Compatible Explanations

Youngji Sohn*

October 12, 2026

Click for the latest version

Abstract

Berk–Nash equilibrium allows players to hold misspecified views of their environment, requiring optimal actions under beliefs supported on hypotheses that minimize Kullback–Leibler divergence. These conditions concern each player's own statistical decision problem: a best-fitting hypothesis may nevertheless conjecture an opponent strategy that cannot be justified by the opponent's inference and incentives. This paper develops a refinement motivated by players' introspection under common knowledge of the game structure and players' subjective model classes. A hypothesis about Nature and the opponent's behavior, together with the player's own strategy, determines the observations the opponent would receive if the hypothesis were correct. Rationalizable Berk–Nash equilibrium requires each player to have a supporting belief such that every hypothesis in its support conjectures an opponent strategy optimal under some belief supported on the opponent's best-fitting hypotheses for those hypothetical observations. The paper characterizes rationalizable Berk–Nash equilibria through families of compatible explanations. The concept extends the recursive strategic justification underlying rationalizable conjectural equilibrium (RCE) from exact observational consistency to best available statistical fit. It coincides with the corresponding RCE benchmark when players share a model class containing the true state distribution. If both model classes contain the truth, every Nash equilibrium of the objective game survives, even when the classes differ. With misspecified models, false explanations can remain recursively justifiable despite positive fitting losses in both actual and hypothetical problems.

Keywords: Berk–Nash equilibrium, misspecified games, identification, conjectures, rationality, rationalizable conjectural equilibrium.
JEL: C72, D83.

*Department of Economics, University of Warwick. youngji.sohn@warwick.ac.uk.

1 Introduction

In games with incomplete information, players' choices depend on beliefs about both the environment and opponents' strategies. When equilibrium is interpreted as a steady state of repeated play, these beliefs require a justification grounded in what players can learn from their observations. The common-prior formulation of Bayes–Nash equilibrium combines a shared prior about Nature with correct conjectures about opponents' strategies. As Dekel et al. (2004) argue, allowing players to disagree about Nature while maintaining correct strategic conjectures is difficult to justify through learning. Observations sufficiently informative to reveal opponents' strategies may also reveal information about Nature, while limited feedback may leave mistaken beliefs about either undetected. An account of steady-state play must therefore explain how beliefs about both Nature and opponents' strategies are jointly sustained by the observations that play generates.

With this regard, the literature has studied equilibrium concept as understood as a steady state of learning in repeated play. In a self-confirming equilibrium, each player's strategy is optimal under beliefs about Nature and opponents' strategies that reproduce the distribution of her observations following the actions she plays (Battigalli, 1987; Fudenberg and Levine, 1993; Dekel et al., 2004). These beliefs may be mistaken when different explanations generate the same observations. Berk–Nash equilibrium provides a general framework that accommodates self-confirming equilibrium as a special case and allows players' subjective models to be misspecified (Esponda and Pouzo, 2016). Each player's subjective model consists of hypotheses describing how consequences are generated conditional on her actions and may exclude the true data-generating process. A strategy profile is a Berk–Nash equilibrium if each player's strategy is optimal under a belief over her hypotheses that assigns probability only to those that best account for her observations at that profile. The best available hypotheses need not reproduce those observations exactly.

These requirements concern each player's own statistical decision problem. A player's equilibrium belief may nevertheless include a conjecture about the opponent's strategy that cannot be justified by the opponent's own inference and incentives. This paper develops a refinement of Berk–Nash equilibrium that uses reasoning about opponents' rationality to restrict its supporting beliefs.

