# Misspecified Learning and Evolutionary Stability<sup>\*</sup>

Kevin  $He^{\dagger}$  Jonathan Libgober<sup>‡</sup>

First version: December 20, 2020 This version: September 22, 2024

#### Abstract

Toward explaining the persistence of biased inferences, we propose a framework to evaluate competing (mis)specifications in strategic settings. Agents with heterogeneous (mis)specifications coexist and draw Bayesian inferences about their environment through repeated play. The relative stability of (mis)specifications depends on their adherents' equilibrium payoffs. A key mechanism is the *learning channel*: the endogeneity of perceived best replies due to inference. We characterize when a rational society is only vulnerable to invasion by some misspecification through the learning channel, and highlight new stability phenomena that arise due to the learning channel. As an application, we show how our framework can be used to endogenize coarse analogy classes in centipede games.

**Keywords**: misspecified Bayesian learning, endogenous misspecifications, evolutionary stability, analogy classes

<sup>\*</sup>We thank Cuimin Ba, Thomas Chaney, Sylvain Chassang, In-Koo Cho, Krishna Dasaratha, Andrew Ellis, Ignacio Esponda, Mira Frick, Drew Fudenberg, Alice Gindin, Ryota Iijima, Yuhta Ishii, Philippe Jehiel, Pablo Kurlat, Elliot Lipnowski, Jonny Newton, Filippo Massari, Andy Postlewaite, Philipp Sadowski, Alvaro Sandroni, Grant Schoenebeck, Joshua Schwartzstein, Philipp Strack, Carl Veller, and various conference and seminar participants for helpful comments. Byunghoon Kim provided excellent research assistance. Kevin He thanks the California Institute of Technology for hospitality when some of the work on this paper was completed, and the University Research Foundation Grant at the University of Pennsylvania for financial support. Jonathan Libgober thanks Yale University and the Cowles foundation for their hospitality.

<sup>&</sup>lt;sup>†</sup>University of Pennsylvania. Email: hesichao@gmail.com

<sup>&</sup>lt;sup>‡</sup>University of Southern California. Email: libgober@usc.edu

# 1 Introduction

In many economic settings, people draw *misspecified inferences* about the world: while they learn from data, they exclude the true data-generating process from consideration. For instance, past work has documented a number of prevalent statistical biases. Reasoning about economic fundamentals under the spell of these biases constitutes misspecified learning. Economists have become increasingly interested in the implications of Bayesian learning under particular misspecifications, for the most part taking them to be exogenously imposed.

Compared with other errors and mistakes, a distinctive component of misspecified *learning* is using data to form beliefs about the world. This raises a natural question: how does the ability to draw inferences affect the viability of such mistakes? We introduce an *evolutionary approach* to answer this question in strategic settings. Specifically, we associate the viability of a particular (mis)specification with the objective payoffs of the individuals who adopt it. In contrast to contemporaneous papers that use the same criterion in single-agent decision problems (Fudenberg and Lanzani, 2022; Frick, Iijima, and Ishii, 2024), our key innovation is to focus on games, where this objective performance depends on strategic behavior in equilibrium.

Our main message is that the *learning channel* — i.e., the ability for agents to learn and draw (possibly wrong) inferences from data — adds new ways for biased individuals to develop strategically beneficial commitments. When these biased agents are initially uncertain about some parameters of their environment, the learning channel endogenously determines their beliefs and hence their perceived best replies through equilibrium feedback. By contrast, the learning channel is absent for agents who have a fixed but distorted belief about all parameters of their environment, since they never use data to update their beliefs. The learning channel thus generates additional flexibility in equilibrium beliefs, and our contribution is to emphasize two implications of such flexibility:

- 1. Due to the greater flexibility, misinference can confer strategic benefits in cases where dogmatic beliefs do not.
- 2. Misspecified learners are *polymorphic*: agents with a fixed bias may be weak in one environment but become stronger in another due to (endogenous) changes in beliefs.

Our main results fall under one of these two themes. On the former, we find general conditions under which no dogmatically wrong belief can persist in a rational society, but some misspecified agents can nevertheless do strictly better than rational incumbents through the learning channel. On the latter, polymorphism makes it harder to predict the viability of a given error across different economic environments. Without the learning channel, we find some sufficient conditions that let us use the welfare of a given error in one society to extrapolate that it will not persist in another society. But these conclusions no longer hold for biased agents who may develop different beliefs in different environments.

Formally, our general framework encodes specifications in *models* that delineate feasible beliefs about the stage game. These models serve as the basic unit of cultural transmission. The model's adherents think that one of the model parameters describes the true stage game. They estimate the best-fitting parameter which determines their subjective preference. Models rise and fall in prominence based on the objective welfare of adherents, as higher payoffs confer greater evolutionary success.

Society consists of the adherents of multiple competing models who match up to play the stage game every period. We introduce the concept of a *zeitgeist* to capture the social interaction structure — the sizes of the subpopulations with different models and the matchmaking technology that pairs up opponents to play the game. Agents can identify which subpopulation their opponent is from, and (correctly) know that the game they play is orthogonal to the type of opponent.<sup>1</sup> Our framework assumes that the agents might face one of several possible games and therefore richer models can in principle help as they allow agents to adapt their behavior more. Conditional on the stage game, in equilibrium each agent forms a Bayesian belief about the game using data from all of her interactions, playing a subjective best response against every type of opponent given this belief.

We define the *evolutionary stability* of model A against model B based on whether model A has a weakly higher average equilibrium payoff than model B when the population share of model A is close to 1, with the average taken over the different stage games. This criterion is familiar from past work following what is known as the *indirect evolutionary approach*. Under this approach, evolution acts on some trait that determines best responses, as opposed to actions. We emphasize that our stability concepts reduce to standard notions under this approach when inference is absent. Rather, our contribution is to apply it to the selection of *models* that contain multiple feasible beliefs about the environment.

<sup>&</sup>lt;sup>1</sup>If the players think that the stage game can change depending on their opponent, then this would give additional channels for biases to invade a rational society. Our framework focuses on how the learning channel that plays a distinctive role in misspecified learning affects the viability of errors.

The ability to draw inferences within a model (as opposed to committing to a fixed belief) is necessary for misspecifications to defeat rationality in some contexts. In Section 3.1, we characterize environments where the correctly specified model is only evolutionarily fragile against invading models that allow for inferences. Our argument constructs an optimal misspecified model for invading a rational society. This misspecification resembles an "illusion of control" bias, where agents think the outcomes they get in a game only depend on their own strategy and not on the opponent's strategy. The model has the property that its adherents end up adopting the optimal commitment against a correctly specified opponent game-by-game. Misinference thus becomes a channel to tailor commitments to the true game. The correctly specified model is evolutionarily fragile against this misspecified model with uniform matching unless the former already gets the Stackelberg payoff in every game.

Our next contribution is to highlight certain stability phenomena that can only emerge with non-dogmatic misspecified models. We highlight the aforementioned polymorphism of misspecified models—that they can appear weak against rational incumbents in one environment and yet grow stronger and successfully invade the rational society in another environment. The reason is that due to the learning channel, an adherent of a misspecified model may come to hold different beliefs about parameters of the underlying stage game, and thus (endogenously) adopt different best-reply functions when facing outcomes generated from different strategy profiles. Thus, changes in the population structure and matching process influence the perceived best replies for adherents of misspecified models.

We explore implications of the endogenous interaction between inference and evolutionary forces that emerges due to polymorphism. First, polymorphism enables a new stability phenomenon that we call *stability reversals*. Two models exhibit stability reversal if:

- 1. Whenever model A is dominant, its adherents strictly outperform model B's adherents not only on average, but even conditional on the opponent's type; and
- 2. Whenever model B is dominant, its adherents strictly outperform model A's adherents on average

In the absence of inference, condition (1) would imply that A outperforms B regardless of the two subpopulations' sizes. But this no longer holds when inference is possible. The reason is that the adherents of model B might make an evolutionarily advantageous inference only when they are matched up with each other sufficiently often. Thus, even if condition (1) held, model B might still drive out model A if model B adherents reach some critical mass.

Polymorphism also manifests in a non-monotonicity of stability with respect to matching assortativity. As discussed in Alger and Weibull (2013), the assortativity parameter can represent the degree of homophily in a society or the frequency of interaction with kin. Various versions of the idea that high assortativity selects for cooperative agents and low assortativity selects for competitive ones date back to at least Hamilton (1964a,b). But this simple dichotomous perspective becomes complicated with misspecifications. The reason is that in our framework, the preferences agents seek to maximize are *endogenously determined in equilibrium*, due to the learning channel. Because the adherents of a misspecified model can draw different misinferences about a fixed game's parameters when facing data generated by different opponent actions, one model may be favored over another only at *intermediate* levels of assortativities, but not favored at either very low or very high levels. Thus, a particular bias might only survive in moderately homophilous societies — a novel empirical implication of misspecified inference.

Our hope is that the extension of the indirect evolutionary approach to accommodate (misspecified) inferences will enable this framework to speak to richer applications. We pursue this agenda in our companion paper, He and Libgober (2024), studying the selection of misspecified higher-order beliefs in a Cournot duopoly game with incomplete information. We showcase the potential applied value of our equilibrium concept in Section 4 by studying the selection of *analogy classes* in extensive form games (Jehiel, 2005), based on the payoffs for players with different analogy classes. Under analogy-based reasoning, players believe (incorrectly) that opponents choose the same action distribution within a given grouping of nodes (i.e., analogy class), inferring this distribution to be the empirical frequency within the analogy class. Our approach predicts not only that analogy-based reasoning may invade a correctly-specified society, but also that the two can coexists. Solving for the corresponding stable population composition, we obtain sharp predictions for the relative prominence of analogy-based reasoning as a function of the underlying interaction. While we believe these results highlight the potential practical value of our framework, we admit our current paper only scratches the surface. We thus hope our framework can guide formal analysis of the possible advantages of misspecifications in applications beyond those considered so far.

# 2 Environment and Stability Concept

We start with our formal stability concept, defining *equilibrium zeitgeist* to determine the evolutionary fitness of specifications that coexist in a society. We consider a separate notion, *equilibrium zeitgeist with strategic uncertainty*, in Section 4, when we allow agents to draw inferences about others' strategies in addition to learning about the fundamentals. Appendix C provides a combined learning foundation for both equilibrium concepts, but in the main text we primarily focus on the steady-state characterization. Section 2.6.1 sketches the framework without inference, which has been studied in past work.

# 2.1 Objective Primitives

Agents in a population repeatedly match to play a stage game, which is a symmetric twoplayer game with a common, metrizable strategy space A. There is a set of possible states of nature  $G \in \mathcal{G}$ , called *situations*. The strategy choices  $a_i, a_{-i} \in \mathbb{A}$  of i and -i, together with the situation, stochastically generate consequences  $y_i, y_{-i} \in \mathbb{Y}$  from a metrizable space  $\mathbb{Y}$ . Each agent i's consequence  $y_i$  determines her utility, according to a common utility function  $\pi : \mathbb{Y} \to \mathbb{R}$ , which we take to be Borel measurable with respect to the sigma algebra generated by the topology on  $\mathbb{Y}$ . The objective distribution over consequences is  $F^{\bullet}(a_i, a_{-i}, G) \in \Delta(\mathbb{Y})$ , with an associated density or probability mass function denoted by  $f^{\bullet}(a_i, a_{-i}, G)$ , where  $f^{\bullet}(a_i, a_{-i}, G)(y) \in \mathbb{R}_+$  for each  $y \in \mathbb{Y}$ . We suppress G from  $f^{\bullet}$  and  $F^{\bullet}$  when  $|\mathcal{G}| = 1$ . Note that we allow for  $\mathbb{Y}$  to be completely general outside of the previous technical restrictions.

This setup captures mixed strategies (if A is the set of mixtures over some pure actions), incomplete-information games (if S is a space of private signals, A a space of actions, and  $A = A^S$  is the set of signal-contingent actions), and even asymmetric games. For the latter, we consider the "symmetrized" version where each player is placed into each role with equal probability (see Section 4 for one application where agents play an asymmetric game).

# 2.2 Models and Parameters

Throughout this paper, we will take the strategy space  $\mathbb{A}$ , the set of consequences  $\mathbb{Y}$ , and the utility function over consequences  $\pi$  to be common knowledge among the agents. But, agents are unsure about how play in the stage game translates into consequences: that is, they have fundamental uncertainty about the function  $(a_i, a_{-i}) \mapsto F^{\bullet}(a_i, a_{-i}, G)$ . While we assume that

the situation G is unobserved, we allow agents to draw inferences about it by observing the consequences from the matches they face.

We focus on the case where society consists of two<sup>2</sup> observably distinguishable groups of agents, A and B, who may behave differently in the stage game due to different beliefs about how y is generated. The two groups of agents entertain different *models* of the world that help resolve their fundamental uncertainty. A model  $\Theta$  is a collection of data-generating processes  $F : \mathbb{A}^2 \to \Delta(\mathbb{Y})$  about how strategy profiles translate into consequences for the agent, with the processes parameterized by some  $\gamma \in \Gamma$ . We thus view each model as a subset of  $(\Delta(\mathbb{Y}))^{\mathbb{A}^2}$ , sometimes writing  $F_{\gamma}$  for the process given by the parameter  $\gamma$ . For every  $(a_i, a_{-i}) \in \mathbb{A}^2$  and every process  $F \in \Theta$ ,  $F(a_i, a_{-i})$  is a Borel measure on  $\mathbb{Y}$ ; we assume it has associated with it a density or probability mass function  $f(a_i, a_{-i}) : \mathbb{Y} \to \mathbb{R}_+$ . We assume the parameter space  $\Gamma$  of each model is metrizable, and that for every fixed  $(a_i, a_{-i})$  the map  $\gamma \mapsto \mathbb{E}_{y \sim F_{\gamma}(a_i, a_{-i})}[\pi(y)]$  is Borel measurable.<sup>3</sup>

Each agent enters society with a persistent model, which depends entirely on whether she is from group A or group B. We refer to the agents who are endowed with a given model as the *adherents* of that model. We call  $\Theta = \{F^{\bullet}(\cdot, \cdot, G) : G \in \mathcal{G}\}$  the *minimal correctly specified* model. A model may exclude the true  $F^{\bullet}(\cdot, \cdot, G)$  that produces consequences, at least in some situation G. In this case, the model is *misspecified*.

One possibility our framework accommodates is that  $|\Theta| = 1$ , in which case the model is a *singleton*. Singleton models are of special interest in our setting because they reflect agents who hold dogmatic beliefs and hence do not draw inferences from data or adapt their preferences differently across different situation. Since the situation is itself unobserved, preferences can only vary by drawing inferences after observing consequences.

## 2.3 Zeitgeists

To study competition between two models, we must describe the social composition and interaction structure in the society where learning takes place. We have in mind a setting where each agent plays the stage game with a random opponent in every period and uses her

 $<sup>^{2}</sup>$ We view the case of two groups of agents with different models as the natural starting point, though it is straightforward to generalize Definition 1 to the case of more than two groups.

<sup>&</sup>lt;sup>3</sup>Note that this measurability property would follow from the measurability of the mapping  $\gamma \mapsto f_{\gamma}(a_i, a_{-i})(y)$  for each fixed  $(a_i, a_{-i}, y)$ , under some further restrictions necessary to apply Fubini's theorem to the function  $(a_i, a_{-i}, y) \mapsto \pi(y)f(a_{-i}, a_{-i})(y)$ 

personal experience in these matches to calibrate the most accurate parameter within her model. A zeitgeist describes the corresponding landscape.

**Definition 1.** Fix models  $\Theta_A$  and  $\Theta_B$ . A zeitgeist  $\mathfrak{Z} = (\mu_A(G), \mu_B(G), p, \lambda, a(G))_{G \in \mathcal{G}}$ consists of: (1) for each situation G, a belief over parameters for each model,  $\mu_A(G) \in \Delta(\Theta_A)$ and  $\mu_B(G) \in \Delta(\Theta_B)$ ; (2) relative sizes of the two groups in the society,  $p = (p_A, p_B)$ with  $p_A, p_B \geq 0$ ,  $p_A + p_B = 1$ ; (3) a matching assortativity parameter  $\lambda \in [0, 1]$ ; (4) for each situation G, each group's strategy when matched against each other group,  $a = (a_{AA}(G), a_{AB}(G), a_{BA}(G), a_{BB}(G))$  where  $a_{g,g'}(G) \in \mathbb{A}$  is the strategy that an adherent of  $\Theta_g$ plays against an adherent of  $\Theta_{q'}$  in situation G.

A zeitgeist outlines the beliefs and interactions among agents with heterogeneous models living in the same society. Part (1) captures the beliefs of each group. Parts (2) and (3) determine social composition and social interaction—the relative prominence of each model and the probability of interacting with one's own group versus with the overall population. In each period,  $\lambda$  is the probability an agent's opponent is from her own group, and  $1 - \lambda$  is the probability the opponent is drawn uniformly from the population. Therefore, an agent from group g has probability  $\lambda + (1 - \lambda)p_g$  of being matched with an opponent from her own group, and a complementary chance of being matched with an opponent from the other group. Part (4) describes behavior in the society. Note that a zeitgeist describes each group's situation-contingent belief and behavior, since agents may infer different parameters and thus adopt different subjective best replies in different situations. However, it is worth emphasizing that since the situation is not observed directly, situations influence strategies by changing the distribution over consequences (and hence the inferences made).

### 2.4 Equilibrium Zeitgeists

A model's fitness corresponds to the equilibrium payoffs of its adherents. An equilibrium zeitgeist (EZ) imposes optimality conditions on inference and behavior in a zeitgeist. Optimality of behavior requires each player to best respond given her beliefs, and optimality of inference requires that the support of each player's belief only contains the "best-fitting" parameter from her model in the sense of minimizing Kullback-Leibler (KL) divergence, using the observed distribution of consequences.

We now formalize this criterion. For two distributions over consequences,  $\Phi, \Psi \in \Delta(\mathbb{Y})$ with density or probability mass functions  $\psi, \phi$ , define the KL divergence from  $\Psi$  to  $\Phi$  as  $D_{KL}(\Phi \parallel \Psi) := \int \phi(y) \ln\left(\frac{\phi(y)}{\psi(y)}\right) dy$ . Recall that every data-generating process F, like the true fundamental  $F^{\bullet}(\cdot, \cdot, G)$ , outputs a distribution over consequences for every profile of own play and opponent's play,  $(a_i, a_{-i}) \in \mathbb{A}^2$ . For data-generating process F, let  $K(F; a_i, a_{-i}, G) := D_{KL}(F^{\bullet}(a_i, a_{-i}, G) \parallel F(a_i, a_{-i}))$  be the KL divergence from the expected distribution  $F(a_i, a_{-i})$  to the objective distribution  $F^{\bullet}(a_i, a_{-i}, G)$  under the play  $(a_i, a_{-i})$  and situation G. For a Borel measure  $\mu$  over parameters, let  $U_i(a_i, a_{-i}; \mu)$  represent *i*'s subjective expected utility under the belief that the true parameter is drawn according to  $\mu$ . That is,  $U_i(a_i, a_{-i}; \mu) := \mathbb{E}_{F \sim \mu}(\mathbb{E}_{y \sim F(a_i, a_{-i})}[\pi(y)]).$ 

**Definition 2.** A zeitgeist  $\mathfrak{Z} = (\mu_A(G), \mu_B(G), p, \lambda, a(G))_{G \in \mathcal{G}}$  is an equilibrium zeitgeist (EZ) if, for every  $G \in \mathcal{G}$  and  $g, g' \in \{A, B\}$ ,  $a_{g,g'}(G) \in \underset{a_i \in \mathbb{A}}{\operatorname{arg\,max}} U_i(a_i, a_{g',g}(G); \mu_g(G))$  and, for every  $g \in \{A, B\}$ , belief  $\mu_g(G)$  is supported on

 $\underset{F \in \Theta_g}{\operatorname{arg\,min}} \left\{ (\lambda + (1 - \lambda)p_g) \cdot K(F; a_{g,g}(G), a_{g,g}(G), G) + (1 - \lambda)(1 - p_g) \cdot K(F; a_{g,-g}(G), a_{-g,g}(G), G) \right\}$ 

where -g means the group other than g.

