Learning and equilibrium in misspecified models*

Filippo Massari† and Jonathan Newton‡

September 24, 2020

Abstract

We consider learning in games that are misspecified in that players are unable to learn the true probability distribution over outcomes. Under misspecification, Bayes' rule might not converge to the model that leads to actions with the highest objective payoff among the models subjectively admitted by the player. From an evolutionary perspective, this renders a population of Bayesians vulnerable to invasion. Drawing on the machine learning literature, we show that learning rules that outperform Bayes' rule suggest a new solution concept for misspecified games: misspecified Nash equilibrium.

Keywords: misspecified learning, evolutionary models.

JEL Codes: C7, D8.

"No statistical model is "true" or "false", "right" or "wrong"; the models just have varying performance, which can be assessed."

– Rissanen (2007).

1. Introduction

A desirable property of an equilibrium concept is that there be no profitable deviation (NPD). In the space of learning rules (or, a fortiori, beliefs), this means that an adopted learning rule should ensure that the learning outcome is not a model that, if believed true, leads a player to choose actions that have lower expected payoff (according to the true distribution) than the actions she would choose if she had learned another model.

*6 minute video summary: https://youtu.be/kilz3oFSYp4. The authors thank Larry Samuelson for his reading and guidance.

†School of Economics, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom. e-mail: f.massari@uea.ac.uk; 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.

		Bob		
		b_1	b_2	b_3
Alice	a_1	2,2	1,0	4,1
	a_2	1,4	1,5	3,3
	a_3	0,1	0,0	5,5

Figure 1: **A game.** For each combination of strategies, entries give payoffs for Alice and Bob respectively.

When a learning problem is correctly specified, a Bayesian learner will learn the true model. In such an environment, no alternative learning rule strictly outperforms Bayesian learning. However, this is not the case when the learning problem is misspecified. To see this, consider a coin which can land either heads or tails and has a true probability of 0.7 of landing heads. Consider Alice, a Bayesian learner whose prior has full support over two models of the coin, one in which the probability of heads is 0.45 and one in which the probability of heads is 0.9. Every period, she earns a dollar if she correctly guesses the outcome of the coin toss. Bayesian updating will lead her, in the limit, to place probability one on the first model. Hence, she will predict tails and earn an average per period payoff of 0.3. However, she would achieve a higher payoff if she placed probability one on the second model, predicted heads and earned an average per period payoff of 0.7. NPD is not satisfied.

From an evolutionary perspective (see, e.g. Weibull, 1995; Sandholm, 2010), a population of Alice-like players who learn using Bayes' rule would thus be vulnerable to invasion by a mutant, say Bob, who follows a learning rule that eventually places probability one on the second model, leading him to predict heads. If payoff is positively related to replication, then over time the share of Bobs in the population will increase as they outperform the Alices in terms of realized payoff. Note that Bob in fact performs exactly as well as another player type, say Colm, who learns the correct belief that the probability of heads is 0.7. In pragmatic terms, Bob and Colm learn perfectly. Alice, in contrast, learns the model in her support that maximizes log-likelihood.

Here, we propose an equilibrium concept that satisfies NPD even if the learning environment is not well specified. The equilibrium we propose, *misspecified Nash equilibrium* (mNE), requires that each player's beliefs attach probability one to the set of subjective distributions over consequences that lead to his taking actions that lead to the highest realized payoff, keeping fixed the strategies of the other players. In the one player coin toss game described above, Bob learns to play an mNE, but Alice does not. In well-specified learning settings, every NE is a mNE. The reverse inclusion does not hold because mNE does not uniquely pin down beliefs: the same actions can be, and typically are, a best response to more than one belief. In the coin toss example above, Bob learns an mNE but not an NE.

One way to understand mNE is as a restricted Nash equilibrium in which the actions available to each player are restricted by the incomplete set of beliefs available to them. Consider the game in Figure 1. If Alice's set of conceivable models does not include any model in which Bob plays b_3 with enough probability, then Alice will never play a_3 as a best response. Similarly, if Bob's set of conceivable models does not include any model in which Alice plays $a_2$with enough probability, then Bob will never play b_2 as a best response. We are left with a restricted game in which the action sets are $\{a_1, a_2\}$ and $\{b_1, b_3\}$. Given that Bob plays some combination of b_1 and b_3, Alice is better off when she learns a set of beliefs that leads her to play a_1 rather than a set of beliefs that leads her to play a_2. Similarly, Bob should learn a set of beliefs that leads him to play b_1. Hence $\{a_1, b_1\}$, which constitutes a Nash equilibrium of the restricted game, is the action profile that will be played at any misspecified Nash equilibrium. Players' beliefs at mNE can be any subjectively possible beliefs that lead to these actions. Beliefs at mNE can thus be considered constrained optimal, where the constraints arise from the restriction to the sets of beliefs that players consider possible.

Optimality itself can be understood as something that is strategically pursued (Brunnermeier and Parker, 2005; Brunnermeier et al., 2007) or understood evolutionarily in the sense of fitness maximization (Johnson and Fowler, 2011; Jouini et al., 2013; Frenkel et al., 2018; Heller, 2014). This relationship between mNE and standard NE suggests that procedures for learning NE such as regret testing (Foster and Young, 2006; Germano and Lugosi, 2007) or interactive trial and error learning (Young, 2009; Pradelski and Young, 2012) can be adapted to this constrained optimization problem to give learning foundations for mNE. Conversely, learning procedures that separate beliefs from payoffs, such as Bayes' rule, can be sub-optimal in such a setting because they ignore the payoff consequences of the learning outcome.

The main limitation of Bayes' rule in misspecified learning problems is that, even when convergence of the prior occurs, Bayes' rule converges to the maximum likelihood model (Berk, 1966; White, 1982), but there is no guarantee that the maximum likelihood model is also the model that maximizes payoffs (Grünwald et al., 2017; Csaba and Szoke, 2018; Massari, 2019). This lack of robustness to model misspecification is the reason why the use of Bayes' rule in misspecified problems is controversial in the (more pragmatic) statistical learning and computer science literatures which are mainly concerned with empirical validation rather than consistency with a set of axioms.[1] Pragmatically, for a decision maker that wishes to maximize payoff, the adoption of Bayes' rule in (possibly) misspecified learning

[1]See Timmermann (2006); Grünwald (2007); Grünwald and Langford (2007). Note that in this literature, maximizing expected payoff is usually described as minimizing an expected loss.

problems is 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).

There are many candidate solutions to "robustify" a learning problem. Most of them are obtained by incorporating the objective of the decision maker into his learning rule to give more weight to models that induce actions that lead to high expected payoffs (according to the objective distribution), rather than to the models with the highest likelihood. Typically, these procedures are model free. However, if a player subjectively believes that only certain models are possible, then a descriptive theory of learning should account for this.[2]

Here we propose an approach which, to the best of the authors' knowledge, has not been explored in the economics literature. We modify the *entropification* procedure of Grünwald (1998) to create a learning procedure for our model that fits the generalized Bayesian learning framework (a.k.a. aggregation algorithm, Vovk, 1990; Rissanen, 1989). We show that the beliefs that a player learns under this procedure will be optimal within the class of models that he considers possible. That is, the beliefs learned are those that support mNE.

After carrying out entropification, our learning procedure closely parallels Bayesian learning. Following this line of reasoning, we can then compare mNE with the corresponding Bayesian equilibrium concept, Berk-Nash equilibrium proposed by Esponda and Pouzo (2016). We discuss how mNE is the appropriate concept when players face an external criterion of success. One such situation is when payoffs correspond to fitness. We show that if a population follows a learning rule that leads to equilibrium behavior that is not a mNE, then the population is vulnerable to invasion by players who follow a different rule. Conversely, if the learning rule leads to a strict mNE, then the population is robust to invasion by other learning rules. The stability concept we use adapts the idea of an evolutionarily stable state (Taylor and Jonker, 1978) for a situation in which the aspect of the environment under evolutionary pressure is neither strategies (Weibull, 1995), nor preferences (Samuelson, 2001), nor agency (Newton, 2017b), but rather the learning rule that players follow.

The paper is organized as follows. Section 2 gives the model. Section 3 defines and discusses mNE. Section 4 describes the learning foundation of mNE. Section 5 compares mNE and Berk-Nash equilibrium. Section 6 provides examples. Section 7 discusses the evolutionary stability of mNE.