Example 1 (Investment)**.** Consider a simple game with no uncertainty about Nature. Player 1 invests, I, or takes an outside option, O, while player 2 assists, A, or blocks, B. Suppose each player considers every strategy of her opponent possible and, after play,

observes only her own action and payoff:

	A	B
I	$(2,1)$	$(0,0)$
O	$(1,1)$	$(1,0)$

Assistance strictly dominates blocking for player 2, and investment is player 1's unique best response to assistance. Thus, (I, A) is the unique Nash equilibrium and a Berk–Nash equilibrium, supported by correct conjectures that reproduce observations exactly. Moreover, player 2 assists in every Berk–Nash equilibrium. The profile (O, A) is also a Berk–Nash equilibrium. When player 1 takes the outside option, she receives one regardless of player 2's action and cannot distinguish A from B. Any conjecture assigning probability at least $1/2$ to blocking therefore reproduces her observations and makes her choice optimal. Yet every belief supporting the outside option must assign positive probability to blocking, which is never optimal for player 2. If player 1 knows player 2's payoff structure and regards her as rational, however, she can rule out blocking without observing player 2's action. She must then expect assistance and strictly prefer investing, ruling out (O, A).

Rationalizability as a selection device The example motivates using the opponent's rationality to discipline conjectures that otherwise fit a player's observations and justify her choice. We extend this logic of rationalizable conjectures (Rubinstein and Wolinsky, 1994; Battigalli and Guaitoli, 1997; Esponda, 2013) to games with incomplete information and possibly misspecified subjective models.

We call a strategy profile a *rationalizable Berk–Nash equilibrium* if it is a Berk–Nash equilibrium for which each player has a supporting belief satisfying the following recursive requirement. Each hypothesis in its support must conjecture a strategy that is optimal for the opponent under some belief over her own hypotheses. That belief must be supported on hypotheses that best account, within the opponent's subjective model, for the observations she would receive if the original hypothesis were true. The same requirement applies to every hypothesis in the support of this further belief, at every subsequent depth.

The hypotheses in the support of a Berk–Nash supporting belief account equally well for the player's observations at the equilibrium profile. The additional restriction is motivated by the player regarding the opponent as Bayesian rational under each of these possible explanations. Every hypothesis used in her belief must therefore accommodate the opponent's rationality. Imposing the restriction at every depth is motivated by common belief in Bayesian rationality: each player regards the opponent as Bayesian rational, regards the opponent as holding the same view of her, and so on.

Example 2 (Cournot game with uncertain demand)**.** Consider now a game in which payoffs depend on an uncertain state of Nature. Two firms choose quantities $q_i \in 1, 2$.

Market demand is low or high, each with true probability 1/2. Writing $Q = q_1 + q_2$, market prices are $P_\ell(Q) = 15 - 3Q$ and $P_h(Q) = 19 - 3Q$. Both firms have marginal cost 3, and firm i earns $q_i(P - 3)$. Neither firm observes the demand state before choosing its quantity. After play, each observes only its own quantity and profit, which reveals the market price. Firm 1 considers only a high-demand probability of 2/3, whereas firm 2 considers probabilities 1/2 and 4/5. Both consider every strategy of their rival possible.

$(1, 2)$ and $(2, 1)$ are Berk–Nash equilibria. At either profile, both firms observe prices 6 and 10 with equal frequency. Their unique best-fitting hypotheses correctly conjecture the rival's quantity, with firm 1 assigning probability 2/3 to high demand and firm 2 assigning probability 1/2. Under these respective beliefs, each firm prefers producing two units against one and one unit against two. Thus, both profiles satisfy statistical fit and optimality, although firm 1's hypothesis predicts the observed price frequencies incorrectly.

At $(2, 1)$, however, firm 1 cannot justify the rival behavior its hypothesis conjectures. If high demand occurred with probability 2/3 and firm 1 produced two units, firm 2 would observe prices 6 and 10 with probabilities 1/3 and 2/3. Firm 2 cannot represent this demand frequency within its model. Its unique best-fitting hypothesis would assign probability 4/5 to high demand, which fits the hypothetical frequency better than 1/2, and correctly conjecture that firm 1 produces two. Under this belief, firm 2 would strictly prefer producing two units, earning 32/5 rather than 31/5 from producing one. Firm 1's only supporting hypothesis therefore fails to justify its conjecture, ruling out $(2, 1)$.