This definition requires agents from group g to choose a subjective best response against their opponents, given the belief  $\mu_g$  about the fundamental uncertainty. No matter which group the agent is matched against, these choices are always made to selfishly maximize her (individual) subjective utility function. Each agent's belief  $\mu_g$  is supported on the parameters in her model that minimize a weighted KL-divergence objective in situation G, with the data from each type of match weighted by the probability of confronting this type of opponent. The use of KL-divergence minimization as the inference procedure is standard in the misspecified Bayesian learning literature, as in Esponda and Pouzo (2016). We note that here we assume inference occurs separately across situations. This reflects situation persistence, with agents having enough data to establish new beliefs and behavior if the situation were to change. Our learning foundation in Appendix C justifies this situation-by-situation updating, but we omit the details here as it otherwise plays no role in our results.

### 2.5 Evolutionary Stability of Models

Given a distribution  $q \in \Delta(\mathcal{G})$  and an EZ, we define the *fitness* of each model as the expected objective payoff of its adherents in the EZ when G is drawn according to q. We have in mind

an evolutionary story where the relative success of the two models depends on their relative fitness: for instance, agents may play a large number of games in different periods possibly facing different situations over time, and models of those agents with higher total objective payoffs are more likely to be adopted in the next generation.<sup>4</sup> Given this notion of fitness, our question of interest is: Can the adherents of a *resident model*  $\Theta_A$ , starting at a position of social prominence, always repel an invasion from a small  $\epsilon$  mass of agents who adhere to a *mutant model*  $\Theta_B$ ?

Evolutionary stability depends on the fitness of models  $\Theta_A$ ,  $\Theta_B$  in EZs with  $p_A = 1 - \epsilon$ ,  $p_B = \epsilon$  for small  $\epsilon > 0$ .

**Definition 3.** Say  $\Theta_A$  is evolutionarily stable [fragile] against  $\Theta_B$  under  $\lambda$ -matching if there exists some  $\bar{\epsilon} > 0$  so that for every  $0 < \epsilon \leq \bar{\epsilon}$ , there is at least one EZ with models  $\Theta_A, \Theta_B$ ,  $p = (1 - \epsilon, \epsilon)$ , matching assortativity  $\lambda$  and, in all such EZs,  $\Theta_A$  has a weakly higher [strictly lower] fitness than  $\Theta_B$ .

Evolutionary stability is when  $\Theta_A$  has higher fitness than  $\Theta_B$  in all EZs, and evolutionary fragility is when  $\Theta_A$  has lower fitness in all EZs.<sup>5</sup> These two cases give sharp predictions about whether a small share of mutant-model invaders might grow in size, across all equilibrium selections. We fix these rather stringent definitions of stability and fragility, and focus on showing in Section 3 how the presence of the learning channel can generate new stability / fragility phenomena. A third possible case, where  $\Theta_A$  has lower fitness than  $\Theta_B$  in some but not all EZs, corresponds to a situation where the mutant model may or may not grow in the society, depending on the equilibrium selection.

### 2.6 Discussion

Before using this framework to illustrate our main contributions—on tailored commitments and polymorphism, mentioned in the introduction—we clarify some important aspects of it.

<sup>&</sup>lt;sup>4</sup>One subtlety is that fitness maximization may require not maximizing expected payoffs, but rather some other function of the distribution of payoffs, if shocks can be correlated (Robson, 1996). However, our microfoundation in Appendix C posits that situations are fixed for long stretches of time, with no correlated shocks across matches, making the expectation an appropriate measurement of fitness.

<sup>&</sup>lt;sup>5</sup>If the set of EZs is empty, then  $\Theta_A$  is neither evolutionarily stable nor evolutionarily fragile against  $\Theta_B$ .

#### 2.6.1 Comparison to Other Evolutionary Frameworks

We apply the "indirect evolutionary approach" (see Robson and Samuelson (2011)) to settings where agents can draw inferences (especially misspecified inferences). When  $|\Theta| = 1$  and  $|\mathcal{G}| = 1$ , our framework reduces to the setup studied by the literature on preference evolution (Alger and Weibull, 2019), since singleton models are equivalent to subjective preferences. But in general, models with multiple parameters allow agents to adapt their beliefs (which determine their subjective preferences) endogenously.

Allowing for multiple situations is the most direct way for inference to be beneficial. With only a single situation, any steady state outcome that emerges for some  $\Theta$  can also emerge when  $|\Theta| = 1$ . That said, one could also study settings with multiple situations without inference (see Güth and Napel (2006) for an example of such an exercise).

#### 2.6.2 Framework Assumptions

An important assumption is that agents (correctly) believe the economic fundamentals (represented by G) do not vary depending on which group they are matched against. That is, the mapping  $(a_i, a_{-i}) \mapsto \Delta(\mathbb{Y})$  describes the stage game that they are playing, and agents know that they always play the same stage game even though opponents from different groups may use different strategies in the game. As a result, the agent's experiences in games against both groups of opponents jointly resolve the same fundamental uncertainty about the environment.<sup>6</sup> If adherents were able to believe the fundamentals changed depending on their opponent, then this would give a trivial way for in-group preferences to emerge and also trivialize the question of which errors could invade. For expositional simplicity, we do not consider this elaboration.

We comment on some other modeling assumptions. First, our framework assumes that agents can identify which group their matched opponent belongs to, though we do not assume that agents know the data-generating processes contained in other models or that they are capable of making inferences using other models. Observability assumptions are common in the literature on the indirect evolutionary approach; see Alger and Weibull (2019) and Dekel et al. (2007) for discussions. While full observability can be relaxed in several ways,

<sup>&</sup>lt;sup>6</sup>We note that play between two groups g and g' is not a Berk-Nash equilibrium (Esponda and Pouzo, 2016), since adherents from one group draw inferences about the game's parameters from the matches against the other group, which may adopt a different strategy. A Berk-Nash equilibrium between groups g and g' would require inferences to *only* be made from data generated in the match between g and g'.

we expect the main insights to carry through given sufficient observability. In our context, one key assumption that makes our approach tractable is that players do not change their inferences in response to seeing their opponents' actions. In other words, players do not "read into" what others do when learning. This particular assumption seems plausible in many cases. Consider hedge funds that regularly trade against each other in a variety of settings. Funds hold differing philosophies, with some focusing on fundamental analysis and others on technical analysis.<sup>7</sup> But, simply observing another fund's actions would not lead a technical analysi are complex forecasting systems that involve calibrating sophisticated models and take many years of training and experience to master. In settings such as these, agents need not know how others' models work even after identifying who they are.

Second, EZs as presented abstract away from the issues surrounding learning others' strategies. However, we study an extension in Section 4 allowing agents to be misspecified about others' strategies and hold wrong beliefs about these strategies in equilibrium.

Lastly, even as agents adjust their beliefs and behavior to achieve optimality, population proportions  $p_A$  and  $p_B$  remain fixed. We imagine a world where the relative prominence of models changes much more slowly than the rate of convergence to an EZ. This assumption about the relative rate of change in the population sizes follows the previous work on evolutionary game theory (See Sandholm (2001) or Dekel, Ely, and Yilankaya (2007)).

# 3 Stability Implications of the Learning Channel

We now illustrate some stability phenomena that distinguish misspecified learning from dogmatic beliefs in our framework. These phenomena underscore our two main contributions mentioned in the introduction. The main novelty of our framework relative to past work on the indirect evolutionary approach is that agents maximize endogenously determined subjective preferences, not exogenously fixed ones. The *learning channel* refers to this endogenous preference formation, and we showcase some of its unique implications in this section, toward making the aforementioned contributions.

The learning channel adds new ways for biased individuals to develop strategic commitments in games. First, unlike agents with fixed subjective preferences, misspecified learners

<sup>&</sup>lt;sup>7</sup>In practice, each fund's model about the financial market is well known to other market participants, as it is always prominently marketed to their clients.

can develop situation-specific commitments that are better tailored to the stage game. We show this mechanism expands the scope of invading rational societies. Second, misspecified learners can exhibit polymorphism as they form different beliefs in different environments. This leads to new stability phenomena and adds nuance to extrapolations of the welfare implications of a misspecified model across different societies, relative to that of a distorted subjective preference.

## 3.1 When Is Learning Necessary to Defeat Rationality?

Our first result characterizes when misspecified models can *only* invade a rational society when inference is possible. More precisely, when does there exist a distribution over situations such that the correctly specified model is not evolutionarily fragile against any singleton model, but it is evolutionarily fragile against some models with multiple parameters? The following example illustrates:

**Example 1.** Suppose there are two situations,  $G_A$  and  $G_B$ , which are equally likely, and consequences  $\mathbb{Y} = \{g, b\}$ , with u(g) = 1 and u(b) = 0. Suppose that the probability a given player obtains g given an action profile and situation is determined by the table below.

$G_A$	$a_1$	$a_2$	$a_3$	
$a_1$	0.1, 0.1	0.1, 0.1	0.1, 0.11	
$a_2$	0.1, 0.1	0.3, 0.3	0.1,  0.1	
$a_3$	0.11, 0.1	0.1, 0.1	0.2,  0.2	

$G_B$	$a_1$	$a_2$	$a_3$
$a_1$	0.11, 0.11	0.5,  0.5	0.12,  0.4
$a_2$	0.5,  0.5	0.12,  0.12	0.14,  0.55
$a_3$	0.4, 0.12	0.55, 0.14	0.4, 0.4

Taking  $\lambda = 0$ , we show the correctly specified model is not evolutionarily fragile against any singleton mutant model  $\Theta = \{F\}$ . Consider the case where the resident correctly specified model has a population size of 1. It will obtain an objective fitness 0.35 if  $(a_2, a_2)$  in situation  $G_A$  and  $(a_3, a_3)$  in situation  $G_B$  are played, as these are Nash equilibria. But under the singleton model  $\{F\}$ , one of the three must hold:

- If  $a_3$  is a best response to  $a_3$  under F, there is an EZ where  $(a_3, a_3)$  is always the outcome, and the expected fitness is 0.3 < 0.35
- If  $a_2$  is a best response to  $a_3$  under F, there is an EZ where  $(a_2, a_3)$  is played by the mutant and resident in  $G_B$ , so the mutant's payoff is at most  $\frac{1}{2} \cdot 0.3 + \frac{1}{2} \cdot 0.14 < 0.35$

• If  $a_1$  is a best response to  $a_3$  under F, then there is an EZ where  $(a_1, a_3)$  is played by the mutant and resident in  $G_A$ , so the mutant's payoff is at most  $\frac{1}{2} \cdot 0.1 + \frac{1}{2} \cdot 0.55 < 0.35$ .

Thus, for all  $\epsilon > 0$  sufficiently small, there is an EZ where the resident model does strictly better, and so the minimal correctly specified model is not evolutionarily fragile against any singleton. However, consider the misspecified model  $\Theta = \{F_A, F_B\}$ , where both  $F_A$  and  $F_B$ depend only on one's own strategies and not the opponent's. Under  $F_A$ ,  $a_1$ ,  $a_2$ , and  $a_3$  lead to consequence g with probabilities 0.1, 0.3, and 0.2 respectively. Under  $F_B$ , playing  $a_1$ ,  $a_2$ , and  $a_3$  lead to consequence g with probabilities 0.5, 0.14, and 0.4 respectively.

The resident minimal correctly specified model is evolutionarily fragile against this misspecified model. Note that the mutants never choose  $a_3$ , since this is dominated under both  $F_A$  and  $F_B$ . Next, note that mutants would play  $a_2$  when believing  $F_A$  and  $a_1$  when believing  $F_B$ . We show these mutants play  $a_2$  in  $G_A$  and  $a_1$  in  $G_B$  against the resident in every EZ when the population share of the mutants is sufficiently small. Indeed, if mutants were to play  $a_1$  in situation  $G_A$ , the correctly specified residents would best respond with  $a_3$  in  $G_A$ . The mutants then learn  $F_A$  in  $G_A$ , and would then deviate to  $a_2$ . If mutants play  $a_2$  in situation  $G_B$ , once again the residents best respond with  $a_3$  in  $G_B$ , and the mutants learn  $F_B$ . But under  $F_B$ , the mutants believe they should deviate to  $a_1$ . These arguments rule out all other EZ behavior, so the mutants must play  $a_2$  in  $G_A$  and  $a_1$  in  $G_B$ . In this EZ, mutant fitness is  $(1/2) \cdot 0.3 + (1/2) \cdot 0.5 = 0.4 > 0.35$ , higher than the resident's fitness.

The previous example has two notable features: (1) A misspecification resembling an "illusion of control" bias whereby individuals believe consequences only depend on their own actions, not their opponent's actions; (2) inferences that lead misspecified agents to misperceive the Stackelberg action in each situation as the strictly dominant action. Models of this form allow us to determine when the ability to draw misinferences strictly expands the scope for invasion against rationality. Intuitively, if mutants can adopt the optimal commitment situation-by-situation, then the learning channel allows the mutants to tailor their commitment. But a mutant with only one model (i.e., an exogenous subjective preference) lacks the flexibility to play differently in different situations.

Some notation is needed to state the general result. Consider an arbitrary situation G. We let  $v_G^{\text{NE}} \in \mathbb{R}$  be the highest symmetric Nash equilibrium payoff in G, when agents choose strategies from  $\mathbb{A}$ . For each  $a_i \in \mathbb{A}$ , we let  $\underline{BR}(a_i, G)$  be a rational best response against the strategy  $a_i$  in situation G, breaking ties *against* the user of  $a_i$ . Let  $\bar{v}_G \in \mathbb{R}$  be the Stackelberg equilibrium payoff in situation G, breaking ties against the Stackelberg leader, i.e.,

$$\bar{v}_G := \max_{a_i} U_i(a_i, \underline{BR}(a_i, G), F^{\bullet}(G)).$$
(1)

Call the strategy  $\bar{a}_G$  that maximizes Equation (1) the Stackelberg strategy in situation G. We assume the Stackelberg strategy is unique in each situation, and furthermore that there is a unique rational best response to  $\bar{a}_G$  in each situation G', where possibly  $G \neq G'$ . Finally, let  $v_G^b$  denote the worst equilibrium payoff of an agent with the subjective best-response correspondence b when she plays against a rational opponent in situation G.<sup>8</sup>

We impose two identifiability conditions:

**Definition 4.** Situation identifiability is satisfied if for every  $a_i, a_{-i} \in \mathbb{A}$  and  $G \neq G'$ , we have  $F^{\bullet}(a_i, a_{-i}, G) \neq F^{\bullet}(a_i, a_{-i}, G')$ . Stackelberg identifiability is satisfied if whenever  $G \neq G'$  and  $a_{-i}, a'_{-i}$  are rational best responses to  $\bar{a}_G$  in situations G and G', we have  $F^{\bullet}(\bar{a}_G, a_{-i}, G) \neq F^{\bullet}(\bar{a}_G, a'_{-i}, G')$ .

Under situation identifiability, a minimal correctly specified agent can identify the true situation. Under Stackelberg identifiability, playing  $\bar{a}_G$  in situation G leads to different consequences than playing the same strategy in situation  $G' \neq G$ , provided the opponent chooses the rational best response to the strategy.

The following result presents our characterization of when the learning channel is required for misspecified models to outperform rationality, for some distribution over situations:

**Theorem 1.** Suppose  $\lambda = 0$ , there are finitely many situations, and there is a symmetric Nash equilibrium in  $\mathbb{A} \times \mathbb{A}$  for every situation G.

- 1. If there is no point  $(u_G)_{G \in \mathcal{G}}$  in the convex hull of  $\{(v_G^b)_{G \in \mathcal{G}} \mid b : \mathbb{A} \rightrightarrows \mathbb{A}\}$  with the property that  $u_G \ge v_G^{NE}$  for every  $G \in \mathcal{G}$ , then there exists a full-support distribution  $q \in \Delta(\mathcal{G})$  so that the correctly specified model is not evolutionarily fragile against any singleton model.
- 2. If  $v_G^{NE} < \bar{v}_G$  for some G, situation identifiability and Stackelberg identifiability hold, and there are finitely many strategies, then there exists a model  $\hat{\Theta}$  such that the correctly

<sup>&</sup>lt;sup>8</sup>More formally, given correspondence  $b : \mathbb{A} \rightrightarrows \mathbb{A}$ , let  $v_G^b \in \mathbb{R}$  be defined as *i*'s lowest payoff across all strategy profiles  $(a_i, a_{-i})$  such that  $a_i \in b(a_{-i})$  and  $a_{-i}$  is a rational response to  $a_i$  in situation *G*. If no such profile exists, let  $v_G^b = -\infty$ .

specified model is evolutionarily fragile against  $\Theta$  under any full-support distribution  $q \in \Delta(\mathcal{G})$ .

Whenever both conditions in Theorem 1 are satisfied, there is some distribution over situations so that the minimal correctly specified model is evolutionarily fragile against some mutant model, but not evolutionarily fragile against any *singleton* mutant model. In these environments, the ability to adapt preferences endogenously to the relevant situation (i.e., the learning channel) is a necessary condition for an invading mutant to displace the rational incumbent. The proof of Theorem 1 also illustrates the *kind* of misspecification that can outperform rationality—specifically, those which yield optimal Stackelberg commitments situation-by-situation.

One environment where  $v_G^{\text{NE}} = \bar{v}_G$  for every situation G is when the agents face decision problems — that is, a player's payoff in every situation G is independent of the action of the matched opponent. When all situations are decision problems, Theorem 1 does not apply: in fact, the correctly specified model is not evolutionarily fragile against any other model, regardless of whether such invaders learn from data. But in general, when the situations G are games that feature strategic interactions between players, Theorem 1 characterizes when the possibility highlighted by Example 1 arises; and indeed, one can check that Example 1 satisfies both conditions of Theorem 1. Hence, this result shows that mutants with misspecified models cannot in general be represented simply as mutants with fixed subjective best-response correspondences.