[2]Furthermore, although model free algorithms optimize average payoffs, they necessarily have slower learning rates than algorithms that only search among a subset of probabilistic models. This trade-off is the reason that information is valuable in economics.

2. Model

A **game** $\mathscr{G} = \langle \mathscr{O}, \mathscr{Q} \rangle$ is composed of a (simultaneous-move) objective game $\mathscr{O}$ and a subjective model $\mathscr{Q}$. The objective game represents the players' true environment. The subjective model represents the players' perception of their environment.

OBJECTIVE GAME. A (simultaneous-move) **objective** game is a tuple

$$\mathscr{O} = \langle I, \Omega, \mathbb{S}, p, \mathbb{X}, \mathbb{Y}, f, \pi \rangle .$$

I is the set of players. Ω is the set of payoff-relevant states. $\mathbb{S} = \times_{i \in I} \mathbb{S}^i$ is the set of profiles of signals, where $\mathbb{S}^i$ is the set of signals for player i. p is a probability distribution over $\Omega \times \mathbb{S}$ and is assumed to have marginals with full support. Standard notation is used to denote marginal and conditional distributions, for example $p_{\Omega|S^i}(\cdot|s^i)$ denotes the conditional distribution over Ω given $S^i = s^i$. $\mathbb{X} = \times_{i \in I} \mathbb{X}^i$ is a set of profiles of actions, where $\mathbb{X}^i$ is the set of actions of player i. $\mathbb{Y} = \times_{i \in I} \mathbb{Y}^i$ is a set of profiles of (observable) consequences, where $\mathbb{Y}^i$ is the set of consequences for player i. $f = (f^i)_{i \in I}$ is a profile of feedback or consequence functions, where $f^i : \mathbb{X} \times \Omega \to \mathbb{Y}^i$ maps outcomes in $\Omega \times \mathbb{X}$ into consequences for player i. $\pi = (\pi^i)_{i \in I}$, where $\pi^i : \mathbb{X}^i \times \mathbb{Y}^i \to \mathbb{R}$ is the payoff function of player i. All of the above sets are finite.

A strategy for player i is a mapping $\sigma^i : \mathbb{S}^i \to \Delta(\mathbb{X}^i)$. The probability that player i chooses action x^i after observing signal s^i is denoted by $\sigma^i(x^i|s^i)$. A strategy profile is a vector of strategies $\sigma = (\sigma^i)_{i \in I}$. Let Σ denote the space of all strategy profiles.

Fix an objective game. For each strategy profile σ, there is an **objective distribution** over player i's consequences, $Q^i_\sigma : \mathbb{S}^i \times \mathbb{X}^i \to \Delta(\mathbb{Y}^i)$, where

$$Q^i_\sigma(y^i|s^i,x^i) = \sum_{\{(\omega,x^{-i}) : f^i(x^i,x^{-i},\omega) = y^i\}} \sum_{s^{-i}} \prod_{j \neq i} \sigma^j(x^j|s^j) p_{\Omega \times S^{-i}|S^i}(\omega, s^{-i}|s^i). \tag{1}$$

That is, when the strategy profile is σ, player i observes signal s^i and takes action x^i, then the distribution over consequences for player i is given by $Q^i_\sigma(\cdot|s^i,x^i)$.

SUBJECTIVE MODEL. The subjective model is the set of distributions over consequences that players consider possible a priori. For a fixed objective game, a **subjective model** is a tuple

$$\mathscr{Q} = \langle \Theta, (Q_\theta)_{\theta \in \Theta} \rangle ,$$

$\Theta = \times_{i \in I} \Theta^i$ and Θ^i is player i's parameter set. $Q_\theta = (Q^i_{\theta^i})_{i \in I}$, where $Q^i_{\theta^i} : \mathbb{S}^i \times \mathbb{X}^i \to \Delta(\mathbb{Y}^i)$ is the conditional distribution over player i's consequences parameterized by $\theta^i \in \Theta^i$. Denote the conditional distribution by $Q^i_{\theta^i}(\cdot|s^i,x^i)$.

3. Misspecified Nash Equilibrium

Misspecified Nash equilibrium requires that each player's beliefs attach probability one to the set of subjective distributions over consequences that lead to his taking actions that lead to the highest realized payoff, keeping fixed the strategies of the other players.

First, we define the set of *best responses* induced by every subjective belief and the set of *subjectively non-dominated* model-response pairs.

Definition 1. The set of **best responses** of player i to $Q^i_{\theta^i}$ is

$$X^*(Q^i_{\theta^i}) = \times_{s^i \in \mathbb{S}^i} X^*(Q^i_{\theta^i}, s^i),$$

where

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

So $(\bar{x}^i_{s^i})_{s^i \in \mathbb{S}^i} \in X^*(Q^i_{\theta^i})$ is a vector, each element of which comprises a best response for some signal. For some $Q^i_{\theta^i}$, it may be that $X^*(Q^i_{\theta^i})$ has multiple elements. When this is the case, it suits to consider each pair $(Q^i_{\theta^i}, \bar{x}^i)$ as a distinct object that can be learned. We define the set of all such model-response pairs.

Definition 2. The set of **subjectively non-dominated** model-response pairs of player i is

$$\Lambda^i = \left\{ (\theta^i, \bar{x}^i) : \theta^i \in \Theta^i, \bar{x}^i \in X^*(Q^i_{\theta^i}) \right\}.$$

It must be that every $\theta^i \in \Theta^i$ appears in at least one element of Λ^i, but the same is not true for $\bar{x}^i \in (\mathbb{X}^i)^{\mathbb{S}^i}$. If $\bar{x}^i$ is not a best response for any subjective model considered by player i, then it will not be part of any element of Λ^i. Conversely, the same actions can occur in multiple elements of Λ^i. For example, considering the coin toss example from our introduction, if there are multiple models that give a probability of heads of at least half, then each of these models paired with the action "predict heads" will be an element of Λ^i.

There is more than one way to consider such pairs $(\theta^i, \bar{x}^i)$ bonded by a best response correspondence. Our preferred interpretation is that subjective beliefs are ancillary to actions in the sense that it is possible to omit beliefs from the decision model and still have a model, but the model without actions would be nonsensical. What the beliefs do is to restrict the set of possible actions to those that are justifiable by some model in the prior. Actions that are unjustifiable are never taken.

Second, given any belief and an associated subjective best response, we calculate the objective (expected) payoff for player i against a given strategy profile σ.

Definition 3. The **objective payoff** of player i from $(\theta^i, \vec{x}^i) \in \Lambda^i$ is

$$\Pi^i_\sigma(Q^i_{\theta^i}, \vec{x}^i) = \sum_{s^i \in \mathbb{S}^i} E_{Q^i_\sigma(\cdot|s^i, \vec{x}^i_{s^i})} \pi^i(\vec{x}^i_{s^i}, Y^i)\, p_{S^i}(s^i).$$

If θ^i is the true model under strategy profile σ, then any choice of $\vec{x}^i$ gives the same objective payoff, so we omit the second argument and write $\Pi^i_\sigma(Q^i_\sigma)$.

Finally, we have that

Definition 4. A profile $(\theta^{i*}, \vec{x}^{i*})_{i \in I}$ is a **misspecified Nash equilibrium** (mNE) if, for all $i \in I$,

$$(\theta^{i*}, \vec{x}^{i*}) \in \operatorname*{argmax}_{(\theta^i, \vec{x}^i) \in \Lambda^i} \Pi^i_\sigma(Q^i_{\theta^i}, \vec{x}^i). \tag{2}$$

mNE is a solution concept for players who (i) care about obtaining as high a payoff as possible for themselves, similarly to all Nash-style concepts, and (ii) learn about what they care about. In equilibrium, there are no (subjective) beliefs that player i could learn that could lead him to act in a way that would increase his (objective) expected payoff. In other words, there does not exist an (objectively) profitable deviation to a different set of beliefs together with (subjectively) optimal actions.