By contrast, $(1, 2)$ admits justification at every depth. Firm 1's hypothesis assigns probability 2/3 to high demand and conjectures that firm 2 produces two. Since firm 1 produces one, firm 2's best-fitting hypothesis for these hypothetical observations uses probability 4/5 and correctly conjectures quantity one, making its own quantity two optimal. This hypothesis, in turn, generates observations for firm 1 whose best-fitting hypothesis again uses 2/3 and conjectures quantity two, making its own quantity one optimal. These hypotheses therefore justify each other recursively. Firm 2's actual supporting hypothesis, using probability 1/2, also leads to firm 1's same hypothesis and hence into this continuing justification. Thus, $(1, 2)$ is a rationalizable Berk–Nash equilibrium: firm 1 remains mistaken about demand, but the rival behavior attributed by its hypothesis can be justified through the rival's own observations, inference, and incentives.

Knowledge assumptions. We maintain common knowledge of the payoff structure and each player's observation rule and subjective model. This is a demanding assumption: players must understand both the opponent's incentives and the restrictions on her interpretation of observations. It nevertheless leaves room for uncertainty and disagreement about Nature. The uncertain state can affect both the consequences of play and what observations reveal, so common knowledge of the observation rule need not reveal what

the opponent observes or infers. Likewise, her subjective model specifies which hypotheses she considers possible without determining her belief over those hypotheses or requiring the other player to know that belief.

Main results Our first result characterizes rationalizable Berk–Nash equilibrium through families of compatible explanations. A family of explanations is compatible if, for every explanation, there is a belief for the opponent that justifies the conjectured opponent strategy and generates only further explanations within the family. A strategy profile is a rationalizable Berk–Nash equilibrium if and only if beliefs supporting it as a Berk–Nash equilibrium can be chosen so that all explanations generated by each player's belief belong to a compatible family, which may differ across players. The characterization generalizes the loop of reasoning illustrated in the second example.

In studying existence, it is useful to distinguish two properties of a player's subjective model: containing the true state distribution and being able to reproduce her observations. Nash equilibrium provides a useful benchmark: when both players' subjective models contain the true state distribution, every Nash equilibrium of the objective game survives the refinement, guaranteeing existence. Correct hypotheses about Nature and the opponent's strategy generate explanations that justify one another in a loop of reasoning of length two. Yet rationalizable Berk–Nash equilibrium may fail to exist even with a common subjective model that reproduces each player's consequence distributions at every strategy profile. Exact fit may require conjectures about the opponent that her own inference and incentives cannot justify. Statistical accuracy alone therefore does not ensure existence.

When models exclude the truth, one sufficient construction uses Nash equilibria evaluated under possibly different state distributions in the players' respective models. Such a strategy profile survives if each player's corresponding hypothesis best accounts for both her actual observations and the hypothetical observations generated by the opponent's hypothesis. The general characterization also allows longer loops and branching families of explanations, with beliefs and conjectures varying across successive hypothetical problems.

Under an all-or-nothing restriction on feedback, the statistical requirements concerning Nature and opponent behavior separate. For a pair of state distributions in the players' respective subjective models, every Nash equilibrium evaluated under that pair survives the refinement if each distribution best accounts, within the corresponding player's model, for both the true distribution and the other distribution. In particular, a common subjective model guarantees existence under this feedback restriction even when it excludes the true state distribution.

1.1 Related Literature

Rationalizable conjectures and related reasoning. The closest precedent for this use of introspection is rationalizable conjectural equilibrium (RCE), developed by Rubinstein and Wolinsky (1994) and extended to uncertainty about fundamentals by Esponda (2013). In Esponda's formulation, players share a common model containing the true fundamental. RCE requires each player's strategy to be optimal under a belief consistent with her observations and applies these requirements recursively to conjectured behavior. When both players share a common model containing the true state distribution, rationalizable Berk–Nash equilibrium coincides with the corresponding RCE benchmark for the same strategies and observations. With different models containing the truth, it includes the RCE profiles defined on their intersection. More generally, rationalizable Berk–Nash equilibrium permits beliefs supported on hypotheses that best account for actual or hypothetical observations without requiring exact consistency.