# 3.2 Stability Reversals

We now illustrate polymorphism and highlight one consequence of it: the potential for a greater indeterminacy in the emergence of stable biases. For expositional simplicity, we assume that  $|\mathcal{G}| = 1$  throughout this section. We will refer to a model's *conditional fitness against group g*, i.e., the expected payoff of the model's adherents in matches against group *g*.

**Definition 5.** Two models  $\Theta_A$ ,  $\Theta_B$  exhibit stability reversal if (i) in every EZ with  $\lambda = 0$  and  $(p_A, p_B) = (1, 0)$ ,  $\Theta_A$  has strictly higher conditional fitness than  $\Theta_B$  against group A opponents and against group B opponents, but also (ii) in every EZ with  $\lambda = 0$  and  $(p_A, p_B) = (0, 1)$ ,  $\Theta_B$  has strictly higher fitness than  $\Theta_A$ .

When  $p_B = 0$ , how  $\Theta_A$  performs against  $\Theta_B$  does not actually affect group A's fitness.

Condition (i) encodes the strong requirement that  $\Theta_A$  outperforms  $\Theta_B$  even on the zeroprobability event of being matched against a  $\Theta_B$  opponent. A stability reversal occurs if this stronger requirement holds (when  $\Theta_A$  dominates in society), and yet  $\Theta_B$  still strictly outperforms  $\Theta_A$  if  $\Theta_B$  starts from a position of prominence.

We begin with two general results on when stability reversals *cannot* emerge. First, it cannot emerge without the learning channel:

**Proposition 1.** Suppose  $|\mathcal{G}| = 1$ . Two singleton models (i.e., two subjective preferences in the stage game) cannot exhibit stability reversal.

Additionally, stability reversals cannot emerge in decision problems. We show this by introducing a class of games where strategic interactions do not matter:

**Definition 6.** A model  $\Theta$  is strategically independent if for all  $\mu \in \Delta(\Theta)$ ,  $\underset{a_i \in \mathbb{A}}{\operatorname{rg max}} U_i(a_i, a_{-i}; \mu)$  is the same for every  $a_{-i} \in \mathbb{A}$ .

The adherents of a strategically independent model believe that while an opponent's action may affect their utility, it does not affect their best response.

**Proposition 2.** Suppose  $|\mathcal{G}| = 1$ , suppose  $\Theta_A, \Theta_B$  exhibit stability reversal and  $\Theta_A$  is the correctly specified singleton model. Then, the beliefs that the adherents of  $\Theta_B$  hold in all EZs with p = (1,0) and the beliefs they hold in all EZs with p = (0,1) form disjoint sets. Also,  $\Theta_B$  is not strategically independent.

The first claim of Proposition 2 underscores that stability reversal requires inference—it cannot happen if group B agents merely have a different subjective preference. The second claim shows that stability reversal can only happen if the misspecified agents respond differently to different rival play, immediately implying they cannot emerge in decision problems.

We now show by example that stability reversal can emerge with models that allow for inference. Consider a two-player investment game where player *i* chooses an investment level  $a_i \in \{1, 2\}$ . A random productivity level *P* is realized according to  $b^{\bullet}(a_i + a_{-i}) + \epsilon$  where  $\epsilon$  is a zero-mean noise term,  $b^{\bullet} > 0$ . Player *i*'s payoffs are  $a_i \cdot P - 1_{\{a_i=2\}} \cdot c$ . Consequences are  $y = (a_i, a_{-i}, P)$ . We record the payoff matrix of this investment game:

	1	2
1	$2b^{\bullet}, 2b^{\bullet}$	$3b^{\bullet}, 6b^{\bullet} - c$
2	$6b^{\bullet} - c, 3b^{\bullet}$	$8b^{\bullet} - c, 8b^{\bullet} - c$

#### Condition 1. $5b^{\bullet} < c < 6b^{\bullet}$ .

In words, we assume that  $a_i = 1$  is a strictly dominant strategy in the stage game, but the investment profile (2,2) Pareto dominates the investment profile (1,1) (so that the corresponding game is a Prisoner's Dilemma). Consider two models in the society. Take  $\Theta_A$ to be a correctly specified singleton (thus knowing the true mapping from actions to payoffs), while  $\Theta_B$  wrongly stipulates  $P = b(a_i + a_{-i}) - m + \epsilon$ , where m > 0 is fixed, while  $b \in \mathbb{R}$  is a parameter that the adherents infer. We impose a condition on  $\Theta_B$ , which holds whenever m > 0 is large enough:

Condition 2.  $c < 4b^{\bullet} + \frac{1}{3}m$  and  $c < 5b^{\bullet} + \frac{1}{4}m$ .

We show that in this example models  $\Theta_A$  and  $\Theta_B$  exhibit stability reversal.

**Example 2.** In the investment game, under Condition 1 and Condition 2,  $\Theta_A$  and  $\Theta_B$  exhibit stability reversal.

The idea is that the adherents of  $\Theta_B$  are polymorphic. They overestimate the complementarity of investments, and this overestimation is more severe when they face data generated from lower investment profiles. As a result, the match between  $\Theta_A$  and  $\Theta_B$  plays out differently depending on which model is resident: it results in the investment profile (1, 2) when  $\Theta_A$  is resident, but results in (1, 1) when  $\Theta_B$  is resident. (We relegate the formal argument to Appendix A.5.) Due to Propositions 1 and 2, we conclude that this example is possible due to the non-trivial strategic interactions and  $\Theta_B$ 's inference about b (polymorphism through the learning channel).

Stability reversals provide a clear demonstration of polymorphism in models that permit inference. A mutant model may appear weak when present in small proportions, doing worse than the incumbent model conditional on every type of opponent. Yet, if the population share of the mutant model reaches a critical mass, its adherents infer a more evolutionarily advantageous model parameter based on their within-group interactions, change their bestresponse correspondence, and hence outperform the adherents of the incumbent model.

## 3.3 Non-Monotonic Stability in Matching Assortativity

Our last general result is also a consequence of the polymorphism of misspecified learners: a mutant model might successfully invade *only* when matching assortativity in the society is

intermediate. This non-monotonicity arises because a misspecified agent can draw different inferences about the game's fundamentals depending on the relative frequency of in-group and out-group interactions, as these two groups of opponents choose different actions. The idea that social interaction structure shapes people's beliefs about the world has empirical support,<sup>9</sup> and our framework shows how this mechanism affects the stability of misspecifications.

We again assume there is only one situation, for simplicity. Note that without inference (i.e., in the setting of preference evolution), the fitness of a group is linear in matching assortativity. Thus, for singleton models,  $\Theta_A$  being evolutionarily stable against  $\Theta_B$  both when  $\lambda = 0$  and when  $\lambda = 1$  implies the same holds for all  $\lambda \in (0, 1)$ .

**Proposition 3.** Suppose  $\Theta_A, \Theta_B$  are singleton models (i.e., subjective preferences in the stage game) and  $\Theta_A$  is evolutionarily stable against  $\Theta_B$  with  $\lambda$ -matching for both  $\lambda = 0$  and  $\lambda = 1$ . Then,  $\Theta_A$  is also evolutionarily stable against  $\Theta_B$  with  $\lambda$ -matching for any  $\lambda \in [0, 1]$ .

Crucially, inference leads to cases where the relevant "preference" changes depending on how frequently a model interacts with different types of opponents. This kind of polymorphism means a model's fitness may be *non-linear* in the matching probabilities. This phenomenon is a distinguishing feature of our framework and we show that the conclusion of Proposition 3 need not hold for models that allow for parameter inferences.

Consider a stage game where each player chooses an action from  $\{a_1, a_2, a_3\}$ . Every player then receives a random prize,  $y \in \{g, b\}$ , with utility values  $\pi(g) = 1$  and  $\pi(b) = 0$ . The payoff matrix below displays the objective expected utilities associated with different action profiles, which also correspond to the probabilities that the row and column players receive the good prize g.

	$a_1$	$a_2$	$a_3$
$a_1$	0.25,  0.25	0.50, 0.20	0.70,  0.15
$a_2$	0.20,  0.50	0.40, 0.40	0.40,  0.20
$a_3$	0.15,0.70	0.20,  0.40	0.20,  0.20

Let  $\Theta_A$  be the correctly specified singleton model. The action  $a_1$  is strictly dominant under the objective payoffs, so an adherent of  $\Theta_A$  always plays  $a_1$  in all matches. Let  $\Theta_B$ 

<sup>&</sup>lt;sup>9</sup>For example, Bazzi et al. (2019) document how ethnic attachment in response to a resettlement policy in Indonesia has varying effects depending on whether a community is "fractionalized" (so that most interactions are not with one's own group members, i.e.,  $\lambda$  is small) versus polarized (so that most interactions are with one's own group, i.e.,  $\lambda$  is large).

be a misspecified model  $\Theta_B = \{F_H, F_L\}$ . Each model  $F_H, F_L$  stipulates that the prize g is generated according to the probabilities in the following table, where b and c are parameters that depend on the model. The model  $F_H$  has (b, c) = (0.8, 0.2) and  $F_L$  has (b, c) = (0.1, 0.4).

	$a_1$	$a_2$	$a_3$
$a_1$	0.10, 0.10	0.10, c	0.10, 0.15
$a_2$	c, 0.10	b, b	b, 0.20
$a_3$	0.15,0.10	0.20, b	0.20,  0.20

The learning channel for the biased mutants leads the correctly specified model to have non-monotonic evolutionary stability in terms of matching assortativity.

**Example 3.** In this stage game,  $\Theta_A$  is evolutionarily stable against  $\Theta_B$  under  $\lambda$ -matching when  $\lambda = 0$  and  $\lambda = 1$ , but it is also evolutionarily fragile under  $\lambda$ -matching when  $\lambda \in (\lambda_l, \lambda_h)$ , where  $0 < \lambda_l < \lambda_h < 1$  are  $\lambda_l = 0.25$ ,  $\lambda_h \approx 0.56$ .

Consider the match between two adherents of  $\Theta_B$ . If they believe in  $F_H$ , they will play the action profile  $(a_2, a_2)$  and payoff profile (0.4, 0.4), a Pareto improvement compared to the correctly specified outcome  $(a_1, a_1)$ . The problem is that the data from play of  $(a_2, a_2)$  fit  $F_L$  better than  $F_H$ , since the objective 40% probability of getting prize g is closer to  $F_L$ 's conjecture (10%) than  $F_H$ 's conjecture (80%). A belief in  $F_H$  — and hence the profile  $(a_2, a_2)$ — cannot be sustained if the mutants only play each other. On the other hand, when an adherent of  $\Theta_B$  plays a correctly specified  $\Theta_A$  adherent, both  $F_H$  and  $F_L$  prescribe a best response of  $a_2$  against the  $\Theta_A$  adherent's play  $a_1$ . The data generated from the  $(a_2, a_1)$  profile lead biased agents to the parameter  $F_H$  that enables cooperative behavior within the mutant community. But, these matches against correctly specified opponents harm the mutant's welfare, as they only get an objective payoff of 0.2.

Therefore, the most advantageous interaction structure for the mutants is one where they can infer  $F_H$  using the data from matches against correctly specified opponents, then extrapolate this optimistic belief about b to coordinate on  $(a_2, a_2)$  in matches against fellow mutants. This requires the mutants to match with intermediate assortativity. Figure 1 depicts the equilibrium fitness of the mutant model  $\Theta_B$  as a function of assortativity in a society where the mutant model's population share is close to zero. While payoffs of  $\Theta_B$  adherents increase in  $\lambda$  at first, eventually they drop when mutant-vs-mutant matches become sufficiently frequent that a belief in  $F_H$  can no longer be sustained. Note that a



Figure 1: The EZ fitness of  $\Theta_B$  for different values of  $\lambda$  when  $p_B = 0$ . (The EZ fitness of the resident model  $\Theta_A$  is always 0.25.) In the blue region, adherents of  $\Theta_B$  infer  $F_H$  and receive linearly increasing average payoffs across all matches as  $\lambda$  increases. In the red region, adherents of  $\Theta_B$  infer  $F_L$  and receive a payoff of 0.2 in all matches, regardless of  $\lambda$ .

similar conclusion holds with fixed  $\lambda$  and varying population sizes: what ultimately matters is the probability with which  $\Theta_B$  interacts with each model. Non-linearity of fitness in the population shares can emerge here as well, also a unique possibility due to inference.<sup>10</sup>

# 4 Evolutionary Stability of Analogy Classes

We now turn to an illustration of the practical value of our approach. Specifically, we apply the stability notions introduced in this paper to study coarse thinking in games. Jehiel (2005) introduced analogy-based expectation equilibrium (ABEE) in extensive-form games, where agents group opponents' nodes into *analogy classes* and only keep track of aggregate statistics of opponents' average behavior within each analogy class. An ABEE is a strategy profile where agents best respond to the belief that at all nodes in every analogy class, opponents behave according to the average behavior in the analogy class. The ensuing literature typically treats analogy classes as exogenously given, interpreted as arising from coarse feedback or agents' cognitive limitations.<sup>11</sup> We showcase the practical value of our approach by using the framework from Section 2 to endogenize analogy classes based on their objective expected

 $<sup>^{10}</sup>$ See Appendix 4.2 for a discussion of stability with intermediate population shares.

<sup>&</sup>lt;sup>11</sup>Section 6.2 of Jehiel (2005) mentions that if players could choose their own analogy classes, then the finest analogy classes need not arise, but also says "it is beyond the scope of this paper to analyze the implications of this approach." In a different class of games, Jehiel (1995) similarly observes that another form of bounded rationality (having a limited forecast horizon about opponent's play) can improve welfare.

payoffs in equilibrium.<sup>12</sup>

# 4.1 Relaxing the Observability of Strategies

To study analogy-based reasoning, we relax the assumption that people correctly know others' strategies in equilibrium. We introduce the concepts of extended parameters and extended models:

**Definition 7.** An extended parameter is a triplet  $(a_A, a_B, F)$  with  $a_A, a_B \in \mathbb{A}$  and  $F : \mathbb{A}^2 \to \Delta(\mathbb{Y})$ . An extended model  $\overline{\Theta}$  is a collection of extended parameters: i.e., a subset of  $\mathbb{A}^2 \times (\Delta(\mathbb{Y}))^{\mathbb{A}^2}$ .

In addition to a conjecture F about how strategy profiles translate into consequences for the agent, extended models also contain conjectures about how group A and group B opponents will act. We assume the marginal of the extended model on  $(\Delta(\mathbb{Y}))^{\mathbb{A}^2}$  is metrizable. As before, we also assume each F is given by a density or probability mass function  $f(a_i, a_{-i}) : \mathbb{Y} \to \mathbb{R}_+$  for every  $(a_i, a_{-i}) \in \mathbb{A}^2$ . We say that an extended model  $\overline{\Theta}$  is *correctly specified* if  $\overline{\Theta} = \mathbb{A}^2 \times \{F^{\bullet}(\cdot, \cdot, G)\}$ , so the agent can make unrestricted inferences about others' play and does not rule out the correct data-generating process  $F^{\bullet}(\cdot, \cdot, G)$  for any situation G.

Defining zeitgeists for extended models is immediate, as we can simply replace "model" with "extended model" in Definition 1. The equilibrium notion, however, is subtly different:

**Definition 8.** A zeitgeist with strategic uncertainty  $\overline{\mathfrak{Z}} = (\overline{\Theta}_A, \overline{\Theta}_B, \mu_A(G), \mu_B(G), p, \lambda, a(G))_{G \in \mathcal{G}}$ is an *equilibrium zeitgeist with strategic uncertainty (EZ-SU)* if for every  $G \in \mathcal{G}$  and  $g, g' \in \{A, B\}, a_{g,g'}(G) \in \underset{\hat{a} \in \mathbb{A}}{\operatorname{arg max}} \mathbb{E}_{(a_A, a_B, F) \sim \mu_g(G)} \left[ \mathbb{E}_{y \sim F(\hat{a}, a_{g'})}(\pi(y)) \right]$  and, for every  $g \in \{A, B\}$ , the belief  $\mu_g(G)$  is supported on

$$\arg \min_{(\hat{a}_{A},\hat{a}_{B},\hat{F})\in\overline{\Theta}_{g}} \left\{ \begin{array}{c} (\lambda + (1-\lambda)p_{g}) \cdot D_{KL}(F^{\bullet}(a_{g,g}(G), a_{g,g}(G), G) \parallel \hat{F}(a_{g,g}(G), \hat{a}_{g}))) \\ + (1-\lambda)(1-p_{g}) \cdot D_{KL}(F^{\bullet}(a_{g,-g}(G), a_{-g,g}(G), G) \parallel \hat{F}(a_{g,-g}(G), \hat{a}_{-g}) \end{array} \right\}$$

where -g means the group other than g.

 $<sup>^{12}</sup>$ Other approaches to endogenizing analogy classes are pursued in Jehiel and Mohlin (2023); Jehiel and Weber (2023).

The difference compared to Definition 2 is that the KL divergence is now taken with respect to the *conjectured opponent's strategy*, part of the extended model. Conjectures now include others' play, in addition to stage game parameters.

# 4.2 Defining Stable Population Shares

In this section, we will also be interested in stable population shares in a society that contains positive fractions of both rational and misspecified players. We briefly introduce the following solution concept.

**Definition 9.** Call population share (p, 1 - p) with  $p \in (0, 1)$  a stable population share if there is an EZ (or EZ-SU) with (p, 1 - p) where both models have the same fitness, and there exists  $\bar{\epsilon}$  such that:

- 1. For any  $0 < \epsilon < \overline{\epsilon}$ , there is an EZ (or EZ-SU) with population share  $(p + \epsilon, 1 p \epsilon)$ where  $\Theta_A$  has strictly lower fitness than  $\Theta_B$
- 2. For any  $0 < \epsilon < \overline{\epsilon}$ , there is an EZ (or EZ-SU) with population share  $(p \epsilon, 1 p + \epsilon)$ where  $\Theta_A$  has strictly higher fitness than  $\Theta_B$ .

Whereas Definition 3's stability notion involves comparing the performance of the two models when one of them is present in an arbitrarily small fraction, stability with an interior population share as in Definition 9 refers to both models co-existing with equal fitness in a way that is robust to local perturbations of population sizes.

# 4.3 Centipede Games and Analogy-Based Reasoning

We now analyze analogy-based reasoning in the centipede game in Figure 2 (there is only one situation, given by the payoffs in this game). P1 and P2 take turns choosing Across (A) or Drop (D). The non-terminal nodes are labeled  $n^k$ ,  $1 \le k \le K$  where K is an even number. P1 acts at odd nodes and P2 acts at even nodes, where choosing Drop at  $n^k$  leads to the terminal node  $z^k$ . If Across is always chosen, then the terminal node  $z^{end}$  is reached. Every time a player *i* chooses Across, the sum of payoffs grows by g > 0, but if the opponent chooses Drop next, *i*'s payoff is  $\ell > 0$  smaller than *i*'s payoff had they chosen Drop, with  $\ell > g$ . Thus, if  $z^{end}$  is reached, both get Kg/2; if  $z^k$  is reached when k is odd, both players obtain  $\frac{g(k-1)}{2}$ ; and if if  $z^k$  is reached when k is even, P1 obtains  $\frac{k-2}{2}g - \ell$ , and P2 obtains  $\frac{k}{2}g + \ell$ .