Our analysis has effectively reduced the problem to a game with player set I, strategy sets Λ^i for $i \in I$, and payoff functions given by the objective payoffs. Each mNE corresponds to a pure Nash equilibrium of the reduced game. The supporting intuition is that, under an appropriate learning procedure, the role of model misspecification is to reduce the choice of strategies available to a decision maker. This reduces the choice of possible profitable deviations and consequently, if strategies that constitute a pure Nash equilibrium of the objective game $\mathscr{O}$ are still available to players in the game $\mathscr{G} = \langle \mathscr{O}, \mathscr{Q} \rangle$, then there exists a mNE in these strategies.

Proposition 1. *If $(\vec{x}^{i*})_{i \in I}$ is a pure Nash Equilibrium of the objective game and, for all $i \in I$, there exists $\theta^{i*} \in \Theta^i$ such that $(\theta^{i*}, \vec{x}^{i*}) \in \Lambda^i$, then $(\theta^{i*}, \vec{x}^{i*})_{i \in I}$ is a mNE.*

Proof. For given i, by definition of Nash equilibrium, $\vec{x}^{i*}$ is a best response under correct beliefs $(Q^i_\sigma)_{i \in I}$. This best response gives an expected payoff of $\Pi^i_\sigma(Q^i_\sigma)$. As $\Pi^i_\sigma(Q^i_\sigma) \geq \Pi^i_\sigma(Q^i_{\theta^i}, \vec{x}^i)$ for all $(\theta^i, \vec{x}^i) \in \Lambda^i$, and $\Pi^i_\sigma(Q^i_\sigma) = \Pi^i_\sigma(Q^i_{\theta^{i*}}, \vec{x}^{i*})$, it must be that $(\theta^{i*}, \vec{x}^{i*})$ solves (2). □

A question that remains is whether model misspecification should reduce the choice of strategies even further. Specifically, should it be permissible to consider mixing over elements of Λ^i? We can think of two interpretations of such a mixture. The first is that

a player mixing between $(\theta^{i1},\bar{x}^{i1})$ and $(\theta^{i2},\bar{x}^{i2})$ should act according to beliefs that are a convex combination of $Q^i_{\theta^{i1}}$ and $Q^i_{\theta^{i2}}$. However, it may be that neither $\bar{x}^{i1}$ nor $\bar{x}^{i2}$ is a best response to such beliefs. The second interpretation is the "mass action" interpretation of John Nash's PhD thesis (Nash, 1950a). Under this interpretation, a mixture between $(\theta^{i1},\bar{x}^{i1})$ and $(\theta^{i2},\bar{x}^{i2})$ 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 $\bar{x}^{i1}$ is a best response, and some chance of player i being of type $i2$, for whom $\bar{x}^{i2}$ is a best response. This latter interpretation motivates the following.

Let Ξ^i be the set of all probability measures over model-response pairs $(\theta^i,\bar{x}^i)$. Let ς^i denote an element of Ξ^i. Note that $\varsigma^i \in \Xi^i$ induces a distribution σ^i on $\mathbb{X}^i$ given by

$$\sigma^i(x^i|s^i) = \sum_{(\theta^i,\bar{x}^i)\in\Lambda^i:\bar{x}^i_{s^i}=x^i} \varsigma^i\left((\theta^i,\bar{x}^i)\right).$$

It follows that if $(\varsigma^i)_{i\in I}$ is given, then probabilities Q^i_σ under the true model are well defined and, consequently, so is Π^i_σ.

Definition 5. $(\varsigma^i)_{i\in I}$ is a (mixed) **misspecified Nash equilibrium** (mmNE) of game $\mathscr{G}$ if, for all players $i \in I$, for all $(\theta^{i*},\bar{x}^{i*})$ in the support of ς^i,

$$(\theta^{i*},\bar{x}^{i*}) \in \operatorname*{argmax}_{(\theta^i,\bar{x}^i)\in\Lambda^i} \Pi^i_\sigma(Q^i_{\theta^i},\bar{x}^i).$$

Proposition 2. *A mixed mNE exists.*

Proof. For all $i \in I$, for all $\bar{x}^i$ such that $(\theta^i,\bar{x}^i) \in \Lambda^i$ for some θ^i, choose one such θ^i. Denote the finite set of $(\theta^i,\bar{x}^i) \in \Lambda^i$ chosen this way by $\tilde{\Lambda}^i \subseteq \Lambda^i$.

The game $\tilde{G}$ with player set I, pure strategies $(\tilde{\Lambda}^i)_{i\in I}$ and payoffs equal to objective payoffs 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 $(\tilde{\varsigma}^{i*})_i$.

Define G to be identical to $\tilde{G}$ except that the strategy sets are Λ^i instead of $\tilde{\Lambda}^i$. For all $i \in I$, let $\varsigma^{i*} = \tilde{\varsigma}^{i*}$ on $\tilde{\Lambda}^i$ and $\varsigma^{i*}(\Lambda^i \setminus \tilde{\Lambda}^i) = 0$.

If $(\varsigma^{i*})_i$ is not a Nash equilibrium of G, there exists a profitable deviation for some player i to some $(\theta^{i1},\bar{x}^{i1}) \in \Lambda^i$. Note that by construction of $\tilde{\Lambda}^i$ there exists $\theta^i \in \Theta^i$ such that $(\theta^i,\bar{x}^{i1}) \in \tilde{\Lambda}^i \subseteq \Lambda^i$. Objective payoffs do not depend directly on beliefs, so if $(\theta^{i1},\bar{x}^{i1})$ is a profitable deviation from $(\varsigma^{i*})_i$, then $(\theta^i,\bar{x}^{i1})$ is also a profitable deviation from $(\varsigma^{i*})_i$. However, as $(\tilde{\varsigma}^{i*})_i$ and $(\varsigma^{i*})_i$ induce the same distributions over consequences, it must be that $(\theta^i,\bar{x}^{i1})$ is also a profitable deviation from $(\tilde{\varsigma}^{i*})_i$. Contradiction. Therefore, $(\varsigma^{i*})_i$, is a Nash equilibrium of G.

By definition of Nash equilibrium, if $(\theta^{i*}, \bar{x}^{i*})$ is in the support of ς^{i*}, then $(\theta^{i*}, \bar{x}^{i*}) \in \operatorname{argmax}_{(\theta^i, \bar{x}^i) \in \Lambda^i} \Pi^i_\sigma(\theta^i, \bar{x}^i)$. Therefore $(\varsigma^{i*})_i$ is a mNE. □

4. Generalized Bayes' rule

In this section, we describe a learning procedure, *generalized Bayes' rule*, that generalizes Bayes' rule, guarantees equally good performance (in terms of convergence speed) in well-specified learning problems, and learns models which give the best payoff in misspecified settings. The mNE we propose can be interpreted as the equilibrium resulting from a population of agents adopting generalized Bayes'. Like many regret-free algorithms, this approach is arguably closer to the way learning occurs in real-world situations because it is both less abstract and more robust than Bayes' rule. Less abstract because players learn directly from and about rewards and punishments rather than learning from observations about a hypothetical parameter characterizing a true distribution. More robust because, unlike Bayes' rule, it guarantees that a player will learn a model that induces an action that leads to as high an objective expected payoff as possible. Unlike other regret-free algorithms in the literature, the generalized Bayesian algorithm we propose allows us to naturally incorporate players' beliefs in the learning problem and it nests Bayes rule as a special case.

The generalized Bayesian algorithm has two steps. First, players transform their original beliefs to a new set of *entropified probabilities* (Grünwald, 1998) which incorporate payoffs that correspond to the best responses induced by each subjective belief. Second, players update their prior beliefs iteratively using (generalized) Bayes' rule on the set of entropified probabilities.

Here we briefly describe the entropification procedure, define generalized Bayes' rule and provide a simple proof of the fact that a player who follows this rule will learn to play actions that correspond to the highest objective payoff that can be justified by some model in his prior. That is, players learn to play as per the definition of mNE. We then illustrate the differing learning outcomes of Bayes' and generalized Bayes' rule by revisiting Example 6.1 (coin tosses).

Definition 6. For each $(\theta^i, \bar{x}^i) \in \Lambda^i$, the **entropified probability** of consequence y^i given s^i is

$$eQ^i_{(\theta^i, \bar{x}^i)}(y^i \mid s^i) = \frac{e^{\beta\, \pi^i(\bar{x}^i_{s^i}, y^i)}}{\int_{Y^i} e^{\beta\, \pi^i(\bar{x}^i_{s^i}, \hat{y}^i)} d\hat{y}^i},$$

where we will fix $\beta = 1$ for the rest of the paper.[3] For given σ, we similarly define eQ^i_σ by replacing $\bar{x}^i$ with an arbitrary best response to Q^i_σ.