The same recursive reasoning appears in rationalizable partition-confirmed equilibrium for extensive-form games, which combines observationally consistent beliefs with optimality conditional on the information available when a player acts (Fudenberg and Kamada, 2015, 2018). Another related paper is Battigalli and Siniscalchi (2003), which develops rationalizability concepts for games with incomplete information. Their Δ-rationalizability imposes exogenous restrictions on the beliefs used to rationalize behavior; in ζ-rationalizability, these restrictions derive from public statistics on payoff types and terminal histories generated by repeated play.

Strategic inference under misspecification. Strategic reasoning under misspecification has also been studied in specific contexts. He and Libgober (2025) study learning from prices under misperceived signal correlation, with an extension in which entrants jointly infer opponents' strategies and persistent market parameters. Frick et al. (2022) study agents who neglect assortativity, treating the behavior observed among their peers as representative of society. These agents infer the population's type distribution by explaining observed behavior as optimal choices in a perceived nonassortative society. Their Appendix C.A allows agents to attribute different perceptions to others, provided those perceptions are consistent with the observations those agents are imagined to receive and make their conjectured behavior optimal. This generalization is closest in spirit to our approach. We study a general refinement of Berk–Nash equilibrium based on such reasoning, without imposing a particular behavioral bias beyond the restrictions encoded in the players' subjective models.

The broader distinction between accurate prediction and correct understanding also arises in single-agent settings. For example, Heidhues et al. (2025) study an agent who persistently underestimates her own present bias and interprets her past choices as evidence about the benefits or costs of consumption. In a stable environment, she can learn to

predict her behavior accurately while continuing to misunderstand the motives driving it.

Analogy-based expectation equilibrium. Analogy-based expectation equilibrium restricts forecasts of opponents' behavior through observations pooled within analogy classes (Jehiel, 2005; Jehiel and Koessler, 2008). Players forecast average behavior within each class and choose optimally given those forecasts. Its learning interpretation does not require knowledge of opponents' payoffs or information structures.[1] Our question is whether the conjectured opponent strategy can be optimal under a belief over hypotheses in the opponent's own model. This belief must assign probability only to hypotheses that best account for the observations she would receive.

The remainder of the paper is organized as follows. Section 2 introduces the model and the Berk–Nash benchmark. Section 3 defines rationalizable Berk–Nash equilibrium recursively and characterizes it through families of compatible explanations. Section 4 presents the existence and nonexistence results. Section 5 compares the refinement with other solution concepts.

2 The Framework and Berk–Nash Equilibrium

2.1 Game

The definition of a game largely follows the framework of Esponda and Pouzo (2016), specialized to two players who receive no private information before choosing their actions. A simultaneous-move game $\mathcal{G} = \langle \mathcal{O}, \mathcal{Q} \rangle$ consists of an objective game $\mathcal{O}$ and a subjective model $\mathcal{Q}$.

Objective game. An objective game is a tuple $\mathcal{O} = \langle I, \Omega, p, X, Y, f, u \rangle$. $I = \{1, 2\}$ is the set of players. Throughout, i and j denote distinct players. Ω is the finite set of payoff-relevant states, and p is a probability distribution over Ω with full support. $X = X^1 \times X^2$ denotes the set of action profiles, where X^i is the finite action set of player i. $Y = Y^1 \times Y^2$ denotes the set of consequence profiles, where Y^i is the finite consequence set of player i. $f = (f^1, f^2)$ is the profile of feedback functions, with $f^i : X^i \times X^j \times \Omega \to Y^i$ giving player i's consequence.[2] Finally, $u = (u^1, u^2)$ is the profile of payoff functions, with $u^i : X^i \times Y^i \to \mathbb{R}$ for player i. Let $\mathcal{S} = \langle I, \Omega, X, Y, f, u \rangle$ denote the underlying game structure, consisting of all components of the objective game except the state distribution p.