(0,0) (-l,g+l) (g,g) g(K-2)/2 gK/2+l) Figure 2: The centipede game. P1 (blue) and P2 (red) alternate in choosing Across (A) or Drop (D). Payoff profiles are shown at the terminal nodes.

While this is an asymmetric stage game, we study a symmetrized version where two matched agents are randomly assigned into the roles of P1 and P2. Let  $\mathbb{A} = \{(d^k)_{k=1}^K \in [0, 1]^K\}$ , so each strategy is characterized by the probabilities of playing Drop at various nodes in the game tree. When assigned into the role of P1, the strategy  $(d^k)$  plays Drop with probabilities  $d^1, d^3, ..., d^{K-1}$  at nodes  $n^1, n^3, ...n^{K-1}$ . When assigned into the role of P2, it plays Drop with probabilities  $d^2, d^4, ..., d^K$  at nodes  $n^2, n^4, ...n^K$ . The set of consequences is  $\mathbb{Y} = \{1, 2\} \times (\{z_k : 1 \leq k \leq K\} \cup \{z_{end}\})$ , where the first dimension of the consequence returns the player role that the agent was assigned into, and the second dimension returns the terminal node reached. Let  $F^{\bullet} : \mathbb{A}^2 \to \Delta(\mathbb{Y})$  be the objective distribution over consequences.

All agents know the game tree (i.e.,  $F^{\bullet}$ ), but some might adhere to a model which mistakenly assumes that their opponent plays Drop with the same probabilities at all of their nodes. Formally, define the restricted space of strategies  $\mathbb{A}^{An} := \{(d^k) \in [0,1]^K : d^k = d^{k'} \text{ if } k \equiv k' \pmod{2}\} \subseteq \mathbb{A}$ . The correctly specified extended model is  $\overline{\Theta}^{\bullet} := \mathbb{A} \times \mathbb{A} \times \{F^{\bullet}\}$ . The misspecified model of interest is  $\overline{\Theta}^{An} := \mathbb{A}^{An} \times \mathbb{A}^{An} \times \{F^{\bullet}\}$ , reflecting a dogmatic belief that opponents play the same mixed action at all nodes in the analogy class. We emphasize these restriction on strategies only exists in the subjective beliefs of the model  $\overline{\Theta}^{An}$  adherents. All agents, regardless of their model, actually have the strategy space  $\mathbb{A}$ .

#### 4.4 Results

The next proposition provides a justification for why we might expect agents with coarse analogy classes given by  $\mathbb{A}^{An}$  to persist in the society.

**Proposition 4.** Suppose  $K \ge 4$  and  $g > \frac{2}{K-2}\ell$ . For any matching assortativity  $\lambda \in [0,1]$ , the correctly specified extended model  $\overline{\Theta}^{\bullet}$  is evolutionarily stable with strategic uncertainty against itself, but it is not evolutionarily stable with strategic uncertainty against the misspecified extended model  $\overline{\Theta}^{An}$ . Also,  $\overline{\Theta}^{An}$  is not evolutionarily stable against  $\overline{\Theta}^{\bullet}$ , unless  $\lambda = 1$ .

Thus, the correctly specified extended model is not evolutionarily stable against a coarse

reasoner for any level of assortativity. Here, the conditional fitness of  $\overline{\Theta}^{An}$  against both  $\overline{\Theta}^{\bullet}$ and  $\overline{\Theta}^{An}$  can strictly improve on the correctly specified residents' equilibrium fitness. This is because the matches between two adherents of  $\overline{\Theta}^{\bullet}$  must result in Dropping at the first move in equilibrium, while matches where at least one player is an adherent of  $\overline{\Theta}^{An}$  either lead to the same outcome or lead to a Pareto dominating payoff profile as the misspecified agent misperceives the opponent's continuation probability and thus chooses Across at almost all of the decision nodes.

However,  $\overline{\Theta}^{An}$  is not evolutionarily stable against  $\overline{\Theta}^{\bullet}$  either. The correctly specified agents can exploit the analogy reasoners' mistake and receive higher payoffs in matches against them than the misspecified agents receive in matches against each other. Hence, no homogeneous population can be stable, as the resident model would have lower fitness than the mutant model in equilibrium. Thus we determine stable shares as defined in Section 4.2, focusing on the EZ-SU where Across is played as often as possible.

We take  $\lambda = 0$  throughout the remainder of this section. Suppose  $K \ge 4$  and  $g > \frac{2}{K-2}\ell$ . Consider the maximal continuation EZ-SU: (1) misspecified agents always play Across except at node K where they choose Drop, and (2) correctly specified agents (i) matched with misspecified agents play Drop at nodes K - 1 and K and Across otherwise, and (ii) matched with correctly specified agents always play Drop. We verify this indeed forms an EZ-SU.

**Proposition 5.** Suppose  $\lambda = 0$ ,  $K \ge 4$  and  $g > \frac{2}{K-2}\ell$ . The only stable population share  $(p_A^*, p_B^*)$  supported by the maximal continuation EZ-SU described above is  $p_B^* = 1 - \frac{\ell}{g(K-2)}$ . We have  $p_B^*$  is strictly increasing in g and K, and strictly decreasing in  $\ell$ .

Intuitively,  $p_B^*$  reflects the fraction of society expected to be analogy reasoners if long run population changes are determined by fitness. Under the maintained assumption  $g > \frac{2}{K-2}\ell$ , the stable population share of misspecified agents is strictly more than 50%, and the share grows with more periods and a larger increase in payoffs from continuation. The main intuition is that the misspecified model has a higher conditional fitness than the rational model against rational opponents. The former leads to many periods of continuation and a high payoff for the biased agent when the rational agent eventually drops, but the latter leads to 0 payoff from immediate dropping. On the other hand, the misspecified model has a lower conditional fitness than the rational model against misspecified opponents. For the two groups to have the same expected fitness, there must be fewer rational opponents (i.e., a smaller stable population share  $p_A^*$ ) when g and K are higher. Note that, when payoffs are specified as above, two successive periods of continuation lead to a strict Pareto improvement in payoffs. Consider instead the dollar game (Reny, 1993) in Figure 3, a variant with a more "competitive" payoff structure, where an agent always gets zero when the opponent plays Drop, at all parts of the game tree. Assume total payoff increases by 1 in each round. If the first player stops immediately, payoffs are (1, 0), and if the second player continues at the final node  $n^{K}$ , payoffs are (K + 2, 0).



Figure 3: The dollar game. Players 1 (blue) and 2 (red) alternate in choosing Across (A) or Drop (D). Payoff profiles are shown at the terminal nodes.

**Proposition 6.** For  $\lambda = 0$  and every population size (p, 1 - p) with  $p \in [0, 1]$ , the maximal continuation EZ-SU is an EZ-SU where the fitness of  $\overline{\Theta}^{\bullet}$  is strictly higher than that of  $\overline{\Theta}^{An}$ .

While maximal continuation remains an EZ-SU, the rational model strictly outperforms the misspecified model for all population shares. Provided the maximal continuation EZ-SU remains focal, we would expect no analogy reasoners in the long run with this stage game. Intuitively, the payoffs imply one player can only do better *at the expense* of the opponent. Since  $\lambda = 0$ , this implies the less cooperative strategy will be selected.

In a recent survey, Jehiel (2020) points out that the misspecified Bayesian learning approach to analogy classes should aim for "a better understanding of how the subjective theories considered by the players may be shaped by the objective characteristics of the environment."<sup>13</sup> Taken together, our analysis in this section provides predictions regarding when coarse reasoning should be more prevalent, specifically when the payoff structure is "less competitive." When this is indeed the case, the bias become more prevalent with a longer horizon and with faster payoff growth.

<sup>&</sup>lt;sup>13</sup>Jehiel (2020) interprets ABEEs as players adopting the "simplest" explanations of observed aggregate statistics of play with coarse feedback. An objectively coarse feedback structure can lead agents to adopt the subjective belief that others behave in the same way in all contingencies in the same coarse analogy class.

# 5 Related Literature

Our paper contributes to the literature on misspecified Bayesian learning by proposing a framework to assess which specifications are more likely to persist based on their objective performance. Our two main contributions are to highlight that misinference in such a framework allows for (1) tailored commitments and (2) polymorphism. Most prior work on misspecified Bayesian learning takes the misspecification as exogenous, studying the subsequent implications in both single-agent decision problems<sup>14</sup> and multi-agent games.<sup>15</sup> A number of papers establish general convergence properties of misspecified learning.<sup>16</sup> Our approach to endogenizing misspecified inference contrasts with those involving subjective expectations of payoffs<sup>17</sup> or goodness-of-fit tests.<sup>18</sup> To our knowledge, past work that *has* used objective payoffs to endogenize misspecified inference has restricted attention to financial markets (Sandroni, 2000; Massari, 2020).

This paper is closest to two independent and contemporaneous papers, Fudenberg and Lanzani (2022) and Frick, Iijima, and Ishii (2024), who consider welfare-based criteria for selecting among misspecifications in single-agent decision problems.<sup>19</sup> We differ in highlighting that the learning channel can *strictly* expand the possibility for misspecifications to invade rational societies in strategic settings (relative to biased invaders who do not draw inferences), and we show that misspecifications can lead to different best responses in different environments and thus induce new stability phenomena.

Our framework of competition between different specifications for Bayesian learning is inspired by the evolutionary game theory literature. Relative to this literature, our contribution is to accommodate misspecified inference. We follow past work that also uses

<sup>&</sup>lt;sup>14</sup>See Nyarko (1991); Fudenberg, Romanyuk, and Strack (2017); Heidhues, Koszegi, and Strack (2018); He (2022). Also related is Fudenberg et al. (2024) who show how memory limitations yield inferences resembling those in misspecified learning models.

<sup>&</sup>lt;sup>15</sup>See Bohren (2016); Bohren and Hauser (2021); Jehiel (2018); Molavi (2019); Dasaratha and He (2020); Ba and Gindin (2022); Frick, Iijima, and Ishii (2020); Murooka and Yamamoto (2023).

<sup>&</sup>lt;sup>16</sup>See Esponda and Pouzo (2016); Esponda, Pouzo, and Yamamoto (2021); Frick, Iijima, and Ishii (2022); Fudenberg, Lanzani, and Strack (2021).

<sup>&</sup>lt;sup>17</sup>See Olea, Ortoleva, Pai, and Prat (2022); Levy, Razin, and Young (2022); Gagnon-Bartsch, Rabin, and Schwartzstein (2021)

<sup>&</sup>lt;sup>18</sup>See Cho and Kasa (2015, 2017); Ba (2022); Schwartzstein and Sunderam (2021); Lanzani (2022).

<sup>&</sup>lt;sup>19</sup>Fudenberg and Lanzani (2022) study a framework where a continuum of agents with heterogeneous misspecifications arrive each period and learn from their predecessors' data. Frick, Iijima, and Ishii (2024) assign a *learning efficiency index* to every misspecified signal structure and conduct a robust comparison of welfare under different misspecifications.

objective payoffs as the selection criterion for subjective preferences in games and decision problems (e.g., Dekel, Ely, and Yilankaya (2007), see also the surveys Robson and Samuelson (2011) and Alger and Weibull (2019)) and the evolution of constrained strategy spaces (Heller, 2015; Heller and Winter, 2016). Like us, Güth and Napel (2006) allow for stage-game heterogeneity, studying the ability to discriminate between these games.

When agents entertain fundamental uncertainty about payoff parameters, our framework applies evolutionary forces to *sets of* preferences (i.e., models with multiple possible parameter values). This allows us to ask our central question: When does the ability to draw inference expand the scope for errors to invade rational societies? Developing a framework that accommodates inference is necessary to answer this question, providing the main point of departure from the literature on the indirect evolutionary approach. Our emphasis on *Bayesian* learning also distinguishes our work from papers that study the evolution of different belief-formation processes (Heller and Winter, 2020; Berman and Heller, 2022), who take a reduced-form (and possibly non-Bayesian) approach and consider arbitrary inference rules.

# 6 Concluding Discussion

We have introduced an evolutionary approach to predict the persistence of misspecified Bayesian learning. We have emphasized the implications and significance of the learning channel for evolutionary stability and the viability of biases. Our contributions are twofold. First, we show that the learning channel may confer strategic benefits in cases where dogmatic beliefs do not. This is because the learning channel enables flexible commitments that are tailored to the realized situation. Second, we show that misspecified agents are polymorphic. For this reason, the performance of a fixed bias may be difficult to extrapolate across environments. More broadly, we hope to have shown that incorporating inference enables the evolutionary approach to speak to new applications and patterns.

The idea that agents' personal experiences (and more broadly, the environments that generate these experiences) shape their preferences *beyond* their individual characteristics is empirically well documented. For instance, recent work studying attitudes toward immigrants (Bursztyn et al. (2022)) or attitudes among immigrants (Bolotnyy et al. (2022)) find that variation in a person's environment—plausibly independent from individual characteristics—can considerably influence their political behavior and preferences. In an experiment with Indian men, Lowe (2021) finds that favoritism for one's own caste changes in response to

cross-caste contacts, in a way that depends on whether interactions are competitive or cooperative. Our framework derives the implications of these kinds of preference-formation mechanisms on the stability of misspecified models.

We acknowledge that our framework does not account for which errors appear in the first place. It is plausible that some first-stage filter prevents certain obvious misspecifications from ever reaching the stage that we study in the evolutionary framework. For this reason, the applications we focused on reflected misspecifications that seem psychologically plausible.

We have used an otherwise off-the-shelf framework to describe the selection of specifications. The goal of this paper is not to identify suitable definitions of fitness to justify particular errors (which is the focus for many of the papers that Robson and Samuelson (2011) survey). Rather, our goal has been to determine what evolutionary forces would suggest about the emergence of misspecified learning, and implications thereof. We have therefore focused more on the implications of the learning channel in an otherwise standard evolutionary setup.

# References

- ALGER, I. AND J. WEIBULL (2013): "Homo Moralis-Preference Evolution Under Incomplete Information and Assortative Matching," *Econometrica*, 81, 2269–2302.
- (2019): "Evolutionary models of preference formation," Annual Review of Economics, 11, 329–354.
- ALIPRANTIS, C. AND K. BORDER (2006): Infinite Dimensional Analysis: A Hitchhiker's Guide, Springer Science & Business Media.
- BA, C. (2022): "Robust model misspecification and paradigm shift," Working Paper.
- BA, C. AND A. GINDIN (2022): "A multi-agent model of misspecified learning with overconfidence," Working Paper.
- BAZZI, S., A. GADUH, A. D. ROTHENBERG, AND M. WONG (2019): "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *American Economic Review*, 109, 3978–4025.
- BERMAN, R. AND Y. HELLER (2022): "Naive Analytics: The Strategic Advantage of Algorithmic Heuristics," *Working Paper*.
- BOHREN, J. A. (2016): "Informational herding with model misspecification," *Journal of Economic Theory*, 163, 222–247.
- BOHREN, J. A. AND D. HAUSER (2021): "Learning with heterogeneous misspecified models: Characterization and Robustness," *Econometrica*, 89, 3025–3077.

- BOLOTNYY, V., M. KOMISARCHIK, AND B. LIBGOBER (2022): "How Does Childhood Environment Shape Political Participation? Evidence from Refugees," *Working Paper*.
- BURSZTYN, L., T. CHANEY, T. HASSAN, AND A. RAO (2022): "The Immigrant Next Door: Long-Term Contact, Generosity, and Prejudice," *Working Paper*.
- CHO, I.-K. AND K. KASA (2015): "Learning and model validation," *Review of Economic Studies*, 82, 45–82.

——— (2017): "Gresham's law of model averaging," American Economic Review, 107, 3589–3616.

- DASARATHA, K. AND K. HE (2020): "Network structure and naive sequential learning," *Theoretical Economics*, 15, 415–444.
- DEKEL, E., J. ELY, AND O. YILANKAYA (2007): "Evolution of preferences," *Review of Economic Studies*, 74, 685–704.
- ESPONDA, I. AND D. POUZO (2016): "Berk–Nash equilibrium: A framework for modeling agents with misspecified models," *Econometrica*, 84, 1093–1130.

ESPONDA, I., D. POUZO, AND Y. YAMAMOTO (2021): "Asymptotic behavior of Bayesian learners with misspecified models," *Journal of Economic Theory*, 195, 105260.

- FRICK, M., R. IIJIMA, AND Y. ISHII (2020): "Misinterpreting others and the fragility of social learning," *Econometrica*, 88, 2281–2328.
  - —— (2022): "Belief convergence under misspecified learning: a martingale approach," *Review of Economic Studies, forthcoming.*

— (2024): "Welfare comparisons for biased learning," American Economic Review, 1612–1649.

- FUDENBERG, D. AND G. LANZANI (2022): "Which misperceptions persist?" Theoretical Economics, forthcoming.
- FUDENBERG, D., G. LANZANI, AND P. STRACK (2021): "Limit Points of Endogenous Misspecified Learning," *Econometrica*, 89, 1065–1098.

— (2024): "Selective Memory Equilibrium," Journal of Political Economy, forthcoming.

- FUDENBERG, D., G. ROMANYUK, AND P. STRACK (2017): "Active learning with a misspecified prior," *Theoretical Economics*, 12, 1155–1189.
- GAGNON-BARTSCH, T., M. RABIN, AND J. SCHWARTZSTEIN (2021): "Channeled attention and stable errors," *Working Paper*.
- GÜTH, W. AND S. NAPEL (2006): "Inequality Aversion in a Variety of Games An Indirect Evolutionary Analysis," *Economic Journal*, 116, 1037–1056.

- HAMILTON, W. D. (1964a): "The Genetical Evolution of Social Behaviour. 2," Journal of Theoretical Biology, 7, 17–52.
- (1964b): "The Genetical Evolution of Social Behaviour. I," *Journal of Theoretical Biology*, 7, 1–16.
- HE, K. (2022): "Mislearning from censored data: The gambler's fallacy in optimal-stopping problems," *Theoretical Economics*, 17, 1269–1312.
- HE, K. AND J. LIBGOBER (2024): "Higher-Order Beliefs and (Mis)learning from Prices," Working Paper.
- HEIDHUES, P., B. KOSZEGI, AND P. STRACK (2018): "Unrealistic expectations and misguided learning," *Econometrica*, 86, 1159–1214.
- HELLER, Y. (2015): "Three steps ahead," Theoretical Economics, 10, 203–241.
- HELLER, Y. AND E. WINTER (2016): "Rule rationality," *International Economic Review*, 57, 997–1026.

— (2020): "Biased-belief equilibrium," American Economic Journal: Microeconomics, 12, 1–40.