Entropified probabilities are defined with reference to the finite set of best responses rather than the possibly infinite set of model-response pairs. Thus, dealing with entropified probabilities effectively reduces the domain of the learning problem to a finite set of classes of model-response pairs indexed by the set of subjectively non-dominated best responses.[4]

Given entropified probabilities, we can define the generalized likelihood of any model response pair $(\theta^i, \bar{x}^i)$ and the generalized Bayesian prior after t observations. These are simply the standard concepts applied to the entropified probabilities.

Definition 7. For each $(\theta^i, \bar{x}^i) \in \Lambda^i$, the **generalized likelihood** after t periods on $(s^i_\tau, y^i_\tau)^t_{\tau=0}$ is

$$gQ^i_{(\theta^i,\bar{x}^i)}\left((y^i_1, s^i_1), ..., (y^i_t, s^i_t)\right) = \prod_{\tau=1}^{t} \left(\frac{e^{\pi^i(\bar{x}^i_{s^i_\tau}, y^i_\tau)}}{\int_{\mathbf{Y}^i} e^{\pi^i(\bar{x}^i_{s^i_\tau}, \hat{y}^i)} d\hat{y}^i} \right).$$

The evolution of generalized Bayes' rule mimics Bayes' rule: the generalized posterior weight of each model is proportional to its generalized likelihood.

Definition 8. Given prior distribution μ^i_0 on Λ^i, the **generalized Bayesian prior** distribution $g\mu^i_t$ given observations $(s_\tau, y_\tau)^t_{\tau=0}$ is given by

$$g\mu^i_t(A) = \frac{\int_A gQ_{(\theta^i,\bar{x}^i)}\left((y^i_1, s^i_1), ..., (y^i_t, s^i_t)\right) d\mu^i_0(\theta^i, \bar{x}^i)}{\int_{\Lambda^i} gQ^i_{(\theta^i,\bar{x}^i)}\left((y^i_1, s^i_1), ..., (y^i_t, s^i_t)\right) d\mu^i_0(\theta^i, \bar{x}^i)},$$

for $A \subseteq \Lambda^i$.

Finally, we define the *entropified Kullback-Leibler divergence* as the measure of distance between entropified beliefs.

Definition 9. **Entropified Kullback-Leibler divergence** (eKLD):

$$eK^i(\sigma, \theta^i, \bar{x}^i) = \Pi^i_\sigma(Q^i_\sigma) - \Pi^i_\sigma(Q^i_{\theta^i}, \bar{x}^i).$$

$\Pi^i_\sigma(Q^i_{\theta^i}, \bar{x}^i)$ is player i's objective expected payoff when he plays the subjective best response $\bar{x}^i$ to beliefs $Q^i_{\theta^i}$ and the other players follow strategies $(\sigma^j)_{j\neq i}$. So the eKLD

[3]The appropriate value of β (a.k.a. the learning rate) is an active topic in the machine learning literature (e.g., Grünwald, 1998). WLOG, we set $\beta = 1$ because any time independent value of the learning rate converges to the same model when convergence occurs if the prior support is finite.

[4]This finiteness guarantees that assumptions **C1** – **C4** of Grünwald (1998) (π is bounded), Assumption 1 of Frick et al. (2019) and Assumption 1 of Esponda and Pouzo (2016) about the learning problem are satisfied.

measures the distance between model-response pairs in terms of differences in true expected payoffs.

It is easy to verify that if we replace $\pi^i(\bar{x}_{s^i}, y^i)$ by $\ln Q^i_{\theta^i}(y^i|s^i, \bar{x}_{s^i})$ in Definition 6, then we obtain the standard likelihood function. This analogy goes further. In fact, if probabilities over outcomes are independent of a player's action, then the entropified Kullback-Leibler divergence of Definition 9 is simply the definition of standard Kullback-Leibler divergence (see Section 5) applied to the entropified probabilities and generalized Bayes' rule coincides with Bayes rule.

The eKLD plays a similar role in the generalized Bayesian framework to the role played by standard Kullback-Leibler divergence in standard Bayes.[5] While Bayes rule converges to the model with the lowest Kullback-Leibler divergence (Berk, 1966), the generalized Bayes' rule converges to the model with the lowest eKLD (Proposition 3).

The following Proposition generalizes the results of (Berk, 1966), showing that generalized Bayes' rule eventually gives positive weight only to beliefs that support actions that minimize eKLD.

Proposition 3. *Let Q_σ be generated by σ. Write $A := \operatorname{argmin}_{(\theta^i,\bar{x}^i)\in\Lambda^i} eK^i(\sigma, \theta^i, \bar{x}^i)$. If $\mu^i_0(A) > 0$, then $g\mu^i_t(A) \to 1$ Q_σ-a.s. as $t \to \infty$.*

Proof. Write

$$a := \max_{(\theta^i,\bar{x}^i)\in\Lambda^i} \Pi^i_\sigma(Q^i_{\theta^i}, \bar{x}^i) \qquad \text{and} \qquad b := \max_{(\theta^i,\bar{x}^i)\in\Lambda^i\setminus A} \Pi^i_\sigma(Q^i_{\theta^i}, \bar{x}^i). \tag{3}$$

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

$$\begin{aligned}
g\mu^i_t(A) &= 1 - g\mu^i_t(\Lambda^i \setminus A) \\
&\underbrace{=}_{\substack{\text{by}\\ \text{Definition 8}}} 1 - \frac{\int_{\Lambda^i\setminus A} gQ_{(\theta^i,\bar{x}^i)}\left((y^i_1,s^i_1),\ldots,(y^i_t,s^i_t)\right) d\mu^i_0(\theta^i,\bar{x}^i)}{\int_{\Lambda^i} gQ^i_{(\theta^i,\bar{x}^i)}\left((y^i_1,s^i_1),\ldots,(y^i_t,s^i_t)\right) d\mu^i_0(\theta^i,\bar{x}^i)} \\
&\geq 1 - \frac{\int_{\Lambda^i\setminus A} gQ_{(\theta^i,\bar{x}^i)}\left((y^i_1,s^i_1),\ldots,(y^i_t,s^i_t)\right) d\mu^i_0(\theta^i,\bar{x}^i)}{\int_{A} gQ^i_{(\theta^i,\bar{x}^i)}\left((y^i_1,s^i_1),\ldots,(y^i_t,s^i_t)\right) d\mu^i_0(\theta^i,\bar{x}^i)} \\
&= 1 - \frac{\int_{\Lambda^i\setminus A} e^{\ln gQ_{(\theta^i,\bar{x}^i)}\left((y^i_1,s^i_1),\ldots,(y^i_t,s^i_t)\right)} d\mu^i_0(\theta^i,\bar{x}^i)}{\int_{A} e^{\ln gQ^i_{(\theta^i,\bar{x}^i)}\left((y^i_1,s^i_1),\ldots,(y^i_t,s^i_t)\right)} d\mu^i_0(\theta^i,\bar{x}^i)}
\end{aligned}$$

[5]Note that no model can give a higher objective payoff than the true model σ. That is, $\Pi^i_\sigma(Q^i_\sigma) \geq \Pi^i_\sigma(Q^i_{\theta^i}, \bar{x}^i)$ for all $(\theta^i, \bar{x}^i) \in \Lambda^i$. Therefore, $eK^i(\sigma, \theta^i, \bar{x}^i) \geq 0$ and equals 0, by definition, if and only if $\bar{x}^i \in \operatorname{argmax} \Pi^i_\sigma(Q^i_\sigma)$. So, the eKLD is a divergence in the space of entropified beliefs

$$\underbrace{=}_{\substack{\text{by}\\ \text{Definition 7}}} 1-\frac{\int_{\Lambda^i\setminus A} e^{t\sum_{\tau=1}^t \frac{1}{t}\pi(\bar{x}^i_{s^i_\tau},y_\tau)}\,d\mu_0^i(\theta^i,\bar{x}^i)}{\int_A e^{t\sum_{\tau=1}^t \frac{1}{t}\pi(\bar{x}^i_{s^i_\tau},y_\tau)}\,d\mu_0^i(\theta^i,\bar{x}^i)}$$

$$\underbrace{\approx}_{\substack{Q^i_\sigma\text{-a.s. for } t \text{ large}\\ \text{by SSLN}}} 1-\frac{\int_{\Lambda^i\setminus A} e^{t\,\Pi_\sigma(Q^i_{\theta^i},\bar{x}^i)}\,d\mu_0^i(\theta^i,\bar{x}^i)}{\int_A e^{t\,\Pi_\sigma(Q^i_{\theta^i},\bar{x}^i)}\,d\mu_0^i(\theta^i,\bar{x}^i)}$$

$$\underbrace{\geq}_{\text{by (3)}} 1-\frac{e^{t\,b}\,\mu_0^i(\Lambda^i\setminus A)}{e^{t\,a}\,\mu_0^i(A)} \quad \underbrace{\xrightarrow{t\to\infty}}_{\text{by } a>b} 1.$$