A state $\omega \in \Omega$ is drawn according to p. The players then simultaneously choose actions $x^i \in X^i$ without receiving a signal about the state. Write $x = (x^1, x^2) \in X$ for the

[1] Jehiel and Weber (2026) provide a steady-state learning foundation for an extension of ABEE in which analogy partitions are endogenously chosen to minimize prediction errors about opponents' behavior, subject to a bound on the number of classes.

[2] Any additional randomness in feedback can be incorporated into the state space Ω.

resulting action profile. Finally, player i observes her consequence $y^i = f^i(x^i, x^j, \omega) \in Y^i$ and receives payoff $u^i(x^i, y^i)$. Apart from her own action, y^i is her only feedback following the outcome (ω, x).

A strategy of player i is $\sigma^i \in \Sigma^i \equiv \Delta(X^i)$. Let $\Sigma = \Sigma^1 \times \Sigma^2$ and denote a strategy profile by $\sigma = (\sigma^1, \sigma^2)$.

At a strategy profile σ, the objective distribution of player i's consequences after her action x^i is

$$Q^i_\sigma(y^i \mid x^i) = \sum_{\omega \in \Omega} \sum_{x^j \in X^j} p(\omega)\sigma^j(x^j)\mathbf{1}\{f^i(x^i, x^j, \omega) = y^i\}$$

for all $(x^i, y^i) \in X^i \times Y^i$.

Subjective models. The subjective model is a tuple $\mathcal{Q} = \langle \Theta, (Q_\theta)_{\theta \in \Theta} \rangle$. Here $\Theta = \Theta^1 \times \Theta^2$, where Θ^i is player i's parameter set. For $\theta = (\theta^1, \theta^2) \in \Theta$, write $Q_\theta = (Q^i_{\theta^i})_{i \in I}$, where $Q^i_{\theta^i} : X^i \to \Delta(Y^i)$ specifies the conditional distribution of player i's consequences under parameter $\theta^i \in \Theta^i$; we write $Q^i_{\theta^i}(\cdot \mid x^i)$ for this distribution following action x^i. The objective game describes the true environment, while the subjective model specifies the conditional distributions of consequences that each player considers possible.

Following the structural approach in Remark 1 of Esponda and Pouzo (2016), player i's subjective parameter set is $\Theta^i \equiv \Pi^i \times \Sigma^j$, where $\Pi^i \subseteq \Delta(\Omega)$ is the set of state distributions that player i considers possible. By definition, player i may rule out some state distributions, but she considers every strategy of player j possible. We write a parameter as $\theta^i = (\pi^i, \widehat{\sigma}^j) \in \Theta^i$.[3] Under this parameter, player i conjectures that Nature draws the state according to π^i and player j uses $\widehat{\sigma}^j$.

The subjective consequence distribution parametrized by $\theta^i = (\pi^i, \widehat{\sigma}^j)$ is

$$Q^i_{\theta^i}(y^i \mid x^i) = \sum_{\omega \in \Omega} \sum_{x^j \in X^j} \pi^i(\omega)\widehat{\sigma}^j(x^j)\mathbf{1}\{f^i(x^i, x^j, \omega) = y^i\}.$$

For each parameter $(\pi^i, \widehat{\sigma}^j)$, the state and the opponent's action are independent, consistent with the absence of information about the state before actions are chosen. For any $\rho \in \Delta(\Omega)$ and $\tau^j \in \Sigma^j$, write Q^i_{ρ,τ^j} for the same expression with (ρ, τ^j) in place of $(\pi^i, \widehat{\sigma}^j)$.

Two regularity conditions on the parameter set, adapted from Assumption 1 of Esponda and Pouzo (2016), are maintained throughout.

Assumption 1 (Regularity conditions)**.** For every player $i \in I$:

(i) Π^i is nonempty and compact;

(ii) for every $\theta^i \in \Theta^i$, there exists a sequence $(\theta^i_n)_n \subseteq \Theta^i$ converging to θ^i such that for all n, $Q^i_{\theta^i_n}(y^i \mid x^i) > 0$ for all $x^i \in X^i$ and for all $y^i \in Y^i$ satisfying $y^i = f^i(x^i, x^j, \omega)$ for some $(x^j, \omega) \in X^j \times \Omega$.