- JEHIEL, P. (1995): "Limited horizon forecast in repeated alternate games," Journal of Economic Theory, 67, 497–519.
- (2005): "Analogy-based expectation equilibrium," Journal of Economic theory, 123, 81–104.
  - (2018): "Investment strategy and selection bias: An equilibrium perspective on overoptimism," American Economic Review, 108, 1582–97.
- (2020): "Analogy-based expectation equilibrium and related concepts: Theory, applications, and beyond," *Working Paper*.
- JEHIEL, P. AND E. MOHLIN (2023): "Categorization in Games: A Bias-Variance Perspective," Working paper.
- JEHIEL, P. AND R. WEBER (2023): "Calibrated Clustering and Analogy-Based Expectation Equilibrium," Working paper.
- LANZANI, G. (2022): "Dynamic Concern for Misspecification," Working Paper.
- LEVY, G., R. RAZIN, AND A. YOUNG (2022): "Misspecified politics and the recurrence of populism," *American Economic Review*, 112, 928–962.
- LOWE, M. (2021): "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *American Economic Review*, 111, 1807–44.
- MASSARI, F. (2020): "Under-reaction: Irrational behavior or robust response to model misspecification?" Working Paper.

- MOLAVI, P. (2019): "Macroeconomics with learning and misspecification: A general theory and applications," *Working Paper*.
- MUROOKA, T. AND Y. YAMAMOTO (2023): "Higher-Order Misspecification and Equilibrium Stability," Working Paper.
- NYARKO, Y. (1991): "Learning in mis-specified models and the possibility of cycles," *Journal of Economic Theory*, 55, 416–427.
- OLEA, J. L. M., P. ORTOLEVA, M. M. PAI, AND A. PRAT (2022): "Competing models," Quarterly Journal of Economics, forthcoming.
- RENY, P. J. (1993): "Common belief and the theory of games with perfect information," *Journal* of Economic Theory, 59, 257–274.
- ROBSON, A. J. (1996): "A biological basis for expected and non-expected utility," *Journal of Economic Theory*, 68, 397–424.
- ROBSON, A. J. AND L. SAMUELSON (2011): "The evolutionary foundations of preferences," in *Handbook of Social Economics*, Elsevier, vol. 1, 221–310.
- SANDHOLM, W. (2001): "Preference evolution, two-speed dynamics, and rapid social change," *Review of Economic Dynamics*, 4, 637–679.
- SANDRONI, A. (2000): "Do markets favor agents able to make accurate predictions?" *Econometrica*, 68, 1303–1341.
- SCHWARTZSTEIN, J. AND A. SUNDERAM (2021): "Using models to persuade," American Economic Review, 111, 276–323.

# Appendix

# A Proofs of Key Results from the Main Text

## A.1 Proof of Theorem 1

Part 1: Let  $\mathcal{V}$  be the convex hull of  $\{(v_G^b)_{G\in\mathcal{G}} \mid b : \mathbb{A} \Rightarrow \mathbb{A}\}$ , and let  $\mathcal{U} = \{(u_G)_{G\in\mathcal{G}} : u_G \leq v_G \text{ for all } G \text{ for some } v \in \mathcal{V}\}$ . Note  $\mathcal{U}$  is closed and convex (since  $\mathcal{V}$  is convex). By hypothesis,  $v^{\text{NE}}$  is not in the interior or on the boundary of  $\mathcal{U}$ . So by the separating hyperplane theorem, there exists a real number c and a vector  $q \in \mathbb{R}^{|\mathcal{G}|}$  with  $q_G \neq 0$  for every G, so that  $q \cdot v^{\text{NE}} > c > q \cdot u$  for every  $u \in \mathcal{U}$ . Furthermore,  $q_G \geq 0$  for every G. This is because if  $q_{G'} < 0$  for some G', then since  $\mathcal{U}$  contains vectors with arbitrarily negative values in the G' dimension, we cannot have  $q \cdot v^{\text{NE}} \geq q \cdot u$  for every  $u \in \mathcal{U}$ . We may then without loss view q as a distribution on  $\mathcal{G}$ . In fact, we can take q to be full support. To see this, note that since  $|\mathcal{G}| < \infty$  and  $\mathcal{U}$  is convex, we have

$$\lim_{\varepsilon \to 0} \max_{v \in \mathcal{U}} \left[ (1 - \varepsilon)q + \frac{\varepsilon}{|\mathcal{G}|} (1, 1, \dots, 1) \right] \cdot v = \max_{v \in \mathcal{U}} q \cdot v,$$

by continuity of the support function of convex sets in  $\mathbb{R}^n$  (given that the support function on  $\mathcal{U}$  is bounded for all  $q \geq 0$ , since  $v_G^b$  is bounded above for every b and every G). Thus, setting  $\tilde{q}(\varepsilon) = (1 - \varepsilon)q + \frac{\varepsilon}{|G|}(1, 1, ..., 1)$ , we have  $\tilde{q}(\varepsilon)$  is a full support distribution with  $\tilde{q}(\varepsilon) \cdot v^{NE} > c > \tilde{q}(\varepsilon) \cdot u$  whenever  $\varepsilon$  is sufficiently small, since we have that these inequalities hold in the limit.

Now consider any singleton model  $\Theta = \{F\}$ , and let  $b : \mathbb{A} \rightrightarrows \mathbb{A}$  be the subjective bestresponse correspondence that F induces. If  $v_G^b \neq -\infty$  for every G, then, for each G we can find a strategy profile  $(a_i^G, a_{-i}^G)$  where  $a_i^G \in b(a_{-i}^G)$ ,  $a_{-i}^G$  is a rational best response to  $a_i^G$  in situation G, and the strategy pair gives utility  $v_G^b$  to the first player. For any population shares of the two models, there is an EZ where the resident correctly specified agents get  $v_G^{\text{NE}}$  in situation G when playing against each other, and the mutants with model  $\Theta$  play  $(a_i, a_{-i})$  in matches against the residents and get utility  $v_G^b$  in the same situation. Under the distribution of situations q, as the fraction of the mutants approaches 0, the residents' fitness approaches  $q \cdot v^{\text{NE}}$  while that of the mutants approaches  $q \cdot v^b$ , and the former is strictly larger by construction of q since  $v^b \in \mathcal{U}$ . This EZ shows the correctly specified model is not evolutionarily fragile against  $\{F\}$ . Otherwise, if we have that  $v_G^b = -\infty$  for some G, then there are no EZs, so the correctly specified model is not evolutionarily fragile against  $\{F\}$  by the emptiness of the set of EZs.

Part 2: Suppose the hypotheses hold and let us construct the misspecified model  $\hat{\Theta} = \{F_G : G \in \mathcal{G}\}$ . To define the parameters  $F_G$ , first consider  $\tilde{F}_G$  where  $\tilde{F}_G(a_i, a_{-i}) := F^{\bullet}(a_i, \underline{BR}(a_i, G), G)$  for every  $a_{-i} \in \mathbb{A}$ . Now for each  $(a_i, a_{-i}, G) \in \mathbb{A} \times \mathbb{A} \times \mathcal{G}$ , define the fullsupport distribution  $F_G(a_i, a_{-i}) \in \Delta(\mathbb{Y})$  as a sufficiently small perturbation of the  $\tilde{F}_G(a_i, a_{-i})$ , such that for every  $a_i, a_{-i} \in \mathbb{A}$  and every  $G \in \mathcal{G}$ ,  $\min_{\hat{G} \in \mathcal{G}} KL(F^{\bullet}(a_i, a_{-i}, G) \parallel F_{\hat{G}}(a_i, a_{-i}))$  has a unique solution. This can be done because there are finitely many strategies and situations.

Consider any EZ 3 with the correctly specified resident,  $\Theta$  as the mutant,  $\lambda = 0$ . By situation identifiability, in  $\mathfrak{Z}$  the correctly specified residents must believe in the true  $F^{\bullet}(\cdot, \cdot, G)$ in every situation G. When the fraction of mutants  $\epsilon > 0$  is sufficiently small, the mutants cannot hold a mixed belief in any situation G, by the construction of the parameters in  $\hat{\Theta}$  to rule out ties in KL divergence. We show further that mutants must believe in  $F_G$  in situation G for  $\epsilon$  small enough. This is because if they instead believed in  $F_{G'}$  for some  $G' \neq G$ , then they must play  $\bar{a}_{G'}$  as the Stackelberg strategy is assumed to be unique. Let  $a_{-i}$  be the rational best response to  $\bar{a}_{G'}$  in situation G and  $a'_{-i}$  be the rational best response to  $\bar{a}_{G'}$  in situation G', both unique by assumption. In their matches against the residents, the mutants' expected distribution of consequences  $F_{G'}(\bar{a}_{G'}, a_{-i})$  is a perturbed version of  $F^{\bullet}(\bar{a}_{G'}, a'_{-i}, G')$ , while the true distribution of consequences  $F^{\bullet}(\bar{a}_{G'}, a_{-i}, G)$  is a perturbed version of  $F_G(\bar{a}_{G'}, a_{-i})$ . We have  $F^{\bullet}(\bar{a}_{G'}, a'_{-i}, G') \neq F^{\bullet}(\bar{a}_{G'}, a_{-i}, G)$  by Stackelberg identifiability, so  $KL(F^{\bullet}(\bar{a}_{G'}, a_{-i}, G) \parallel F_G(\bar{a}_{G'}, a_{-i})) < KL(F^{\bullet}(\bar{a}_{G'}, a_{-i}, G) \parallel F_{G'}(\bar{a}_{G'}, a_{-i}))$ when the perturbations are sufficiently small. When  $\epsilon > 0$  is small enough, this contradicts the mutants believing in  $F_{G'}$  in situation G as the parameter  $F_G$  generates smaller weighted KL divergence across all of the mutant's data (since data from matches against mutants get weighted by  $\epsilon$  and the full-support nature of all processes in the model implies that KL divergence of the data from such matches is bounded). So the mutants get the Stackelberg payoff in each situation when playing the resident, which means they have higher fitness than the residents in every EZ for  $\epsilon$  small enough, since  $\bar{v}_G > v_G^{\text{NE}}$  for at least one situation and q has full support. Finally, there exists at least one EZ: for  $\epsilon > 0$  small enough, it is an EZ for the residents to believe in  $F^{\bullet}(\cdot, \cdot, G)$  in every situation G, to play the symmetric Nash profile that results in  $v_G^{\text{NE}}$  when matched with other residents (this profile exists by hypothesis of the theorem), and for the mutants to believe in  $F_G$  and play  $(\bar{a}_G, \underline{BR}(\bar{a}_G, G))$  in matches against residents in situation G.

#### A.2 Proof of Proposition 1

Proof. Let two singleton models  $\Theta_A$ ,  $\Theta_B$  be given. By contradiction, suppose they exhibit stability reversal. Let  $\mathfrak{Z} = (\mu_A, \mu_B, p = (0, 1), \lambda = 0, (a))$  be any EZ where  $\Theta_B$  is resident. By the definition of EZ,  $\mathfrak{Z}' = (\mu_A, \mu_B, p = (1, 0), \lambda = 0, (a))$  is also an EZ where  $\Theta_A$  is resident. Let  $u_{g,g'}$  be model  $\Theta_g$ 's conditional fitness against group g' in the EZ  $\mathfrak{Z}'$ . Part (i) of the definition of stability reversal requires that  $u_{AA} > u_{BA}$  and  $u_{AB} > u_{BB}$ . These conditional fitness levels remain the same in  $\mathfrak{Z}$ . This means the fitness of  $\Theta_A$  is strictly higher than that of  $\Theta_B$  in  $\mathfrak{Z}$ , a contradiction.  $\Box$ 

### A.3 Proof of Proposition 2

Proof. To show the first claim, suppose  $\mathfrak{Z} = (\mu_A, \mu_B, p = (1, 0), \lambda = 0, (a_{AA}, a_{AB}, a_{BA}, a_{BB}))$ is an EZ, and  $\tilde{\mathfrak{Z}} = (\mu_A, \mu_B, p = (0, 1), \lambda = 0, (\tilde{a}_{AA}, \tilde{a}_{AB}, \tilde{a}_{BA}, \tilde{a}_{BB}))$  is another EZ where the adherents of  $\Theta_B$  hold the same belief  $\mu_B$  (group A's belief cannot change as  $\Theta_A$  is the correctly specified singleton model). By the optimality of behavior in  $\mathfrak{Z}$ ,  $a_{BA}$  best responds to  $a_{AB}$  under the belief  $\mu_B$ , and  $a_{AB}$  best responds to  $a_{BA}$  under the belief  $\mu_A$ , therefore  $\tilde{\mathfrak{Z}}' = (\mu_A, \mu_B, p = (0, 1), \lambda = 0, (\tilde{a}_{AA}, a_{AB}, a_{BA}, \tilde{a}_{BB}))$  is another EZ. This holds because the distributions of observations for the adherents of  $\Theta_B$  are identical in  $\tilde{\mathfrak{Z}}$  and  $\tilde{\mathfrak{Z}}'$ , since they only face data generated from the profile  $(\tilde{a}_{BB}, \tilde{a}_{BB})$ . At the same time, since  $\tilde{a}_{BB}$  best responds to itself under the belief  $\mu_B$ , we have that  $\mathfrak{Z}' = (\mu_A, \mu_B, p = (1, 0), \lambda = 0, (a_{AA}, a_{AB}, a_{BA}, \tilde{a}_{BB}))$  is an EZ. Part (i) of the definition of stability reversal applied to  $\mathfrak{Z}'$  requires that  $U^{\bullet}(a_{AB}, a_{BA}) > U^{\bullet}(\tilde{a}_{BB}, \tilde{a}_{BB})$  (where  $U^{\bullet}$  is the objective expected payoffs), but part (ii) of the same definition applied to  $\tilde{\mathfrak{Z}'}$  requires  $U^{\bullet}(\tilde{a}_{BB}, \tilde{a}_{BB}) \geq U^{\bullet}(a_{AB}, a_{BA})$ , a contradiction.

To show the second claim, by way of contradiction suppose  $\Theta_B$  is strategically independent and  $\mathfrak{Z} = (\mu_A, \mu_B, p = (0, 1), \lambda = 0, (a_{AA}, a_{AB}, a_{BA}, a_{BB}))$  is an EZ. By strategic independence, the adherents of  $\Theta_B$  find it optimal to play  $a_{BB}$  against any opponent strategy under the belief  $\mu_B$ . So, there exists another EZ of the form  $\mathfrak{Z}' = (\mu'_A, \mu_B, p = (0, 1), \lambda =$  $0, (a_{AA}, a'_{AB}, a_{BB}, a_{BB}))$ , where  $a'_{AB}$  is an objective best response to  $a_{BB}$ . The belief  $\mu_B$  is sustained because in both  $\mathfrak{Z}$  and  $\mathfrak{Z}'$ , the adherents of  $\Theta_B$  have the same data: from the strategy profile  $(a_{BB}, a_{BB})$ . In  $\mathfrak{Z}', \Theta_A$  is fitness is  $U^{\bullet}(a'_{AB}, a_{BB})$  and  $\Theta_B$  is fitness is  $U^{\bullet}(a_{BB}, a_{BB})$ . We have  $U^{\bullet}(a'_{AB}, a_{BB}) \geq U^{\bullet}(a_{BB}, a_{BB})$  since  $a'_{AB}$  is an objective best response to  $a_{BB}$ .

#### A.4 Proof of Proposition 3

Proof. By the hypotheses, find  $\bar{\epsilon} > 0$  so that for every  $0 < \epsilon \leq \bar{\epsilon}$ , there exists at least one EZ for  $\lambda = 0$  and for  $\lambda = 1$ , and furthermore  $\Theta_A$ 's fitness is weakly higher than that of  $\Theta_B$  in all such EZs. Let  $\lambda \in [0, 1]$  be given and for any  $0 < \epsilon \leq \bar{\epsilon}$ , let  $\mathfrak{Z} = (\mu_A, \mu_B, p = (1 - \epsilon, \epsilon), \lambda, (a))$ be an EZ. Since  $\Theta_A, \Theta_B$  are singleton models,  $\mathfrak{Z}_0 = (\mu_A, \mu_B, p = (1 - \epsilon, \epsilon), \lambda = 0, (a))$ and  $\mathfrak{Z}_1 = (\mu_A, \mu_B, p = (1 - \epsilon, \epsilon), \lambda = 1, (a))$  are also EZs. Let  $u_{g,g'}$  represent model  $\Theta_g$ 's conditional fitness against group g' in each of these three EZs. In the EZs with  $\lambda = 0$  and  $\lambda = 1$ , stability implies  $(1 - \epsilon)u_{AA} + \epsilon u_{AB} \geq \epsilon u_{BB} + (1 - \epsilon)u_{BA}$  and  $u_{AA} \geq u_{BB}$ . But this means  $(1 - \lambda) \cdot [(1 - \epsilon)u_{AA} + \epsilon u_{AB}] + \lambda \cdot [u_{AA}] \geq (1 - \lambda) \cdot [\epsilon u_{BB} + (1 - \epsilon)u_{BA}] + \lambda \cdot [u_{BB}]$ , which says  $\Theta_A$  has weakly higher fitness than  $\Theta_B$  in this EZ with the arbitrary  $\lambda \in [0, 1]$ . Also, at least one such EZ exists with assortativity  $\lambda$ , for at least one EZ exists when  $\lambda = 0$ , and the same equilibrium belief and behavior also constitutes an EZ for any other assortativity.  $\Box$ 

## A.5 Details Behind Example 2

Let  $b^*(a_i, a_{-i})$  solve  $\min_{b \in \mathbb{R}} D_{KL}(F^{\bullet}(a_i, a_{-i}) \parallel \hat{F}(a_i, a_{-i}; b, m)))$ , where  $F^{\bullet}(a_i, a_{-i})$  is the objective distribution over observations under the investment profile  $(a_i, a_{-i})$ , and  $\hat{F}(a_i, a_{-i}; b, m)$  is the distribution under the same investment profile in the model where productivity is given by  $P = b(x_i + x_{-i}) - m + \epsilon$ . We find that  $b^*(a_i, a_{-i}) = b^{\bullet} + \frac{m}{a_i + a_{-i}}$ . That is, adherents of  $\Theta_B$  end up with different beliefs about the game parameter b depending on the behavior of their typical opponents, which in turn affects how they respond to different rival investment levels. Stability reversal happens because when  $\Theta_A$  is resident and the adherents of  $\Theta_B$  always meet opponents who play  $a_i = 1$ , they end up with a more distorted belief about the fundamental than when  $\Theta_B$  is resident.

### A.6 Proof of Example 2

*Proof.* Define  $b^*(a_i, a_{-i}) := b^{\bullet} + \frac{m}{a_i + a_{-i}}$ . It is clear that  $D_{KL}(F^{\bullet}(a_i, a_{-i}) \parallel \hat{F}(a_i, a_{-i}; b^*(a_i, a_{-i}), m))) = 0$ , while this KL divergence is strictly positive for any other choice of b.