□

It follows that the generalized Bayesian prior will identify responses $\bar{x}^i$ which give the highest objective expected payoff, but that each such response may correspond to a multiplicity of beliefs. By using generalized Bayes, players pragmatically learn how to act to maximize their average payoff according to the true distribution, rather than which of their probabilistic models is the most accurate in some abstract sense. Comparing Definition 4 and Definition 9, it is clear that these learned model-response pairs are exactly those that occur in mNE.

4.1 Coin tosses revisited

Again consider the example from the introduction in which a decision maker learns a probabilistic model of coin tosses from amongst the models θ^{i1} (probability of heads is 0.45) and θ^{i2} (probability of heads is 0.9).

Bayes' rule. By standard arguments, the prior probability of θ^{i1} after t periods, calculated via Bayes' rule is

$$\begin{aligned}\mu_t^i(\theta^{i1}) &= \left(\frac{\mu_0^i(\theta^{i1})\prod_{\tau=1}^t\prod_{\omega=H,T} Q^i_{\theta^{i1}}(\omega)^{I_{y_\tau=\omega}}}{\mu_0^i(\theta^{i1})\prod_{\tau=1}^t\prod_{\omega=H,T} Q^i_{\theta^{i1}}(\omega)^{I_{y_\tau=\omega}}+\mu_0^i(\theta^{i2})\prod_{\tau=1}^t\prod_{\omega=H,T} Q^i_{\theta^{i2}}(\omega)^{I_{y_\tau=\omega}}}\right)\\ &= \frac{1}{1+\frac{\mu_0^i(\theta^{i2})}{\mu_0^i(\theta^{i1})}e^{\sum_{\tau=1}^t\sum_{\omega=H,T} I_{y_\tau=\omega}\ln\frac{Q^i_{\theta^{i2}}(\omega)}{Q^i_{\theta^{i1}}(\omega)}}}\\ &= \frac{1}{1+\frac{\mu_0^i(\theta^{i2})}{\mu_0^i(\theta^{i1})}e^{t\left(\frac{1}{t}\sum_{\tau=1}^t\sum_{\omega=H,T} I_{y_\tau=\omega}\ln\frac{Q^i_\sigma(\omega)}{Q^i_{\theta^{i1}}(\omega)}-\frac{1}{t}\sum_{\tau=1}^t\sum_{\omega=H,T} I_{y_\tau=\omega}\ln\frac{Q^i_\sigma(\omega)}{Q^i_{\theta^{i2}}(\omega)}\right)}}\end{aligned}$$

$$\underset{\text{for t large}}{\approx^{Q^i_\sigma\text{-a.s.}}} \left(\frac{1}{1+\frac{\mu^i_0(\theta^{i2})}{\mu^i_0(\theta^{i1})} e^{t\left(K(\sigma,\theta^{i1})-K(\sigma,\theta^{i2})\right)}} \right)$$
$$\underset{}{\to^{Q^i_\sigma\text{-a.s.}}} \begin{cases} 1 & \text{if} \quad K(\sigma,\theta^{i1})-K(\sigma,\theta^{i2})<0 \\ 0 & \text{if} \quad K(\sigma,\theta^{i1})-K(\sigma,\theta^{i2})>0 \end{cases}.$$

For $Q^i_\sigma(H)=0.7$, a quick calculation shows that $K(\sigma,\theta^{i1})-K(\sigma,\theta^{i2})<0$, so that $\mu^i_t(\theta^{i1})\to 1$ as $t\to\infty$. Accordingly, the Bayesian player becomes certain that tails is more likely than heads and bets on tails for all large t. These learned beliefs ensure him an objective expected payoff of 0.3, which is lower than the objective expected utility of 0.7 that he would have obtained had he learned the θ^{i2} model.

Generalized Bayes' rule. Adapting the previous argument (see also the proof of Lemma 3), the generalized Bayesian prior probability of θ^{i1} after t periods is

$$\mu^e_t(\theta_1) \underset{\text{for t large}}{\approx^{Q^i_\sigma\text{-a.s.}}} \left(\frac{1}{1+\frac{\mu_0(\theta_2)}{\mu_0(\theta_2)} e^{t\left(eK(\sigma,\theta^{i1},T)-eK(\sigma,\theta^{i2},H)\right)}} \right)$$
$$\underset{}{\to^{Q^i_\sigma\text{-a.s.}}} \begin{cases} 1 & \text{if} \quad eK(\sigma,\theta^{i1},T)-eK(\sigma,\theta^{i2},H)<0 \\ 0 & \text{if} \quad eK(\sigma,\theta^{i1},T)-eK(\sigma,\theta^{i2},H)>0 \end{cases},$$

which implies that the generalized Bayesian prior converges to a Dirac distribution on the parameter with the lowest eKL divergence from the truth. For $Q^i_\sigma(H)=0.7$, our decision maker correctly learns that betting on heads is more profitable than betting on tails and that he is better off acting under the beliefs $Q_{\theta^{i2}}$ than he is acting under the beliefs $Q_{\theta^{i1}}$.

5. Comparing mNE with Berk-Nash

In this section, we formally define the Berk-Nash equilibrium concept of Esponda and Pouzo (2016) and provide an alternative definition of mNE, equivalent to our original definition, which eases the comparison between the two equilibrium concepts.

Esponda and Pouzo (2016) define an equilibrium concept for mispecified models with Bayesian players by leveraging the observation that a Bayesian learner would learn the models that are "closest" to the objective distribution in terms of minimizing Kullback-Leibler divergence (Berk, 1966). Berk-Nash equilibrium places probability one on the set of subjective beliefs that minimize Kullback-Leibler divergence.

Formally, these are the relevant definitions.

Definition 10. **Weighted Kullback-Leibler divergence** (wKLD):

$$K^i(\sigma,\theta^i)=\sum_{(s^i,x^i)\in\mathbb{S}^i\times\mathbb{X}^i} E_{Q^i_\sigma(\cdot|s^i,x^i)}\left[\ln\frac{Q^i_\sigma(Y^i|s^i,x^i)}{Q^i_{\theta^i}(Y^i|s^i,x^i)}\right]\sigma^i(x^i|s^i)\,p_{S^i}(s^i). \tag{4}$$

Definition 11. A strategy profile σ is a **Berk-Nash equilibrium** (BNE) of game $\mathscr{G}$ if, for all players $i\in I$, there exists $\mu^i\in\Delta(\Theta^i)$ such that

(i) σ^i is optimal given μ^i, and

(ii) If $\hat{\theta}^i$ is in the support of μ^i then $\hat{\theta}^i\in\operatorname{argmin}_{\theta^i\in\Theta^i}K^i(\sigma,\theta^i)$.

In well-specified learning settings with a proper information structure, BNE coincides with NE and every NE is a mNE. The reverse inclusion does not hold because mNE does not uniquely pin down beliefs: the same actions can be, and typically are, a best response to more than one belief. In misspecified learning settings, a BNE is observationally equivalent to some mNE if and only if the beliefs of each player are a *useful* model in that they lead to the highest payoff amongst the models subjectively believed possible by the decision maker.

To further compare BNE and mNE, we give a definition of mNE, equivalent to our original definition, that is directly linked to eKLD, and thus to the generalized Bayesian approach that was described in Section 4.