[3] More explicitly, the parameter can be written as $(\pi^i_{\theta^i}, \widehat{\sigma}^j_{\theta^i})$. For notational convenience, we suppress the subscript θ^i and write $(\pi^i, \widehat{\sigma}^j)$.

The subjective consequence distributions are continuous in θ^i because they are bilinear in $(\pi^i, \widehat{\sigma}^j)$. Condition (ii) requires every parameter to be approximated arbitrarily closely by parameters that assign positive probability to every feasible consequence following each action. In their framework, Esponda and Pouzo (2016, Section 5) show that without this condition, Berk–Nash equilibrium may fail to exist and steady-state behavior need not be characterized by Berk–Nash equilibrium.

2.2 Berk–Nash equilibrium

In a Berk–Nash equilibrium (Esponda and Pouzo, 2016), each player's strategy is optimal under a belief over her subjective parameters where the belief is supported on parameters whose associated subjective distributions are "closest" to the objective distribution given the strategy profile. We define these two requirements below.

Statistical fit. The distance between the objective distribution over i's consequences under $\sigma \in \Sigma$ and the subjective distribution parametrized by $\theta^i \in \Theta^i$ is measured by *weighted Kullback–Leibler divergence* (wKLD)[4]

$$K^i(\sigma, \theta^i) = \sum_{x^i \in X^i} \sum_{y^i \in Y^i} \sigma^i(x^i) Q^i_\sigma(y^i \mid x^i) \ln \frac{Q^i_\sigma(y^i \mid x^i)}{Q^i_{\theta^i}(y^i \mid x^i)}.$$

The parameters that minimize player i's wKLD form the set[5]

$$\Theta^{i,0}(\sigma) \equiv \arg\min_{\theta^i \in \Theta^i} K^i(\sigma, \theta^i).$$

The interpretation is that $\theta^i \in \Theta^{i,0}(\sigma)$ is a parameter that player i can consider possible after observing consequences consistent with strategy profile σ. The wKLD function compares the objective and subjective consequence distributions only following actions that player i plays with positive probability. It adds no restrictions on the subjective distributions following unplayed actions beyond those imposed by Θ^i (Esponda and Pouzo, 2016, Remark 3).

Optimality. The expected payoff from action $x^i \in X^i$ under the subjective distribution parametrized by θ^i is

$$U^i(x^i; \theta^i) = \sum_{y^i \in Y^i} Q^i_{\theta^i}(y^i \mid x^i) u^i(x^i, y^i).$$

[4] wKLD is equivalent to the standard Kullback–Leibler divergence applied to distributions over (x^i, y^i) generated by σ^i with Q^i_σ and by σ^i with $Q^i_{\theta^i}$. We use the conventions $0 \ln(0/b) = 0$ for every $b \geq 0$, $a \ln(a/0) = +\infty$ for every $a > 0$, and $0 \cdot (+\infty) = 0$.

[5] Under Assumption 1, $K^i(\sigma, \cdot)$ attains a finite minimum and $\Theta^{i,0}(\sigma)$ is nonempty and compact at every profile σ.

Player i's strategy σ^i is optimal given a belief $\mu^i \in \Delta(\Theta^i)$ if every action x^i with $\sigma^i(x^i) > 0$ satisfies[6]

$$x^i \in \arg\max_{\tilde{x}^i \in X^i} \mathbb{E}_{\mu^i}\left[U^i(\tilde{x}^i; \theta^i)\right].$$

Optimality compares expected payoffs from all actions, including those that are not played. Denote the set of i's optimal strategies given μ^i by $\mathrm{BR}^i(\mu^i)$.