In every EZ with  $\lambda = 0$  and p = (1,0), we must have  $a_{AA} = a_{AB} = 1$ . If  $a_{BA} = 2$ , then the adherents of  $\Theta_B$  infer  $b^*(1,2) = b^{\bullet} + \frac{m}{3}$ . With this inference, the biased agents expect  $1 \cdot (2(b^{\bullet} + \frac{m}{3}) - m) = 2b^{\bullet} - \frac{m}{3}$  from playing 1 against rival investment 1, and expect  $2 \cdot (3(b^{\bullet} + \frac{m}{3}) - m) - c = 6b^{\bullet} - c$  from playing 2 against rival investment 1. Since  $4b^{\bullet} + \frac{m}{3} - c > 0$ 

from Condition 2, there is an EZ with  $a_{BA} = 2$  and  $\mu_B$  puts probability 1 on  $b^{\bullet} + \frac{m}{3}$ . It is impossible to have  $a_{BA} = 1$  in EZ. This is because  $b^*(1,1) > b^*(1,2)$ , and under the inference  $b^*(1,2)$  we already have that the best response to 1 is 2, so the same also holds under any higher belief about complementarity. Also, we have  $a_{BB} = 2$ , since 2 must best respond to both 1 and 2. So in every such EZ,  $\Theta_A$ 's conditional fitness against group A is  $2b^{\bullet}$  and  $\Theta_B$ 's conditional fitness against group A is  $6b^{\bullet} - c$ , with  $2b^{\bullet} > 6b^{\bullet} - c$  by Condition 1. Also,  $\Theta_A$ 's conditional fitness against group B is  $3b^{\bullet}$ , while  $\Theta_B$ 's conditional fitness against group B is  $8b^{\bullet} - c$ . Again,  $3b^{\bullet} > 8b^{\bullet} - c$  by Condition 1.

Next, we show  $\Theta_B$  has strictly higher fitness than  $\Theta_A$  in every EZ with  $\lambda = 0, p_B = 1$ . There is no EZ with  $a_{BB} = 1$ . This is because  $b^*(1, 1) = b^{\bullet} + \frac{m}{2}$ . As discussed before, under this inference the best response to 1 is 2, not 1. Now suppose  $a_{BB} = 2$ . Then  $\mu_B$  puts probability 1 on  $b^*(2, 2) = b^{\bullet} + \frac{m}{4}$ . With this inference, the biased agents expect  $1 \cdot (3(b^{\bullet} + \frac{m}{4}) - m) = 3b^{\bullet} - \frac{m}{4}$  from playing 1 against rival investment 2, and expect  $2 \cdot (4(b^{\bullet} + \frac{m}{4}) - m) - c = 8b^{\bullet} - c$  from playing 2 against rival investment 2. We have  $5b^{\bullet} + \frac{m}{4} - c > 0$  from Condition 2, so 2 best responds to 2. We must have  $a_{AA} = a_{AB} = 1$ . We conclude the unique EZ behavior is  $(a_{AA}, a_{AB}, a_{BA}, a_{BB}) = (1, 1, 1, 2)$ , since the biased agents expect  $1 \cdot (2(b^{\bullet} + \frac{m}{4}) - m) = 2b^{\bullet} - \frac{m}{2}$  from playing 1 against rival investment 1, and expect  $2 \cdot (3(b^{\bullet} + \frac{m}{4}) - m) - c = 6b^{\bullet} - \frac{m}{2} - c$  from playing 2 against rival investment 1. We have  $4b^{\bullet} - c < 0$  from Condition 1, so 1 best responds to 1. In the unique EZ with  $\lambda = 0$  and p = (0, 1), the fitness of  $\Theta_A$  is  $2b^{\bullet}$  and the fitness of  $\Theta_B$  is  $8b^{\bullet} - c$ , where  $8b^{\bullet} - c > 2b^{\bullet}$  by Condition 1.

### A.7 Proof of Example 3

Proof. Let  $KL_{4,1} := 0.4 \cdot \ln \frac{0.4}{0.1} + 0.6 \cdot \ln \frac{0.6}{0.9} \approx 0.3112$ ,  $KL_{4,8} := 0.4 \cdot \ln \frac{0.4}{0.8} + 0.6 \cdot \ln \frac{0.6}{0.2} \approx 0.3819$ , and  $KL_{2,4} := 0.2 \cdot \ln \frac{0.2}{0.4} + 0.8 \cdot \ln \frac{0.8}{0.6} \approx 0.0915$ . Let  $\lambda_h$  be the unique solution to  $(1 - \lambda)KL_{2,4} - \lambda(KL_{4,8} - KL_{4,1}) = 0$ , so  $\lambda_h \approx 0.564$ .

We show for any  $\lambda \in [0, \lambda_h)$ , for all sufficiently small  $\epsilon > 0$ , there exists a unique EZ  $\mathfrak{Z} = (\Theta_A, \Theta_B, \mu_A, \mu_B, p = (1 - \epsilon, \epsilon), \lambda, (a))$ , and that this EZ has  $\mu_B$  putting probability 1 on  $F_H$ ,  $a_{AA} = a_1$ ,  $a_{AB} = a_1$ ,  $a_{BA} = a_2$ ,  $a_{BB} = a_2$ . First, we may verify that under  $F_H$ ,  $a_2$  best responds to both  $a_1$  and  $a_2$ . Also, in the limit of  $\epsilon \to 0$ , the KL divergence of  $F_H$  approaches  $\lambda \cdot KL_{4,8}$  while that of  $F_L$  approaches  $\lambda \cdot KL_{4,1} + (1 - \lambda) \cdot KL_{2,4}$ . Since  $\lambda < \lambda_h$ , we see that for all sufficiently small  $\epsilon > 0$ ,  $F_H$  has strictly lower KL divergence. Finally, to check that there are no other EZs, note we must have  $a_{AA} = a_1$ ,  $a_{AB} = a_1$ ,  $a_{BA} = a_2$  in every EZ. In an EZ where  $a_{BB}$  puts probability  $q \in [0, 1]$  on  $a_2$ , as  $\epsilon \to 0$  the KL divergence of  $F_H$  approaches  $\lambda p \cdot KL_{4,8}$  and the KL divergence of  $F_L$  approaches  $\lambda p \cdot KL_{4,1} + (1-\lambda) \cdot KL_{2,4}$ . We have

$$\lambda q \cdot KL_{4,1} + (1-\lambda) \cdot KL_{2,4} - \lambda q \cdot KL_{4,8} = \lambda q \cdot (KL_{4,1} - KL_{4,8}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} = \lambda q \cdot (KL_{4,1} - KL_{4,8}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{2,4} \ge (1-\lambda) KL_{2,4} - \lambda (KL_{4,8} - KL_{4,1}) + (1-\lambda) KL_{4,8} - \lambda (KL_{4,8} - KL_{4,8}) + (1-\lambda) KL_{4,8} - KL_{4,8}) + (1-\lambda) KL_{4,8} - KL_{4,8}) + (1-\lambda) KL_{4,8} - KL$$

Since  $\lambda < \lambda_h$ , this is strictly positive. Therefore we must have  $\mu_B$  put probability 1 on  $F_H$ (which in turn implies q = 1) when  $\epsilon > 0$  is sufficiently small.

When  $\Theta_A$  is dominant and  $p_A = 1$ , the equilibrium fitness of  $\Theta_A$  is always 0.25 for every  $\lambda$ . The equilibrium fitness of  $\Theta_B$ , as a function of  $\lambda$ , is  $0.4\lambda + 0.2(1 - \lambda)$ . Let  $\lambda_l$  solve  $0.25 = 0.4\lambda + 0.2(1 - \lambda)$ , that is  $\lambda_l = 0.25$ . This shows  $\Theta_A$  is evolutionarily fragile against  $\Theta_B$  for  $\lambda \in (\lambda_l, \lambda_h)$ , and it is evolutionarily stable against  $\Theta_B$  for  $\lambda = 0$ .

Now suppose  $\lambda = 1$ . For sufficiently small  $\epsilon > 0$ , if there is an EZ with  $p_A = 1$  where  $a_{BB}$ plays  $a_2$  with positive probability, then  $\mu_B$  must put probability 1 on  $F_L$ , since  $KL_{4,1} < KL_{4,8}$ . This is a contradiction, since  $a_2$  does not best respond to itself under  $F_L$ . So the unique EZ for all small enough  $\epsilon > 0$  involves  $a_{AA} = a_1$ ,  $a_{AB} = a_1$ ,  $a_{BA} = a_2$ ,  $a_{BB} = a_3$ . In the EZ, the fitness of  $\Theta_A$  approaches 0.25 and the fitness of  $\Theta_B$  approaches 0.2 as  $\epsilon \to 0$ . This shows  $\Theta_A$ is evolutionarily stable against  $\Theta_B$  for  $\lambda = 1$ .

### A.8 Proof of Proposition 4

Proof. When  $\Theta_A = \Theta_B = \Theta^{\bullet}$ , for any matching assortativity  $\lambda$  and with any  $(p_A, p_B)$ , we show adherents of both models have 0 fitness in every EZ. Suppose instead that the match between groups g and g' reach a terminal node other than  $z_1$  with positive probability. Let  $n_L$  be the last non-terminal node reached with positive probability, so we must have  $L \geq 2$ , and also that nodes  $n_1, \ldots, n_{L-1}$  are also reached with positive probability. So Drop must be played with probability 1 at  $n_L$ . Since  $n_L$  is reached with positive probability, correctly specified agents hold correct beliefs about opponent's play at  $n_L$ , which means at  $n_{L-1}$  it cannot be optimal to play Across with positive probability since this results in a loss of  $\ell$ compared to playing Drop, a contradiction.

Now let  $\Theta_A = \Theta^{\bullet}$ ,  $\Theta_B = \Theta^{An}$ . Suppose  $\lambda \in [0, 1]$  and let  $p_B \in (0, 1)$ . We claim there is an EZ where  $d_{AA}^k = 1$  for every k,  $d_{AB}^k = 0$  for every even k with k < K,  $d_{AB}^k = 1$  for every other k,  $d_{BA}^k = 0$  for every odd k and  $d_{BA}^k = 1$  for every even k, and  $d_{BB}^k = 0$  for every kwith k < K,  $d_{BB}^K = 1$ . It is easy to see that the behavior  $(d_{AA})$  is optimal under correct belief about opponent's play. In the  $\Theta_A$  vs.  $\Theta_B$  matches, the conjecture about A's play  $\hat{d}_{AB}^k = 2/K$ for k even,  $\hat{d}_{AB}^k = 1$  for k odd minimizes KL divergence among all strategies in  $\mathbb{A}^{An}$ , given B's play. To see this, note that when B has the role of P2, opponent Drops immediately. When B has the role of P1, the outcome is always  $z_K$ . So a conjecture with  $\hat{d}_{AB}^k = x$  for every even k has the conditional KL divergence of:

$$\sum_{k \le K-1 \text{ odd}} \underbrace{\underbrace{0 \cdot \ln\left(\frac{0}{0}\right)}_{(1,z_k) \text{ for } k \le K-1 \text{ odd}} + \sum_{k \le K-1 \text{ even}} \underbrace{0 \cdot \ln\left(\frac{0}{(1/2) \cdot (1-x)^{(k/2)-1} \cdot x}\right)}_{(1,z_k) \text{ for } k \le K-1 \text{ even}} + \underbrace{\frac{1}{2} \ln\left(\frac{1/2}{(1/2) \cdot (1-x)^{(K/2)-1} \cdot x}\right)}_{(1,z_K)} + \underbrace{0 \cdot \ln\left(\frac{0}{(1-x)^{(K/2)}}\right)}_{(1,z_{end})}$$

when matched with an opponent from  $\Theta_A$ . Using  $0 \cdot \ln(0) = 0$ , the expression simplifies to  $\frac{1}{2} \ln \left( \frac{1}{(1-x)^{(K/2)-1} \cdot x} \right)$ , which is minimized among  $x \in [0,1]$  by x = 2/K. Against this conjecture, the difference in expected payoff at node  $n_{K-1}$  from Across versus Drop is  $(1-2/K)(g)+(2/K)(-\ell)$ . This is strictly positive when  $g > \frac{2}{K-2}\ell$ . This means the continuation value at  $n_{K-1}$  is at least g larger than the payoff of Dropping at  $n_{K-3}$ , so again Across has strictly higher expected payoff than Drop. Inductively,  $(d_{BA}^k)$  is optimal given the belief  $(\hat{d}_{AB}^k)$ . Also,  $(d_{AB}^k)$  is optimal as it results in the highest possible payoff. We can similarly show that the conjecture  $\hat{d}_{BB}^k$  with  $\hat{d}_{BB}^k = 2/K$  for k even,  $\hat{d}_{BB}^k = 0$  for k odd minimizes KL divergence conditional on  $\Theta_B$  opponent, and  $(d_{BB}^k)$  is optimal given this conjecture.

As  $p_B \to 0$ , we find an EZ where adherents of A have fitness approaching 0, whereas the adherents of B have fitness approaching at least  $\frac{1}{2}(((K/2) - 1)g - \ell) > 0$  since  $g > \frac{2}{K-2}\ell$ . This shows  $\Theta_A$  is not evolutionarily stable against  $\Theta_B$ .

But consider the same  $(d_{AA}, d_{AB}, d_{BA})$  and suppose  $d_{BB}^k = 1$  for every k. Taking  $p_B \to 1$ , with  $\lambda < 1$ , we find an EZ where adherents of B have fitness 0, adherents of A have fitness  $(1 - \lambda) \cdot \frac{1}{2} \cdot ((K/2)g + \ell) > 0$ . This shows  $\Theta_B$  is not evolutionarily stable against  $\Theta_A$ .  $\Box$ 

### A.9 Proof of Proposition 5

*Proof.* In the centipede game, suppose  $g > \frac{2}{K-2}\ell$ . the misspecified agent thinks a group B agent in the role of P2 and a group A agent in either role has a probability 2/K of stopping at every node. Under this belief, choosing to continue instead of drop means there is a (K-2)/K chance of gaining g, but a 2/K chance of losing  $\ell$ . Since we assume  $g > \frac{2}{K-2}\ell$ , it is strictly better to continue. When p fraction of the agents are correctly

specified, the fitness of  $\Theta^{\bullet}$  is  $p \cdot 0 + (1-p) \cdot (\frac{1}{2}\frac{g(K-2)}{2} + \frac{1}{2}(\frac{gK}{2} + \ell))$ , while the fitness of  $\Theta^{An}$  is  $p \cdot [\frac{1}{2}(\frac{g(K-2)}{2} - \ell) + \frac{1}{2}\frac{g(K-2)}{2}] + (1-p)[\frac{1}{2}(\frac{g(K-2)}{2} - \ell) + \frac{1}{2}(\frac{gK}{2} + \ell)]$ . The difference in fitness is  $-p[\frac{1}{2}(\frac{g(K-2)}{2} - \ell) + \frac{1}{2}\frac{g(K-2)}{2}] + (1-p)\frac{1}{2}\ell$ . Simplifying, this is  $\frac{1}{2}\ell - p \cdot \frac{g(K-2)}{2}$ , a strictly decreasing function in p. When  $p = \frac{\ell}{g(K-2)}$ , which is a number strictly between 0 and 1/2 from the assumption  $g > \frac{2}{K-2}\ell$  in the centipede game, the two models have the same fitness. Furthermore, since the payoff difference is linear in p with a negative slope, the difference in fitness is negative when  $p > \frac{\ell}{g(K-2)}$ —so that  $\Theta^{An}$  outperforms  $\Theta^{\bullet}$  under these population shares—and conversely, the difference in fitness is positive when  $p < \frac{\ell}{g(K-2)}$ . Thus, we have this fraction of the population being correctly specified forms a stable population share.  $\Box$ 

# A.10 Proof of Proposition 6

Proof. In the  $\overline{\Theta}^{An}$  vs.  $\overline{\Theta}^{An}$  match, the adherents of  $\overline{\Theta}^{An}$  hold the belief that  $\hat{d}_{BB}^k = 2/K$  for every even k. In the role of P1, at node k for  $k \leq K-3$ , stopping gives them k but continuing gives them a (K-2)/K chance to get at least k+2, and we have  $k \leq \frac{K-2}{K}(k+2) \iff 2k \leq 2K-4 \iff k \leq K-2$ . At node K-1, the agent gets K-1 from dropping but expects  $(K+2) \cdot \frac{K-2}{K}$  from continuing, and  $(K+2) \cdot \frac{K-2}{K} - (K-1) = \frac{K^2-4-K^2+K}{K} = \frac{K-4}{K} > 0$  since  $K \geq 6$ .

In the  $\overline{\Theta}^{\bullet}$  vs.  $\overline{\Theta}^{An}$  match, the adherents of  $\Theta^{An}$  hold the belief that  $\hat{d}_{AB}^{k} = 2/K$  for every k. By the same arguments as before, the behavior of the adherents of  $\Theta^{An}$  are optimal given these beliefs. Also, the adherents of  $\Theta^{\bullet}$  have no profitable deviations since they are best responding both as P1 and P2.

When p fraction of the agents are correctly specified, in the dollar game the fitness of  $\overline{\Theta}^{\bullet}$  is  $p \cdot 0.5 + (1-p) \cdot (\frac{1}{2}(K-1) + \frac{1}{2}K)$ , while the fitness of  $\overline{\Theta}^{An}$  is  $p \cdot 0 + (1-p) \cdot (\frac{1}{2} \cdot 0 + \frac{1}{2}K)$ . For any p, the fitness of  $\overline{\Theta}^{\bullet}$  is strictly higher than that of  $\overline{\Theta}^{An}$ .

# **B** Existence and Continuity of EZ

We provide a few technical results about the existence of EZ and the upper-hemicontinuity of the set of EZs with respect to population share. We suppose that  $|\mathcal{G}| = 1$  for simplicity, but analogous results would hold for environments with multiple situations. Note that the same learning channel that generates new stability phenomena in Section 3 also leads to some difficulty in establishing existence and continuity results, as agents draw different inferences with different interaction structures.

Let two models,  $\Theta_A, \Theta_B$  be fixed. Also fix population shares p and matching assortativity  $\lambda$ . Let  $U_A : \mathbb{A}^2 \times \Theta_A \to \mathbb{R}$  be such that  $U_A(a_i, a_{-i}; F) = U_i(a_i, a_{-i}; \delta_F)$  and let  $U_B : \mathbb{A}^2 \times \Theta_B \to \mathbb{R}$  be such that  $U_B(a_i, a_{-i}; F) = U_i(a_i, a_{-i}; \delta_F)$ .

Assumption A.1.  $\mathbb{A}, \Theta_A, \Theta_B$  are compact metrizable spaces.

Assumption A.2.  $U_A, U_B$  are continuous.

**Assumption A.3.** For every  $F \in \Theta_A \cup \Theta_B$  and  $a_i, a_{-i} \in \mathbb{A}$ ,  $K(F; a_i, a_{-i})$  is well-defined and finite.

Under Assumption A.3, we have the well-defined functions  $K_A : \Theta_A \times \mathbb{A}^2 \to \mathbb{R}_+$  and  $K_B : \Theta_B \times \mathbb{A}^2 \to \mathbb{R}_+$ , where  $K_g(F; a_i, a_{-i}) := D_{KL}(F^{\bullet}(a_i, a_{-i}) \parallel F(a_i, a_{-i})).$ 

Assumption A.4.  $K_A$  and  $K_B$  are continuous.

**Assumption A.5.** A *is convex and, for all*  $a_{-i} \in A$  *and*  $\mu \in \Delta(\Theta_A) \cup \Delta(\Theta_B)$ *,*  $a_i \mapsto U_i(a_i, a_{-i}; \mu)$  *is quasiconcave.* 

We show existence of EZ using the Kakutani-Fan-Glicksberg fixed point theorem, applied to the correspondence which maps strategy profiles and beliefs over parameters into best replies and beliefs over KL-divergence minimizing parameter. We start with a lemma.