Definition 12. A profile $(\theta^{i*},\bar{x}^{i*})_{i\in I}$, is a (pure) **misspecified Nash equilibrium** (mNE) of game $\mathscr{G}$ if, for all players $i\in I$,[6]

(i) $(\theta^{i*},\bar{x}^{i*})\in\Lambda^i$, and

(ii) $(\theta^{i*},\bar{x}^{i*})\in\operatorname{argmin}_{(\theta^i,\bar{x}^i)\in\Lambda^i} eK^i(\sigma,\theta^i,\bar{x}^i)$.

Definition 12 highlights that mNE is similar to BNE in requiring that players' beliefs attach probability one to the set of subjective distributions over consequences that are "closest" to the objective distribution. The difference between BNE and mNE is the way in which the respective concepts define the distance between distributions. Specifically, BNE uses wKLD whereas mNE uses eKLD.

These notions of distance, wKLD and eKLD, and consequently BNE and mNE respectively, rest on different learning paradigms. Bayes' rule has become the standard in economics due to its simplicity and the sound axiomatic foundation (Ghirardato 2002; Gilboa and Marinacci 2011). Furthermore, in well-specified learning problems, its derivation is almost tautological when we consider empirical frequencies. A natural question is why should a player deviate from Bayes' rule to favor other rules such as generalized Bayes? The main

[6]It is easy to verify that this definition is equivalent to Definition 4 because, by Definition 7 $\operatorname{argmin}_{(\theta^i,\bar{x}^i)\in\Lambda^i} eK^i(\sigma,\theta^i,\bar{x}^i)=\operatorname{argmax}_{(\theta^i,\bar{x}^i)\in\Lambda^i}\Pi^i_\sigma(Q^i_{\theta^i},\bar{x}^i)$.

reason is the misspecification risk. When the learning environment is potentially misspecified, Bayes' rule does not guarantee robust choices. As Dawid (1982) eloquently put it:

> *"If a subjective distribution P attaches probability zero to a non-ignorable event, and if this event happens, then P must be treated with suspicion and modified or replaced."*

The generalized Bayesian approach we describe has the pragmatic advantage of being as good as regular Bayes' (comparable convergence rate to the truth) in well specified learning problems, as well as ensuring minimum regret choices even in misspecified learning environments.

The axiomatic and the pragmatic rationality criteria coincide in well-specified learning environments but differ in misspecified problems. These two criteria should be seen as complementary, with the preferred criterion depending on the specifics of the decision environment. The former is appropriate in situations in which an agent is not subject to an external criterion of performance. In this case, a set of axioms can jointly determine an agent's preferences and beliefs. The latter is appropriate for cases in which an agent's decisions are evaluated according to an external criterion of performance (e.g. Sharpe ratio for portfolio managers, calibration for weather forecasters). Because the criterion pins down agent preferences, a pragmatic agent should internalize this constraint in his decision problem and choose a prediction rule that is optimal for his preferences (Massari, 2020).

One important external criterion is fitness. If we consider an evolutionary model in which payoffs correspond to fitness, then players with high payoffs reproduce more than players with low payoffs. In this case, the pragmatic criterion is most appropriate when it comes to identifying the long-run equilibria of the population dynamic. We show this formally in Section 7.

6. Examples

6.1 Coin tosses

Here we discuss the illustrative example from the introduction. A decision maker guesses the outcome of a coin toss, $\mathbb{X}^i = \{H, T\}$. 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. There are no signals. Hence we have $Q^i_\sigma(H|x^i) = Q^i_\sigma(H) = 0.7$ for all σ, x^i. Payoffs are given by $\pi^i(x,y) = 1$ if $x = y$ and $\pi^i(x,y) = 0$ if $x \neq y$. The parameter set is $\Theta^i = \{\theta^{i1}, \theta^{i2}\}$ and we let $Q^i_{\theta^{i1}}(H|x^i) = Q^i_{\theta^{i1}}(H) = 0.45$ and

$Q^i_{\theta^{i2}}(H|x^i) = Q^i_{\theta^{i2}}(H) = 0.9$ for all x^i. Note that T is the unique best response to beliefs $Q^i_{\theta^{i1}}$, whereas H is the unique best response to beliefs $Q^i_{\theta^{i2}}$ or to the true model Q^i_σ.

Misspecified Nash equilibrium. The set of subjectively non-dominated model-response pairs is given by $\Lambda^i = \{(\theta^{i1}, T), (\theta^{i2}, H)\}$. We obtain objective expected payoffs

$$\Pi^i_\sigma(Q^i_{\theta^{i1}}, T) = 0.3, \qquad \Pi^i_\sigma(Q^i_{\theta^{i2}}, H) = 0.7, \qquad \Pi^i_\sigma(Q^i_\sigma) = 0.7.$$

It follows that the unique mNE is (θ^{i2}, H). At this equilibrium, the decision maker obtains an expected objective payoff of 0.7 and thus he would correctly guess the outcome of the coin toss 0.7 of the time.

An easy calculation shows that, for all σ, we have

$$eK^i(\sigma, \theta^{i1}, T) = 0.7 - 0.3 = 0.4, \qquad eK^i(\sigma, \theta^{i2}, H) = 0.7 - 0.7 = 0.$$

So, this is the pair that will be learned by applying generalized Bayes' rule. The player learns what he cares about: payoffs.

Berk-Nash equilibrium. Substituting into (10) we obtain, for $\theta^i \in \Theta^i$,

$$K^i(\sigma, \theta^i) = E_{Q^i_\sigma(\cdot)}\left[\ln \frac{Q^i_\sigma(Y^i)}{Q^i_{\theta^i}(Y^i)}\right]. \tag{5}$$

Therefore, as Q^i_σ is independent of σ, we have that, for all σ,

$$K^i(\sigma, \theta^{i1}) = 0.7\left(\ln \frac{0.7}{0.45}\right) + 0.3\left(\ln \frac{0.3}{0.55}\right) \approx 0.13$$

and

$$K^i(\sigma, \theta^{i2}) = 0.7\left(\ln \frac{0.7}{0.9}\right) + 0.3\left(\ln \frac{0.3}{0.1}\right) \approx 0.15.$$

So, at BNE, it must be that model θ^{i1} is believed with probability one. These are the beliefs that would be learned by applying Bayes' rule. The unique best response for θ^{i1} is T, so the unique BNE has $\sigma^i(H) = 0$, $\sigma^i(T) = 1$, supported by the belief $\mu^i(\theta^i) = 1$. At this equilibrium, the player obtains an expected objective payoff of 0.3 and thus he would correctly guess the outcome of the coin toss only 0.3 of the time. The player learns the model with the highest likelihood, not the model that grants him higher payoffs if believed true.

6.2 Arrow-Debreu securities

We extend the example of the preceding subsection so that the decision maker chooses a share $x^i \in \mathbb{X}^i = \{0, 0.01, \ldots, 0.99, 1\}$ of a unit of Arrow-Debreu security to invest in outcome H. The remainder is invested in outcome T. Similar to before, $y = f(x, \omega) = \omega$, where $\omega = H$ with probability p_H and $\omega = T$ with probability $1 - p_H$. There are no signals and the decision maker's action does not affect outcome probabilities. Hence we have $Q^i_\sigma(H|x^i) = Q^i_\sigma(H) = p_H$ for all σ, x^i. The decision maker is aware that his action does not affect outcome probabilities and has Bernoulli beliefs parametrized by $\Theta = \{0, 0.01, \ldots, 0.99, 1\}$, so that $\forall \theta^i \in \Theta^i, Q^i_{\theta^i}(H) = \theta^i$. Payoffs are given by $\pi^i(x^i, H) = u(x^i)$ and $\pi^i(x^i, T) = u(1 - x^i)$, where u is a utility function.

Misspecified Nash equilibrium. The set of mNE are all pairs $(\theta^i, x^i) \in \Lambda^i$ that maximize $E_{Q^i_\sigma(\cdot)} \pi^i(x^i, Y^i)$. By definition, these are also the pairs that minimize the eKLD.

Berk-Nash equilibrium. To find a BNE, choose θ^i to minimize (5), then choose a strategy in which any action x^i played with positive probability maximizes $E_{Q^i_{\theta^i}(\cdot)} \pi^i(x^i, Y^i)$.