Definition 1 (Berk–Nash equilibrium, Esponda and Pouzo, 2016)**.** A strategy profile $\sigma \in \Sigma$ is a Berk–Nash equilibrium if, for every player $i \in I$, there exists a belief $\mu^i \in \Delta(\Theta^i)$ such that

$$\mu^i\left(\Theta^{i,0}(\sigma)\right) = 1 \qquad \text{and} \qquad \sigma^i \in \mathrm{BR}^i(\mu^i).$$

We call a belief satisfying the two conditions in Definition 1 a supporting belief for player i at σ. Let $\mathcal{B}^{i,0}(\sigma) \equiv \left\{\mu^i \in \Delta(\Theta^i) : \mu^i\left(\Theta^{i,0}(\sigma)\right) = 1,\ \sigma^i \in \mathrm{BR}^i(\mu^i)\right\}$ denote the set of these beliefs. The set of Berk–Nash equilibria is $\mathcal{E}^0 \equiv \{\sigma \in \Sigma : \mathcal{B}^{i,0}(\sigma) \neq \varnothing \text{ for every } i \in I\}$.

Existence. Under Assumption 1, the existence result of Esponda and Pouzo (2016, Theorem 1) applies, so $\mathcal{E}^0 \neq \varnothing$.

Remark 1 (Bayesian rationality in steady state). Under repeated play at a fixed strategy profile, parameters with lower wKLD have higher expected log likelihood for the player's observations. This connects statistical fit to the likelihood comparisons underlying Bayesian updating.[7] Given a belief supported on the wKLD minimizers, choosing a best response to the belief is expected-payoff maximization. Therefore, the statistical-fit and optimality requirements in Definition 1 describe the beliefs and strategy of a *Bayesian rational* player in steady state.

Remark 2 (Role of the structural specification). Berk–Nash equilibrium depends on each parameter only through its induced consequence distributions. Our formulation specifies these distributions through a state distribution and an opponent strategy, without adding an equilibrium condition beyond statistical fit and optimality. In Section 3, we use these structural components to ask whether $\widehat{\sigma}^j$ can be justified as a Bayesian rational strategy for player j when Nature draws the state according to π^i and player i uses σ^i.

[6] Although the state and the opponent's action are independent under each parameter, player i's belief can produce dependence.

[7] Esponda and Pouzo (2016, Remark 2) motivate wKLD through a likelihood argument. They develop this connection in a learning model where players update beliefs over their subjective parameters by Bayes' rule and maximize current expected payoff under those beliefs. Under their Assumption 1 and the learning setup of Section 4.2, including full-support priors over Θ^i, Lemma 2 shows that, if behavior converges to a strategy profile σ, each player i's beliefs become increasingly concentrated on $\Theta^{i,0}(\sigma)$. The posterior belief itself need not converge.

3 Rationalizable Berk–Nash equilibrium

The structural specification allows us to interpret each parameter $(\pi^i, \widehat{\sigma}^j)$ as a hypothesis player i considers: Nature draws the state according to π^i, and player j uses $\widehat{\sigma}^j$. Together with player i's strategy σ^i, this hypothesis provides an explanation $(\pi^i, \sigma^i; \widehat{\sigma}^j)$ of how both players' observations would be generated. Under this explanation, player j faces the state distribution π^i and player i's strategy σ^i. We call (π^i, σ^i) player j's *hypothetical problem*. We ask whether the conjectured strategy $\widehat{\sigma}^j$ is optimal under some belief over player j's subjective parameters that is supported on parameters minimizing her wKLD in the hypothetical problem (π^i, σ^i) when she uses $\widehat{\sigma}^j$.

To interpret these requirements as players' reasoning about each other's strategies, we assume that the underlying game structure $\mathcal{S}$ and the subjective model $\mathcal{Q}$ are common knowledge. We do not assume that either player knows p. Common belief in the players' Bayesian rationality motivates imposing the statistical and optimality requirements recursively in successive hypothetical problems, leading to rationalizable Berk–Nash equilibrium.

