1982). Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences*, 79(4):1331–1335.

Esponda, I. and Pouzo, D. (2016). Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130.

Esponda, I., Pouzo, D., and Yamamoto, Y. (2021). Asymptotic behavior of bayesian learners with misspecified models. *Journal of Economic Theory*, 195:105260.

Frenkel, S., Heller, Y., and Teper, R. (2018). The endowment effect as blessing. *International Economic Review*. (online first).

Frick, M., Iijima, R., and Ishii, Y. (2023). Belief convergence under misspecified learning: A martingale approach. *The Review of Economic Studies*, 90(2):781–814.

Fudenberg, D. and Lanzani, G. (2022). Which misperceptions persist? *Theoretical Economics*.

Fudenberg, D., Lanzani, G., and Strack, P. (2021). Limit points of endogenous misspecified learning. *Econometrica*, 89(3):1065–1098.

Fudenberg, D. and Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica*, pages 523–545.

Gilboa, I. (2009). *Theory of decision under uncertainty*, volume 1. Cambridge university press.

Gray, J. (2018). *Seven types of atheism*. Penguin UK.

Greenberg, J., Gupta, S., and Luo, X. (2009). Mutually acceptable courses of action. *Economic Theory*, 40(1):91–112.

Grünwald, P., Van Ommen, T., et al. (2017). Inconsistency of bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Analysis*, 12(4):1069–1103.

He, K. and Libgober, J. (2021). Evolutionarily stable (mis) specifications: Theory and applications. *PIER Working Paper*.

Heifetz, A., Shannon, C., and Spiegel, Y. (2007). What to maximize if you must. *Journal of Economic Theory*, 133(1):31–57.

Heller, Y. (2014). Overconfidence and diversification. *American Economic Journal: Microeconomics*, 6(1):134–153.

Heller, Y. and Nehama, I. (2022). Evolutionary foundation for heterogeneity in risk aversion. *Available at SSRN 3942389*.

Johnson, D. D. and Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, 477(7364):317–320.

Jouini, E., Napp, C., and Viossat, Y. (2013). Evolutionary beliefs and financial markets. *Review of Finance*, 17(2):727–766.

Lewontin, R. C. and Cohen, D. (1969). On population growth in a randomly varying environment. *Proceedings of the National Academy of Sciences*, 62(4):1056–1060.

Murooka, T. and Yamamoto, Y. (2023). Higher-order misspecification and equilibrium stability. *mimeo*.

Nash, J. (1950a). *Non-cooperative games*. PhD thesis, Princeton University, USA.

Nash, J. F. (1950b). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49.

Newton, J. (2017). The preferences of homo moralis are unstable under evolving assortativity. *International Journal of Game Theory*, 46(2):583–589.

Newton, J. (2018). Evolutionary game theory: A renaissance. *Games*, 9(2):31.

Nyarko, Y. (1991). Learning in mis-specified models and the possibility of cycles. *Journal of Economic Theory*, 55(2):416–427.

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

Page, L. (2021). *Optimally irrational*. Academic Press.

Robson, A. J. (1996). A biological basis for expected and non-expected utility. *Journal of economic theory*, 68(2):397–424.

Rubinstein, A. and Wolinsky, A. (1994). Rationalizable conjectural equilibrium: Between nash and rationalizability. *Games and Economic Behavior*, 6(2):299–311.