In general, a Bayesian will learn different beliefs from those held at mNE. So, mNE and BNE can differ. However, when the model is correctly specified, a Bayesian will learn correct beliefs and thus his best responses will be those described above for mNE. In the case of log utility, minimizing eKLD is equivalent to minimizing wKLD and therefore beliefs at mNE will be identical to those that would be learned by a Bayesian.

Correctly specified model. If there exists $\theta^{i*} \in \Theta^i$ such that $Q^i_{\theta^{i*}} = Q^i_\sigma$, then wKLD is minimized at θ^{i*}. Then BNE is a NE. Furthermore, $E_{Q^i_{\theta^{i*}}(\cdot)} u(x^i) = E_{Q^i_\sigma(\cdot)} u(x^i)$, therefore we also have that (θ^{i*}, x^i) is a mNE. Note that there may also exist other mNE that choose the same actions but are based on incorrect beliefs. However, if (θ^i, x^i) is a mNE, then (θ^{i*}, x^i) is also a mNE and σ^i, $\sigma^i(x^i) = 1$, is a BNE supported by the belief $\mu^i(\theta^{i*}) = 1$.

Log utility. Now, let $u(\cdot) = \ln(\cdot)$. When this is the case, for any $(\theta^i, x^i) \in \Lambda^i$, we have that $x^i = Q^i_{\theta^i}(H)$ and $1 - x^i = Q^i_{\theta^i}(T)$. That is, the share of asset invested in H equals the subjective probability of H. Readers will recognize this as the celebrated Kelly criterion. Consequently, eKLD will be minimized by $(\theta^i, x^i) \in \Lambda^i$ that maximize $E_{Q^i_\sigma(\cdot)} \ln Q^i_{\theta^i}(Y^i)$. This is equivalent to minimizing (5), therefore (θ^i, x^i) is a mNE if and only if σ^i, $\sigma^i(x^i) = 1$, is a BNE supported by the belief $\mu^i(\theta^i) = 1$.

6.3 Monopoly with unknown demand

Here we consider Example 2.1 of Esponda and Pouzo (2016). A monopolist chooses a price $x^i \in \mathbb{X}^i = \{2, 10\}$ that generates demand $y^i = f(x^i, \omega) = \phi_0(x^i) + \omega$, where ω is a mean-zero shock with distribution $p \in \Delta(\Omega)$. It is assumed that $\phi_0(2) = 34$ and $\phi_0(10) = 2$.

There are no signals. The payoff is $\pi^i(x^i, y^i) = x^i y^i$.

The monopolist's uncertainty about p and f is described by a parametric model f_θ, p_θ, where $y = f_\theta(x^i, \omega) = a - bx^i + \omega$ is the demand function, $\theta = (a,b) \in \Theta$ 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 $\phi_0(x^i) = \phi_{\theta_0}(x^i)$ for all $x^i \in \mathbb{X}^i$. This gives $\theta_0 = (a_0, b_0) = (42,4) \notin \Theta$ and therefore the monopolist has a misspecified model. Note that, as there are no other players, the conditional objective distribution $Q^i_\sigma(\cdot|x^i)$ does not depend on σ and is normal with mean $\phi_0(x^i)$ and unit variance. Similarly, $Q_\theta(\cdot|x^i)$ is normal with mean $\phi_\theta(x^i) = a - bx$ and unit variance.

Misspecified Nash equilibrium. By substituting the true model parameters into the payoff function, we obtain expected objective payoffs from playing model-response pairs $(Q^i_{\theta^i}, x^i)$.

$$\begin{aligned}\Pi^i_\sigma(Q^i_{\theta^i}, x^i) &= E[\pi^i(x^i, y^i)] = E\left[x^i(42 - 4x^i + \omega))\right] \\ &= x^i(42 - 4x^i) + x^i E(\omega) = \begin{cases} 68 \; if \; x^i = 2 \\ 20 \; if \; x^i = 10 \end{cases}\end{aligned}$$

Consequently, the set of mNE is the set of all pairs $(\theta^i, 2) \in \Lambda^i$. As there is only one player, the set of mmNE is the set of all mixtures on these pairs. As there are no actions other than 2 and 10, with correct beliefs the monopolist should always choose action 2, so $\Pi^i_\sigma(Q^i_\sigma) = 68$. Applying Definition 9, we see that $eK^i(\sigma, \theta^i, 2) = 68 - 68 = 0$ for all pairs $(\theta^i, 2) \in \Lambda^i$ and $eK^i(\sigma, \theta^i, 2) = 68 - 20 = 48$ for all pairs $(\theta^i, 10) \in \Lambda^i$. So, these are exactly the pairs that can be learned by applying generalized Bayes' rule.

Comparison to a Bayesian learner. Esponda and Pouzo (2016) show that if a Bayesian monopolist consistently chooses action 2, then he will learn beliefs to which the only best response is action 10. Conversely, if he consistently chooses action 10, then he will learn beliefs to which the only best response is action 2. Consequently, the only way to reconcile Bayesian learning with best response is for the monopolist to mix between actions 2 and 10. If he does this in the right proportions, then he can learn beliefs that give both actions as best responses.[7]

The Bayesian disconnect. The reason that Bayesian learning gives rise to the complexity above is a fundamental disconnect between (i) the idea of payoff maximization that underpins Nash-style concepts, and (ii) the learning procedure, which ignores payoffs and maximizes log-likelihood. Even when the Bayesian monopolist plays both actions 2 and 10,

[7]The frequencies with which each action is played under σ^i are chosen so that the θ^i that minimizes $K^i(\sigma, \theta^i)$ makes the monopolist (subjectively) indifferent between the actions. The unique Berk-Nash equilibrium of this game is $\sigma^i = (35/36, 1/36)$ with supporting beliefs given by the parameters $\theta^i = (40, 10/3)$.

his learning procedure does not pay attention to the difference in expected objective payoffs from each action. This is in contrast to the ideas and learning procedures that underpin belief selection for mNE.

7. Evolutionary stability

In this section we show that if a population follows a learning rule that leads to equilibrium behavior that is not a mNE, then the population is vulnerable to invasion by players who follow a different rule. Conversely, if the learning rule leads to a strict mNE, then the population is robust to invasion by other learning rules. The stability concept we use adapts the idea of an evolutionarily stable state (Taylor and Jonker, 1978) for a situation in which the aspect of the environment under evolutionary pressure is neither strategies (Weibull, 1995), nor preferences (Samuelson, 2001), nor agency (Newton, 2017b), but rather the learning rule that players follow.

Formally, consider a population comprising a unit mass of players. Every period, members of the population are matched into groups of size I to play a game $\mathscr{G}$ that we assume, without loss of generality, to be symmetric.[8] Members of the population follow a *learning rule*, a function that maps a player's *history* of actions $x_0^i,\ldots,x_{t-1}^i$, outcomes $y_0^i,\ldots,y_{t-1}^i$ and payoffs $\pi^i(x_0^i,y_0^i),\ldots,\pi^i(x_{t-1}^i,y_{t-1}^i)$ to a distribution ς_t^i over model-response pairs $\lambda_t^i=(\theta_t^i,\vec{x}_t^i)\in\Lambda^i$.

Every period, let each member of the population generate a model-response pair according to the learning rule, before being matched to play the game. Matching is uniform, so that the probability of any given opponent of a player following a given model-response pair equals the share of the players in the population with that model-response pair. Write ς_t as the aggregate distribution over model-response pairs in the population at time t.

Consider an alternative, *mutant*, learning rule that would have led to ς_t^m as the distribution over model-response pairs in the population at time t. For given $\varepsilon\in[0,1)$, let $\varsigma_t^\varepsilon=(1-\varepsilon)\varsigma_t+\varepsilon\varsigma_t^m$. That is, ς_t^ε is the aggregate distribution over model-response pairs in the population at time t when $1-\varepsilon$ of the population follows the incumbent learning rule and ε of the population follows the mutant learning rule. Let *fitness* $F(\varsigma,\varsigma_t^\varepsilon)$ be the expected payoff that a player in the population would obtain from following ς when the population as a whole follows ς_t^ε.

[8] To accomodate asymmetric games, we need only consider a game that is ex-ante symmetric but has players' signals assign them to player positions in the asymmetric game. Further, note that the arguments leading to Proposition 4 do not rely in any substantive way on the pre-mutant population containing only a single learning rule, an assumption which is maintained to avoid a considerable amount of distracting further notation.