**Lemma A.1.** For  $g \in \{A, B\}$ ,  $a = (a_{AA}, a_{AB}, a_{BA}, a_{BB}) \in \mathbb{A}^4$ , and  $0 \le m_q \le 1$ , let

$$\Theta_g^*(a, m_g) := \underset{\hat{F} \in \Theta_g}{\arg\min} \left\{ m_g \cdot K(\hat{F}; a_{g,g}, a_{g,g}) + (1 - m_g) \cdot K(\hat{F}; a_{g,-g}, a_{-g,g}) \right\}.$$

Then,  $\Theta_g^*$  is upper hemicontinuous in its arguments.

This lemma says the set of KL-minimizing parameters is upper hemicontinuous in strategy profile and matching assortativity. This leads to the existence result.

**Proposition A.1.** Under Assumptions A.1, A.2, A.3, A.4, and A.5, an EZ exists.

Next, upper hemicontinuity in  $m_g$  in Lemma A.1 allows us to deduce the upper hemicontinuity of the EZ correspondence in population shares.

**Proposition A.2.** Fix two models  $\Theta_A, \Theta_B$ . Also fix matching assortativity  $\lambda \in [0, 1]$ . The set of EZ is an upper hemicontinuous correspondence in  $p_B$  under Assumptions A.1, A.2, A.3, and A.4.

## B.1 Proofs of Results in Appendix B

#### B.1.1 Proof of Lemma A.1

*Proof.* Write the minimization objective as

$$W(a, F, m_g) := m_g K_g(F; a_{g,g}, a_{g,g}) + (1 - m_g) K_g(F; a_{g,-g}, a_{-g,g}),$$

a continuous function of  $(a, F, m_g)$  by Assumption A.4. Suppose we have a sequence  $(a^{(n)}, m_g^{(n)}) \rightarrow (a^*, m_g^*) \in \mathbb{A}^4 \times [0, 1]$  and let  $F^{(n)} \in \Theta_g^*(a^{(n)}, m_g^{(n)})$  for each n, with  $F^{(n)} \rightarrow F^* \in \Theta_g$ . For any other  $\hat{F} \in \Theta_g$ , note that  $W(a^*, m_g^*, \hat{F}) = \lim_{n \to \infty} W(a^{(n)}, m_g^{(n)}, \hat{F})$  by continuity. But also by continuity,  $W(a^*, m_g^*, F^*) = \lim_{n \to \infty} W(a^{(n)}, m_g^{(n)}, F^{(n)})$  and  $W(a^{(n)}, m_g^{(n)}, F)$  for every n. It therefore follows  $W(a^*, m_g^*, F^*) \leq W(a^*, m_g^*, \hat{F})$ .  $\Box$ 

#### B.1.2 Proof of Proposition A.1

*Proof.* Consider the correspondence  $\Gamma : \mathbb{A}^4 \times \Delta(\Theta_A) \times \Delta(\Theta_B) \rightrightarrows \mathbb{A}^4 \times \Delta(\Theta_A) \times \Delta(\Theta_B)$ ,

$$\Gamma(a_{AA}, a_{AB}, a_{BA}, a_{BB}, \mu_A, \mu_B) := (BR(a_{AA}, \mu_A), BR(a_{BA}, \mu_A), BR(a_{AB}, \mu_B), BR(a_{BB}, \mu_B), \Delta(\Theta_A^*(a)), \Delta(\Theta_B^*(a))),$$

where  $BR(a_{-i}, \mu_g) := \underset{\hat{a}_i \in \mathbb{A}}{\operatorname{arg max}} U_g(\hat{a}_i, a_{-i}; \mu_g)$  and, for each  $g \in \{A, B\}$ , the correspondence  $\Theta_g^*$  is defined with  $m_g = \lambda + (1 - \lambda)p_g$ ,  $m_{-g} = 1 - m_g$ . It is clear that fixed points of  $\Gamma$  are EZ.

We apply the Kakutani-Fan-Glicksberg theorem (see, e.g, Corollary 17.55 in Aliprantis and Border (2006)). By Assumptions A.1 and A.5,  $\mathbb{A}$  is accompact and convex metric space, and each  $\Theta_g$  is a compact metric space, so it follows the domain of  $\Gamma$  is a nonempty, compact and convex metric space. We need only verify that  $\Gamma$  has closed graph, non-empty values, and convex values.

To see that  $\Gamma$  has closed graph, the previous lemma shows the upper hemicontinuity of  $\Theta_A^*(a)$  and  $\Theta_B^*(a)$  in a, and Theorem 17.13 of Aliprantis and Border (2006) then implies  $\Delta(\Theta_A^*(a))$  and  $\Delta(\Theta_B^*(a))$  are also upper hemicontinuous in a. It is a standard argument that since Assumption A.2 supposes  $U_A, U_B$  are continuous, it implies the best-response correspondences BR $(a_{AA}, \mu_A)$ , BR $(a_{BA}, \mu_A)$ , BR $(a_{AB}, \mu_B)$ , BR $(a_{BB}, \mu_B)$  have closed graphs.

To see that  $\Gamma$  is non-empty, recall that each  $\hat{a}_i \mapsto U_g(\hat{a}_i, a_{-i}; \mu_g)$  is a continuous function on a compact domain, so it must attain a maximum on  $\mathbb{A}$ . Similarly, the minimization problem that defines each  $\Theta_g^*(a)$  is a continuous function of F over a compact domain of possible F's, so it attains a minimum. Thus each  $\Delta(\Theta_g^*(a))$  is the set of distributions over a non-empty set.

To see that  $\Gamma$  is convex valued, clearly  $\Delta(\Theta_A^*(a))$  and  $\Delta(\Theta_B^*(a))$  are convex valued by definition. Also,  $\hat{a}_i \mapsto U_A(\hat{a}_i, a_{AA}; \mu_A)$  is quasiconcave by Assumption A.5. That means if  $a'_i, a''_i \in BR(a_{AA}, \mu_A)$ , then for any convex combination  $\tilde{a}_i$  of  $a'_i, a''_i$ , we have  $U_A(\tilde{a}_i, a_{AA}; \mu_A) \ge$  $\min(U_A(a'_i, a_{AA}; \mu_A), U_A(a''_i, a_{AA}; \mu_A)) = \max_{\hat{a}_i \in \mathbb{A}} U_A(\hat{a}_i, a_{AA}; \mu_A)$ . Therefore,  $BR(a_{AA}, \mu_A)$  is convex. For similar reasons,  $BR(a_{BA}, \mu_A)$ ,  $BR(a_{AB}, \mu_B)$ ,  $BR(a_{BB}, \mu_B)$  are convex.  $\Box$ 

#### B.1.3 Proof of Proposition A.2

Proof. Since  $\mathbb{A}^4 \times \Delta(\Theta_A) \times \Delta(\Theta_B)$  is compact by Assumption A.1, we need only show that for every sequence  $(p_B^{(k)})_{k\geq 1}$  and  $(a^{(k)}, \mu^{(k)})_{k\geq 1} = (a_{AA}^{(k)}, a_{AB}^{(k)}, a_{BB}^{(k)}, a_{BB}^{(k)}, \mu_A^{(k)}, \mu_B^{(k)})_{k\geq 1}$  such that for every k,  $(a^{(k)}, \mu^{(k)})$  is an EZ with  $p = (1 - p_B^{(k)}, p_B^{(k)}), p_B^{(k)} \to p_B^*$ , and  $(a^{(k)}, \mu^{(k)}) \to (a^*, \mu^*)$ , then  $(a^*, \mu^*)$  is an EZ with  $p = (1 - p_B^*, p_B^*)$ .

We first show for all  $g, g' \in \{A, B\}$ ,  $a_{g,g'}^*$  is optimal against  $a_{g',g}^*$  under the belief  $\mu_g^*$ . Assortativity does not matter here, since optimality applies within all type match-ups. By Assumption A.2,  $U_g(a_i, a_{-i}; F)$  is continuous, so by property of convergence in distribution,  $U_g(a_{g,g'}^{(k)}, a_{g',g}^{(k)}; \mu_g^{(k)}) \to U_g(a_{g,g'}^*, a_{g',g}^*; \mu_g^*)$ . For any other  $\hat{a}_i \in \mathbb{A}$ ,  $U_g(\hat{a}_i, a_{g',g}^{(k)}; \mu_g^{(k)}) \to$  $U_g(\hat{a}_i, a_{g',g}^*; \mu_g^*)$  and for every k,  $U_g(a_{g,g'}^{(k)}, a_{g',g}^{(k)}; \mu_g^{(k)}) \ge U_g(\hat{a}_i, a_{g',g}^{(k)}; \mu_g^{(k)})$ . Therefore  $a_{g,g'}^*$  best responds to  $a_{g',g}^*$  under belief  $\mu_g^*$ .

Next, we show parameters in the support of  $\mu_g^*$  minimize weighted KL divergence for group g. First consider the correspondence  $H : \mathbb{A}^4 \times [0,1] \rightrightarrows \Theta_g$  where  $H(a, p_g) := \Theta_g^*(a, \lambda + (1-\lambda)(p_g))$ . Then H is upper hemicontinuous by Lemma A.1. Since  $H(a, p_g)$  represents the minimizers of a continuous function on a compact domain, it is non-empty and closed. By Theorem 17.13 of Aliprantis and Border (2006), the correspondence  $\tilde{H} : \mathbb{A}^4 \times [0,1] \rightrightarrows \Delta(\Theta_g)$ defined so that  $\tilde{H}(a, p_g) := \Delta(H(a, p_g))$  is also upper hemicontinuous. For every  $k, \mu_g^{(k)} \in \tilde{H}(a^{(k)}, p_g^{(k)})$ , and  $\mu_g^{(k)} \to \mu_g^*, a^{(k)} \to a^*, p_g^{(k)} \to p_g^*$ . Therefore,  $\mu_g^* \in \tilde{H}(a^*, p_g^*)$ , that is to say  $\mu_g^*$ is supported on the minimizers of weighted KL divergence.

# C Learning Foundation of EZ and EZ-SU

We provide a unified foundation for EZ and EZ-SU as the steady state of a learning system. This foundation considers a world where agents have prior beliefs over extended parameters in an extended model, as in Appendix 4. At the end of every match, each agent observes her consequence and a noisy signal about the matched opponent's strategy. We show that under any asymptotically myopic policy, if behavior and beliefs converge, then the limit steady state must be an EZ-SU when the noisy signals about opponent's strategy are uninformative. Sufficiently accurate signals about opponent's play cause the steady states to be EZs, if the extended models allow agents to make rich enough inferences about opponents' strategies. Finally, if the true situation is redrawn every T periods and the agents reset their beliefs over extended parameters to their prior belief when the situation is redrawn, then their average payoffs approach their fitness in the EZ or EZ-SU when T is large.

### C.1 Regularity Assumptions

We make some regularity assumptions on the objective environments and on the extended models  $\overline{\Theta}_A, \overline{\Theta}_B$ . These are similar to the regularity assumptions from Appendix B.

Suppose the strategy set  $\mathbb{A}$  is finite. Suppose the marginals of the extended models  $\overline{\Theta}_A, \overline{\Theta}_B$  on the dimension of fundamental uncertainty, denoted as  $\Theta_A, \Theta_B$ , are compact and metrizable spaces. Endow  $\overline{\Theta}_A$  and  $\overline{\Theta}_B$  with the product metric. Suppose that every  $(a_A, a_B, F) \in \overline{\Theta}_A \cup \overline{\Theta}_B$  is so that for every  $(a_i, a_{-i}) \in \mathbb{A}^2$  and every situation G, whenever  $f^{\bullet}(a_i, a_{-i}, G)(y) > 0$ , we also get  $f(a_i, a_A)(y) > 0$  and  $f(a_i, a_B)(y) > 0$ , where f is the density or probability mass function for F.

For each  $g, g' \in \{A, B\}$ , define  $K_{g,g'} : \mathbb{A}^2 \times \mathcal{G} \times \overline{\Theta}_g \to \mathbb{R}$  by  $K_{g,g'}(a_i, a_{-i}, G; (a_A, a_B, F)) = D_{KL}(F^{\bullet}(a_i, a_{-i}, G) \parallel F(a_i, a_{g'}))$ . This is the KL divergence of the parameter  $(a_A, a_B, F) \in \overline{\Theta}_g$  in situation G based on the data generated from the strategy profile  $(a_i, a_{-i})$ . Suppose each  $K_{q,q'}$  is well defined and a continuous function of the extended parameter  $(a_A, a_B, F)$ .

For  $g \in \{A, B\}$ ,  $F \in \Theta_g$ , let  $U_g(a_i, a_{-i}; F)$  be the expected payoffs of the strategy profile  $(a_i, a_{-i})$  for *i* when consequences are drawn according to *F*. Assume  $U_A, U_B$  are continuous.

Suppose for every extended model  $\overline{\Theta}_g$  and every  $(a_A, a_B, F) \in \overline{\Theta}_g$  and  $\epsilon > 0$ , there exists an open neighborhood  $V \subseteq \overline{\Theta}_g$  of  $(a_A, a_B, F)$ , so that for every  $(\hat{a}_A, \hat{a}_B, \hat{F}) \in V$ ,  $1 - \epsilon \leq f(a_i, a_A)(y)/\hat{f}(a_i, \hat{a}_A)(y) \leq 1 + \epsilon$  and  $1 - \epsilon \leq f(a_i, a_B)(y)/\hat{f}(a_i, \hat{a}_B)(y) \leq 1 + \epsilon$  for all  $a_i \in \mathbb{A}, y \in \mathbb{Y}$ . Also suppose there is some M > 0 so that  $\ln(f(a_i, a_A)(y))$  and  $\ln(f(a_i, a_B)(y))$  are bounded in [-M, M] for all  $(a_A, a_B, F) \in \overline{\Theta}_g$ ,  $a_i, a_{-i} \in \mathbb{A}, y \in \mathbb{Y}$ .

## C.2 Learning Environment

We first consider an environment with only one true situation,  $|\mathcal{G}| = 1$ . Time is discrete and infinite, t = 0, 1, 2, ... A unit mass of agents,  $i \in [0, 1]$ , enter the society at time 0. A  $p_A \in (0, 1)$  measure of them are assigned to model A and the rest are assigned to model B. Each agent born into model g starts with the same full support prior over the extended model,  $\mu_g^{(0)} \in \Delta(\overline{\Theta}_g)$ , and believes there is some  $(a_A, a_B, F) \in \overline{\Theta}_g$  so that every group g opponent always plays  $a_g$  and the consequences are always generated by F.

In each period t, agents are matched up partially assortatively to play the stage game. Assortativity is  $\lambda \in (0, 1)$ . Each person in group g has  $\lambda + (1 - \lambda)p_g$  chance of matching with someone from group g, and matches with someone from group -g with the complementary chance. Each agent i observes their opponent's group membership and chooses a strategy  $a_i^{(t)} \in \mathbb{A}$ . At the end of the match, the agent observes own consequence  $y_i^{(t)}$  and a signal  $x_i^{(t)} \in \mathbb{A}$  about the opponent's play, where  $x_i^{(t)}$  equals the matched opponent's strategy  $a_{-i}$  with probability  $\tau \in [0, 1)$ , and it is uniformly random on  $\mathbb{A}$  with the complementary probability. To give a foundation for a EZ-SU, we consider  $\tau = 0$ , so the signal  $x_i$  is uninformative. To give a foundation for EZ, we consider  $\tau$  close to 1.

Thus, the space of histories from one period is  $\{A, B\} \times \mathbb{A} \times \mathbb{Y} \times \mathbb{A}$ , with typical element  $(g_i^{(t)}, a_i^{(t)}, y_i^{(t)}, x_i^{(t)})$ . It records the group membership of *i*'s opponent  $g_i^{(t)}$ , *i*'s strategy  $a_i^{(t)}$ , *i*'s consequence  $y_i^{(t)}$ , and *i*'s ex-post signal about the matched opponent's play,  $x_i^{(t)}$ . Let  $\mathbb{H}$  denote the space of all finite-length histories.

Given the assumption on the two models, there is a well-defined Bayesian belief operator for each model  $g, \mu_g : \mathbb{H} \to \Delta(\overline{\Theta}_g)$ , mapping every finite-length history into a belief over extended parameters in  $\overline{\Theta}_g$ , starting with the prior  $\mu_q^{(0)}$ .

We also take as exogenously given policy functions for choosing strategies after each history. That is,  $\mathfrak{a}_{g,g'} : \mathbb{H} \to \mathbb{A}$  for every  $g, g' \in \{A, B\}$  gives the strategy that a group g agent uses against a group g' opponent after every history. Assume these policy functions are asymptotically myopic.

Assumption A.6. For every  $\epsilon > 0$ , there exists N so that for any history h containing at least N matches against opponents of each group,  $\mathfrak{a}_{g,g'}(h)$  is an  $\epsilon$ -best response to the Bayesian belief  $\mu_g(h)$ .

From the perspective of each agent *i* in group *g*, *i*'s play against groups A and B, as well as *i*'s belief over  $\overline{\Theta}_g$ , is a stochastic process  $(\tilde{a}_{iA}^{(t)}, \tilde{a}_{iB}^{(t)}, \tilde{\mu}_i^{(t)})_{t\geq 0}$  valued in  $\mathbb{A} \times \mathbb{A} \times \mathbb{A}$ 

 $\Delta(\Theta_g)$ . The randomness is over the groups of opponents matched with in different periods, the strategies they play, and the random consequences and ex-post signals drawn at the end of the matches. Since there is a continuum of agents, the distribution over histories within each population in each period is deterministic. As such, there is a deterministic sequence  $(\alpha_{AA}^{(t)}, \alpha_{AB}^{(t)}, \alpha_{BA}^{(t)}, \alpha_{A}^{(t)}, \nu_{B}^{(t)}) \in \Delta(\mathbb{A})^4 \times \Delta(\Delta(\overline{\Theta}_A)) \times \Delta(\Delta(\overline{\Theta}_B))$  that describes the distributions of play and beliefs that prevail in the two sub-populations in every period t.

#### C.3 Steady State Limits are EZ-SUs and EZs

We state and prove the learning foundation of EZ-SU and EZ. For  $(\alpha^{(t)})_t$  a sequence valued in  $\Delta(\mathbb{A})$  and  $a^* \in \mathbb{A}$ ,  $\alpha^{(t)} \to a^*$  means  $\mathbb{E}_{\hat{a} \sim \alpha^{(t)}} \parallel \hat{a} - a^* \parallel \to 0$  as  $t \to \infty$ . For  $(\nu^{(t)})_t$  a sequence valued in  $\Delta(\Delta(\overline{\Theta}_g))$  and  $\mu^* \in \Delta(\overline{\Theta}_g)$ ,  $\nu^{(t)} \to \mu^*$  means  $\mathbb{E}_{\hat{\mu} \sim \nu^{(t)}} \parallel \hat{\mu} - \mu^* \parallel \to 0$  as  $t \to \infty$ .

**Proposition A.3.** Suppose the regularity assumptions in Appendix C.1 hold, and suppose Assumption A.6 holds.

Suppose  $\tau = 0$ . Suppose there exists  $(a_{AA}^*, a_{AB}^*, a_{BA}^*, a_{BB}^*, \mu_A^*, \mu_B^*) \in \mathbb{A}^4 \times \Delta(\overline{\Theta}_A) \times \Delta(\overline{\Theta}_B)$ so that  $(\alpha_{AA}^{(t)}, \alpha_{AB}^{(t)}, \alpha_{BA}^{(t)}, \alpha_{A}^{(t)}, \nu_A^{(t)}, \nu_B^{(t)}) \to (a_{AA}^*, a_{AB}^*, a_{BA}^*, a_{BB}^*, \mu_A^*, \mu_B^*)$  and for each agent *i* in group *g*, almost surely  $(\tilde{a}_{iA}^{(t)}, \tilde{a}_{iB}^{(t)}, \tilde{\mu}_i^{(t)}) \to (a_{gA}^*, a_{gB}^*, \mu_g^*)$ . Then,  $(a_{AA}^*, a_{AB}^*, a_{BA}^*, a_{BB}^*, \mu_A^*, \mu_B^*)$  is an EZ-SU.

Suppose for each g, the extended model  $\overline{\Theta}_g = \mathbb{A}^2 \times \Theta_g$  for some model  $\Theta_g$  – that is, each group can make any inference about opponents' strategies. There exists some  $\underline{\tau} < 1$ so that for every  $\tau \in (\underline{\tau}, 1)$  and  $(a_{AA}^*, a_{AB}^*, a_{BA}^*, a_{BB}^*, \mu_A^*, \mu_B^*)$  satisfying the above conditions, we have that  $\mu_A^*$  puts probability 1 on  $(a_{AA}^*, a_{AB}^*)$ ,  $\mu_B^*$  puts probability 1 on  $(a_{BA}^*, a_{BB}^*)$ , and  $(a_{AA}^*, a_{AB}^*, a_{BA}^*, a_{BB}^*, \mu_A^*|_{\Theta_A}, \mu_B^*|_{\Theta_B})$  is an EZ, where  $\mu_g^*|_{\Theta_g}$  is the marginal of the belief  $\mu_g^*$  on the model  $\Theta_g$ .

Proof. We first consider the case of  $\tau = 0$ , so the uninformative ex-post signals may be ignored. For  $\mu$  a belief and  $g \in \{A, B\}$ , let  $u^{\mu}(a_i; g)$  represent subjective expected payoff from playing  $a_i$  against group g. Suppose  $a_{AA}^* \notin \operatorname{argmax}_{\hat{a} \in \mathbb{A}} u^{\mu_A^*}(\hat{a}; A)$  (the other cases are analogous). By the continuity assumptions on  $U_A$  (which is also bounded because  $\Theta_A$  is bounded), there are some  $\epsilon_1, \epsilon_2 > 0$  so that whenever  $\mu_i \in \Delta(\overline{\Theta}_A)$  with  $\| \mu_i - \mu_A^* \| < \epsilon_1$ , we also have  $u^{\mu_i}(a_{AA}^*; A) < \max_{\hat{a} \in \mathbb{A}} u^{\mu_i}(\hat{a}; A) - \epsilon_2$ . By the definition of asymptotically empirical best responses, find N so that  $\mathfrak{a}_{A,A}(h)$  must be a myopic  $\epsilon_2$ -best response when there are at least N periods of matches against A and B. Agent i has a strictly positive chance to match with groups A and B in every period. So, at all except a null set of points in the probability space, *i*'s history eventually records at least N periods of play by groups A and B. Also, by assumption, almost surely  $\tilde{\mu}_i^{(t)} \to \mu_A^*$ . This shows that by asymptotically myopic best responses, almost surely  $\tilde{a}_{iA}^{(k)} \not\rightarrow a_{AA}^*$ , a contradiction.

Now suppose some  $\theta_A^* = (a_A^*, a_B^*, f^*)$  in the support of  $\mu_A^*$  does not minimize the weighted KL divergence in the definition of EZ-SU (the case of a parameter  $\theta_B^*$  in the support of  $\mu_B^*$  not minimizing is similar). Then we have

$$\theta_A^* \notin \underset{\hat{\theta} \in \overline{\Theta}_A}{\operatorname{argmin}} \left[ \begin{array}{c} (\lambda + (1 - \lambda)p_A) \cdot D_{KL}(F^{\bullet}(a_{AA}^*, a_{AA}^*) \parallel \hat{F}(a_{AA}^*, \hat{a}_A)) \\ + (1 - \lambda)(1 - p_A) \cdot D_{KL}(F^{\bullet}(a_{AB}^*, a_{BA}^*) \parallel \hat{F}(a_{AB}^*, \hat{a}_B)) \end{array} \right]$$

where  $\hat{\theta} = (\hat{a}_A, \hat{a}_B, \hat{F}).$ 

This is equivalent to:

$$\theta_A^* \notin \operatorname*{argmax}_{\hat{\theta} \in \overline{\Theta}_A} \left[ \begin{array}{c} (\lambda + (1 - \lambda)p_A) \cdot \mathbb{E}_{y \sim F^{\bullet}(a_{AA}^*, a_{AA}^*)} \ln(\hat{f}(a_{AA}^*, \hat{a}_A)(y)) \\ + (1 - \lambda)(1 - p_A) \cdot \mathbb{E}_{y \sim F^{\bullet}(a_{AB}^*, a_{BA}^*)} \ln(\hat{f}(a_{AB}^*, \hat{a}_B)(y)) \end{array} \right]$$

Let this objective, as a function of  $\hat{\theta}$ , be denoted  $WL(\hat{\theta})$ . There exists  $\theta_A^{opt} = (a_A^{opt}, a_B^{opt}, f^{opt}) \in \overline{\Theta}_A$  and  $\delta, \epsilon > 0$  so that  $(1 - \delta)WL(\theta_A^{opt}) - 2\delta M - 3\epsilon > (1 - \delta)WL(\theta_A^*)$ . By assumption on the primitives, find open neighborhoods  $V^{opt}$  and  $V^*$  of  $\theta_A^{opt}, \theta_A^*$  respectively, so that for all  $a_i \in \mathbb{A}$ ,  $g \in \{A, B\}, y \in \mathbb{Y}, 1 - \epsilon \leq f^{opt}(a_i, a_g^{opt})(y)/\hat{f}(a_i, \hat{a}_g)(y) \leq 1 + \epsilon$ , for all  $\hat{\theta} = (\hat{a}_A, \hat{a}_B, \hat{f}) \in V^{opt}$ , and also  $1 - \epsilon \leq f^*(a_i, a_g^*)(y)/\hat{f}(a_i, \hat{a}_g)(y) \leq 1 + \epsilon$  for all  $\hat{\theta} = (\hat{a}_A, \hat{a}_B, \hat{f}) \in V^*$ . Also, by convergence of play in the populations, find  $T_1$  so that in all periods  $t \geq T_1, \alpha_{AA}^{(t)}(a_{AA}^*) \geq 1 - \delta$ .

For  $T_2 \geq T_1$ , consider a probability space defined by  $\Omega := (\{A, B\} \times \mathbb{A}^2 \times (\mathbb{Y})^{\mathbb{A}^2})^{\infty}$  that describes the randomness in an agent's learning process starting with period  $T_2 + 1$ . For a point  $\omega \in \Omega$  and each period  $T_2 + s$ ,  $s \geq 1$ ,  $\omega_s = (g, a_{-i,A}, a_{-i,B}, (y_{a_i,a_{-i}})_{(a_i,a_{-i})\in\mathbb{A}^2})$  specifies the group g of the matched opponent, the play  $a_{-i,A}, a_{-i,B}$  of hypothetical opponents from groups A and B, and the hypothetical consequence  $y_{a_i,a_{-i}}$  that would be generated for every pair of strategies  $(a_i, a_{-i})$  played. As notation, let  $opp(\omega, s)$ ,  $a_{-i,A}(\omega, s)$ ,  $a_{-i,B}(\omega, s)$ , and  $y_{a_i,a_{-i}}(\omega, s)$  denote the corresponding components of  $\omega_s$ . Define  $\mathbb{P}_{T_2}$  over this space in the natural way. That is, it is independent across periods, and within each period, the density (or probability mass function if  $\mathbb{Y}$  is finite) of  $\omega_s = (g, a_{-i,A}, a_{-i,B}, (y_{a_i,a_{-i}})_{(a_i,a_{-i})\in\mathbb{A}^2})$  is

$$m_g \cdot \alpha_{AA}^{(T_2+s)}(a_{-i,A}) \alpha_{BA}^{(T_2+s)}(a_{-i,B}) \cdot \prod_{(a_i,a_{-i}) \in \mathbb{A}^2} f^{\bullet}(a_i,a_{-i})(y_{a_i,a_{-i}}),$$

where  $m_g$  is the probability of *i* from group A being matched up against an opponent of group *g*, that is  $m_A = (\lambda + (1 - \lambda)p_A), m_B = (1 - \lambda)(1 - p_A).$ 

For  $\theta = (a_A^{\theta}, a_B^{\theta}, F^{\theta}) \in \overline{\Theta}_A$  with  $f^{\theta}$  the density of  $F^{\theta}, \omega \in \Omega$ , consider the process

$$\ell_{s}(\theta,\omega) := \frac{1}{s} \sum_{t=T_{2}+1}^{T_{2}+s} \ln(f^{\theta}(a_{AA}^{*}, a_{opp(\omega,t)}^{\theta})(y_{a_{AA}^{*}, a_{-i,opp(\omega,t)}(\omega,t)}(\omega,t)).$$

By choice of the neighborhood  $V^*$ ,

$$\begin{split} \limsup_{s} \sup_{\theta_{A} \in V^{*}} \ell_{s}(\theta_{A}, \omega) &\leq \epsilon + \frac{1}{s} \sum_{t=T_{2}+1}^{T_{2}+s} \ln(f^{*}(a^{*}_{AA}, a^{*}_{opp(\omega,t)})(y_{a^{*}_{AA}, a_{-i,opp(\omega,t)}(\omega,t)}(\omega, t)) \\ &\leq \epsilon + \frac{1}{s} \sum_{t=T_{2}+1}^{T_{2}+s} \frac{1_{\{a_{-i,opp(\omega,t)}(\omega,t)=a^{*}_{opp(\omega,t),A}\}} \cdot \ln(f^{*}(a^{*}_{AA}, a^{*}_{opp(\omega,t)})(y_{a^{*}_{AA}, a^{*}_{opp(\omega,t),A}}(\omega, t))}{(1-1_{\{a_{-i,opp(\omega,t)}(\omega,t)=a^{*}_{opp(\omega,t),A}\}}) \cdot M. \end{split}$$

Since  $T_2 \geq T_1$ , in every period t,  $\mathbb{P}_{T_2}(a_{-i,opp(\omega,t)}(\omega,t) = a^*_{opp(\omega,t),A}) \geq 1-\delta$ . Let  $(\xi_k)_{k\geq 1}$  a related stochastic process: it is i.i.d. such that each  $\xi_k$  has  $\delta$  chance to be equal to M,  $(1-\delta)m_A$  chance to be distributed according to  $\ln(f^*(a^*_{AA}, a^*_A)(y))$  where  $y \sim f^{\bullet}(a^*_{AA}, a^*_{AA})$ , and  $(1-\delta)m_B$  chance to be distributed according to  $\ln(f^*(a^*_{AB}, a^*_B)(y))$  where  $y \sim f^{\bullet}(a^*_{AB}, a^*_{BA})$ . By law of large numbers,  $\frac{1}{s} \sum_{k=1}^s \xi_k$  converges almost surely to  $\delta M + (1-\delta)WL(\theta^*_A)$ . By this comparison,  $\limsup_{\theta_A \in V^*} \ell_s(\theta_A, \omega) \leq \epsilon + \delta M + (1-\delta)WL(\theta^*_A) \mathbb{P}_{T_2}$ -almost surely. By a similar argument,  $\liminf_{s \in V^{opt}} \ell_s(\theta_A, \omega) \geq -\epsilon - \delta M + (1-\delta)WL(\theta^{opt}_A) \mathbb{P}_{T_2}$ -almost surely.

Along any  $\omega$  where we have both  $\limsup_s \sup_{\theta_A \in V^*} \ell_s(\theta_A, \omega) \leq \epsilon + \delta M + (1 - \delta)WL(\theta_A^*)$ and  $\liminf_s \inf_{\theta_A \in V^{opt}} \ell_s(\theta_A, \omega) \geq -\epsilon - \delta M + (1 - \delta)WL(\theta_A^{opt})$ , if  $\omega$  also leads to i always playing  $a_{AA}^*$  against group A and  $a_{AB}^*$  against group B in all periods starting with  $T_2 + 1$ , then the posterior belief assigns to  $V^*$  must tend to 0, hence  $\tilde{\mu}_i^{(t)} \not\rightarrow \mu_A^*$ . Starting from any length  $T_2$  history h, there exists a subset  $\hat{\Omega}_h \subseteq \Omega$  that leads to i not playing the EZ-SU strategy in at least one period starting with  $T_2 + 1$ . So conditional on h, the probability of  $\tilde{\mu}_i^{(t)} \rightarrow \mu_A^*$  is no larger than  $1 - \mathbb{P}_{T_2}(\hat{\Omega}_h)$ . The unconditional probability is therefore no larger than  $\mathbb{E}_h[1 - \mathbb{P}_{T_2}(\hat{\Omega}_h)]$ , where  $\mathbb{E}_h$  is taken with respect to the distribution of period  $T_2$ histories for i. But this term is also the probability of i playing non-EZ-SU action at least once starting with period  $T_2$ . Since there are finitely many actions and  $(\tilde{a}_{iA}^{(t)}, \tilde{a}_{iB}^{(t)}) \rightarrow (a_{AA}^*, a_{AB}^*)$ almost surely,  $\mathbb{E}_h[1 - \mathbb{P}_{T_2}(\hat{\Omega}_h)]$  tends to 0 as  $T_2 \rightarrow \infty$ . We have a contradiction as this shows  $\tilde{\mu}_i^{(t)} \not\rightarrow \mu_A^*$  with probability 1.

Now consider the foundation for EZs. Suppose Let  $\bar{K} < \infty$  be an upper bound on  $K_{g,g'}(a_i, a_{-i}; (a_A, a_B, F))$  across all  $g, g' \in \{A, B\}, a_i, a_{-i} \in \mathbb{A}, (a_A, a_B, F) \in \overline{\Theta}_g$ . Here  $\bar{K}$  is finite because  $\mathbb{A}$  is finite and  $K_{g,g'}$  is continuous in the extended parameter, which is from a compact domain. Let  $F_{\tau}^X(a_{-i}) \in \Delta(\mathbb{A})$  represent the distribution of ex-post signals given precision  $\tau$ , when opponent plays  $a_{-i} \in \mathbb{A}$ . It is clear that there exists some  $\underline{\tau} < 1$  so that for any  $a_{-i} \neq a'_{-i}, \tau \in (\underline{\tau}, 1)$ , we get  $\min(m_A, m_B) \cdot D_{KL}(F_{\tau}^X(a_{-i}) \parallel F_{\tau}^X(a'_{-i})) > \overline{K}$ . Therefore, given any  $(a_{AA}^*, a_{AB}^*, a_{BA}^*) \in \mathbb{A}^3$ , the solution to

$$\min_{\hat{\theta}\in\overline{\Theta}_{A}} \left[ \begin{array}{c} (\lambda + (1-\lambda)p_{A}) \cdot \left[ D_{KL}(F^{\bullet}(a_{AA}^{*}, a_{AA}^{*}) \parallel \hat{F}(a_{AA}^{*}, \hat{a}_{A})) + D_{KL}(F_{\tau}^{X}(a_{AA}^{*}) \parallel F_{\tau}^{X}(\hat{a}_{A})) \right] \\ + (1-\lambda)(1-p_{A}) \cdot \left[ D_{KL}(F^{\bullet}(a_{AB}^{*}, a_{BA}^{*}) \parallel \hat{F}(a_{AB}^{*}, \hat{a}_{B})) + D_{KL}(F_{\tau}^{X}(a_{BA}^{*}) \parallel F_{\tau}^{X}(\hat{a}_{B})) \right] \end{array} \right]$$

must satisfy  $\hat{a}_A = a_{AA}^*$ ,  $\hat{a}_B = a_{BA}^*$ , because  $(a_{AA}^*, a_{BA}^*, F)$  for any  $F \in \Theta_A$  has a KL divergence no larger than  $\bar{K}$ . On the other hand, any  $(\hat{a}_A, \hat{a}_B, \hat{F})$  with either  $\hat{a}_A \neq a_{AA}^*$  or  $\hat{a}_B \neq a_{BA}^*$  has KL divergence strictly larger than  $\bar{K}$  by the choice of  $\tau$ . The rest of the argument is similar to the case of EZ-SU.

## C.4 Multiple Situations

Now suppose there are multiple situations  $G \in \mathcal{G}$  and a distribution  $q \in \Delta(\mathcal{G})$ , with  $\mathcal{G}$  finite. At the start of period t = 1, Nature draws a situation  $G^{(1)}$  from  $\mathcal{G}$  according to q, and consequences are generated according to  $F^{\bullet}(\cdot, \cdot, G^{(1)})$  until period t = T + 1. In period T + 1, Nature again draws a situation  $G^{(2)}$  from  $\mathcal{G}$  according to q, and consequences are generated according to  $F^{\bullet}(\cdot, \cdot, G^{(2)})$  until period t = 2T + 1, and so forth. Agents start with a prior over their group's extended model,  $\mu_g^{(0)} \in \Delta(\overline{\Theta}_g)$ . In periods  $T + 1, 2T + 1, \ldots$  agents reset their belief to  $\mu_g^{(0)}$ , and their belief in each period over the extended parameters in their extended model only use histories since the last reset. This belief corresponds to agents thinking that the data-generating process is redrawn according to  $\mu_g^{(0)}$  every T periods.

Suppose  $\tau = 0$  and suppose for every  $G \in \mathcal{G}$ , the hypotheses of Proposition A.3 hold in a society where G is the only true situation. Denote  $(a_{AA}^*(G), a_{AB}^*(G), a_{BA}^*(G), a_{BB}^*(G), \mu_A^*(G), \mu_B^*(G))$ as the limit of the agents' behavior and beliefs with situation G. Then it is straightforward to see that in a society with the situation redrawn every T periods, the expected undiscounted average payoff of an agent in group g approaches the fitness of g in the EZ-SU characterized by the behavior and beliefs  $(a_{AA}^*(G), a_{AB}^*(G), a_{BA}^*(G), a_{BB}^*(G), \mu_A^*(G), \mu_B^*(G))_{G \in \mathcal{G}}$  with the distribution q over situations, as  $T \to \infty$ . This provides a foundation for fitness in EZ-SU as the agents' objective payoffs when the true situation changes sufficiently slowly (a similar foundation applies for the fitness in EZ.)