Definition 13. A **population learning evolutionarily stable equilibrium** (PLESE) is a learning rule and a history of play at some arbitrary period, say t, such that when the learning rule is used to extend the history, (i) ς_τ remains constant for all $\tau > t$, and (ii) there does not exist $\bar{\tau} > t$, $\bar{\varepsilon} > 0$ and a mutant learning rule such that for all $\varepsilon < \bar{\varepsilon}$, we have $F(\varsigma_{\bar{\tau}}^m, \varsigma_{\bar{\tau}}^\varepsilon) > F(\varsigma_{\bar{\tau}}, \varsigma_{\bar{\tau}}^\varepsilon)$.

Condition (i) of the definition specifies that we are indeed concerning ourselves with an equilibrium. Condition (ii) specifies that it is impossible for a small number of mutants that follow a different learning rule to invade the population and obtain higher average fitness than the incumbents. Aside from considerations of time, the second part of the definition is the standard definition of an evolutionarily stable state (Taylor and Jonker, 1978).

It turns out that there is a close relationship between play that can be sustained in a PLESE and our concept of mNE. The reason for this is that non-mNE play can be destabilized by the invasion of minimum regret learning algorithms such as the generalized Bayes' rule.

Proposition 4.
(i) If a PLESE is such that $\varsigma_\tau = \varsigma^*$ for all sufficiently large τ, then $(\varsigma^*)_{i\in I}$ is an mNE. (ii) If $(\varsigma^*)_{i\in I}$ is a strict mNE (i.e. the maximizers in the definition are unique), then for any given t, there exists a PLESE such that $\varsigma_\tau = \varsigma^*$ for all $\tau \geq t$.

Proof. Part (i). Consider a PLESE such that $\varsigma_\tau = \varsigma^*$ for all sufficiently large τ, but $(\varsigma^*)_{i\in I}$ is not an mNE. The definition of (mixed) mNE (Definition 5) then implies that ς^* places strictly positive weight on some model-response pairs that do not maximize payoff when every other player follows ς^*.

Consider generalized Bayes' rule as discussed in Section 4. As the assumed PLESE eventually has ς^* played by the population every period, Proposition 3 implies that generalized Bayes' rule eventually places almost all weight on model-response pairs that maximize payoff when every other player follows ς^*. Write the output of generalized Bayes' rule as ς_t^m. That is, generalized Bayes' rule will be our mutant rule. Write $\varsigma_t^\varepsilon = (1-\varepsilon)\varsigma^* + \varepsilon\varsigma_t^m$. Note that $\varsigma_t^0 = \varsigma^*$. The above argument shows that for large enough t, ς_t^m gives a strictly higher payoff than ς^* when every other player plays ς^*. That is, $F(\varsigma_t^m, \varsigma_t^0) > F(\varsigma^*, \varsigma_t^0)$. Fix such a t.

As $\varepsilon \to 0$, the probability of any player in the population being matched only with non-mutants approaches 1. Hence, $F(\varsigma_t^m, \varsigma_t^\varepsilon) \to F(\varsigma_t^m, \varsigma_t^0)$ and $F(\varsigma^*, \varsigma_t^\varepsilon) \to F(\varsigma^*, \varsigma_t^0)$. Therefore, for small enough ε, we have $F(\varsigma_t^m, \varsigma_t^\varepsilon) > F(\varsigma^*, \varsigma_t^\varepsilon)$. This violates Definition 13(ii), so we cannot have started with a PLESE. Contradiction.

Part (ii). Let $(\varsigma^*)_{i\in I}$ be a strict mNE and choose an arbitrary history of play until time t such that $\varsigma_t = \varsigma^*$. Let the incumbent learning rule be such that from periods $\tau = t+1$ onwards, $\varsigma^i_\tau = \varsigma^i_t$ for each player i in the population. Therefore, in the absence of mutants, the aggregate distribution of model-response pairs in the population is ς^* from period t onwards. Thus Definition 13(i) is satisfied.

As the mNE is strict, for any mutant rule, any $\tau > t$, we must have $F(\varsigma^m_\tau, \varsigma^0_\tau) < F(\varsigma^*, \varsigma^0_\tau)$. Hence, for small enough ε, we have $F(\varsigma^m_t, \varsigma^\varepsilon_t) < F(\varsigma^*, \varsigma^\varepsilon_t)$. Thus Definition 13(ii) is satisfied.

□

Some remarks are in order regarding the limits of this result. As remarked earlier in the paper, there are clear analogies between the study of NE in correctly specified games and the study of mNE in misspecified games. Consequently, insights gained from the study of the evolutionary stability of NE (see, e.g. Dekel et al., 2007; Heifetz et al., 2007; Ok and Vega-Redondo, 2001) are applicable to the study of mNE. Crucially, when a player is matched with a mutant that has newly appeared in the population, he does not recognize the mutant. That is, he does not condition his play on whether or not an opponent is a mutant. Relating this to Proposition 4(i), this means that incumbents cannot punish invading regret-minimizing mutants and prevent them from destabilizing non-mNE behavior. Regarding Proposition 4(ii), it means that when an mNE is being played, invading mutants cannot attain higher payoffs than the population average due to their matched opponents adjusting their strategies.

Another important assumption is that of uniform matching. In correctly specified models, it is known that assortativity in matching can lead to non-NE behavior (Alger and Weibull, 2013; Bergstrom, 1995; Eshel and Cavalli-Sforza, 1982). If mutants match more frequently with other mutants due to either exogenous factors or an evolved homophily (Newton, 2017a), then there is a bias towards efficiency in play. This is because mutants can evolve that are both highly likely to interact with others like themselves and behave in a way that obtains high payoffs when one's opponent does similarly. Such arguments can also be applied to misspecified settings.

References

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

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.

Brunnermeier, M. K., Gollier, C., and Parker, J. A. (2007). Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review*, 97(2):159–165.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal expectations. *American Economic Review*, 95(4):1092–1118.

Csaba, D. and Szoke, B. (2018). Learning with misspecified models. *mimeo*.

Dawid, A. P. (1982). The well-calibrated bayesian. *Journal of the American Statistical Association*, 77(379):605–610.

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

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.

Foster, D. P. and Young, H. P. (2006). Regret testing: Learning to play Nash equilibrium without knowing you have an opponent. *Theoretical Economics*, 1(3):341–367.

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. (2019). Stability and robustness in misspecified learning models.

Germano, F. and Lugosi, G. (2007). Global Nash convergence of Foster and Young's regret testing. *Games and Economic Behavior*, 60(1):135–154.

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

Grünwald, P. and Langford, J. (2007). Suboptimal behavior of Bayes and MDL in classification under misspecification. *Machine Learning*, 66(2-3):119–149.

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.

Grünwald, P. D. (1998). *The minimum description length principle and reasoning under uncertainty*. PhD thesis, Quantum Computing and Advanced System Research.

Grünwald, P. D. (2007). *The minimum description length principle*. MIT press.

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.

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.

Massari, F. (2019). Ambiguity, robust statistics, and Raiffa's critique. SSRN Working Paper Series 3388410.

Massari, F. (2020). Price probabilities: A class of bayesian and non-bayesian prediction rules. *Economic Theory*.

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

Newton, J. (2017b). Shared intentions: The evolution of collaboration. *Games and Economic Behavior*, 104:517 – 534.

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.

Pradelski, B. S. and Young, H. P. (2012). Learning efficient Nash equilibria in distributed systems. *Games and Economic behavior*, 75(2):882–897.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry*. World Scientific.

Rissanen, J. (2007). *Information and complexity in statistical modeling*. Springer Science & Business Media.

Samuelson, L. (2001). Introduction to the evolution of preferences. *Journal of Economic Theory*, 97(2):225–230.

Sandholm, W. H. (2010). *Population games and evolutionary dynamics*. Economic learning and social evolution. Cambridge, Mass. MIT Press.

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

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting*, 1:135–196.

Vovk, V. G. (1990). Aggregating strategies. *Proc. of Computational Learning Theory, 1990*.

Weibull, J. (1995). *Evolutionary game theory*. MIT Press.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Young, H. P. (2009). Learning by trial and error. *Games and economic behavior*, 65(2):626–643.