3.1 The hypothetical problem

Consider player j's hypothetical problem $h^j = (\rho, \tau^i) \in \Delta(\Omega) \times \Sigma^i$, where Nature draws the state according to ρ and player i uses τ^i. Following action x^j, player j's consequence has distribution $Q^j_{\rho,\tau^i}(\cdot \mid x^j)$.

To test a candidate strategy $\tau^j \in \Sigma^j$, we compare these consequence distributions with those in player j's own subjective model, weighting the comparison following each action x^j by $\tau^j(x^j)$. For $\theta^j \in \Theta^j$, define player j's wKLD in the hypothetical problem by

$$K^j_\rho(\tau; \theta^j) = \sum_{x^j \in X^j} \tau^j(x^j) \sum_{y^j \in Y^j} Q^j_{\rho,\tau^i}(y^j \mid x^j) \ln \frac{Q^j_{\rho,\tau^i}(y^j \mid x^j)}{Q^j_{\theta^j}(y^j \mid x^j)},$$

where $\tau = (\tau^1, \tau^2)$. The parameters that minimize player j's wKLD form the set $\Theta^j_\rho(\tau) = \arg\min_{\theta^j \in \Theta^j} K^j_\rho(\tau; \theta^j)$.[8]

Definition 2. For every $\rho \in \Delta(\Omega)$ and $\tau^i \in \Sigma^i$, define player j's Berk–Nash response correspondence[9]

$$\mathcal{R}^{j,0}(\rho, \tau^i) = \left\{ \tau^j \in \Sigma^j : \exists \nu^j \in \Delta(\Theta^j) \text{ such that } \nu^j\left(\Theta^j_\rho(\tau)\right) = 1 \text{ and } \tau^j \in \mathrm{BR}^j(\nu^j) \right\}.$$

[8] Under Assumption 1, $\Theta^j_\rho(\tau)$ is nonempty and compact; see Appendix A.1.

[9] Optimality under a belief over opponents' strategies is familiar from rationalizability (Bernheim, 1984; Pearce, 1984). Here, we require player j's belief over Θ^j to be supported on $\Theta^j_\rho(\tau)$. For related restrictions on rationalizing beliefs based on introspection, see Rubinstein and Wolinsky (1994); Battigalli and Siniscalchi (2003).

At a strategy profile σ, the hypothetical problem (p, σ^j) coincides with player i's actual problem, so $\Theta_p^i(\sigma) = \Theta^{i,0}(\sigma)$. Thus, σ is a Berk–Nash equilibrium if and only if each player's strategy is a Berk–Nash response to her opponent's strategy under the objective state distribution p, that is, $\sigma^i \in \mathcal{R}^{i,0}(p, \sigma^j)$ for every $i \in I$.

Lemma 1. *Under Assumption 1, for every pair of distinct players i and j, every $\rho \in \Delta(\Omega)$, and every $\tau^i \in \Sigma^i$, $\mathcal{R}^{j,0}(\rho, \tau^i) \neq \varnothing$.*

The proof is in Appendix A.2. The lemma guarantees that in every hypothetical problem there exists at least one Berk–Nash response, but the conjectured strategy $\widehat{\sigma}^j$ in a particular explanation need not be one.

3.2 The first cross-player restriction

Fix a profile $\sigma \in \Sigma$. For each player i, we restrict her belief to parameters in $\Theta^{i,0}(\sigma)$ for which $\widehat{\sigma}^j$ is a Berk–Nash response to σ^i under the state distribution π^i. These parameters form the set $\Theta^{i,1}(\sigma)$.[10]

$$\Theta^{i,1}(\sigma) = \left\{(\pi^i, \widehat{\sigma}^j) \in \Theta^{i,0}(\sigma) : \widehat{\sigma}^j \in \mathcal{R}^{j,0}(\pi^i, \sigma^i)\right\}.$$

Let $\mathcal{B}^{i,1}(\sigma)$ be the set of beliefs over Θ^i that are supported on $\Theta^{i,1}(\sigma)$ and under which σ^i is optimal. $\mathcal{B}^{i,1}(\sigma) = \left\{\mu^i \in \mathcal{B}^{i,0}(\sigma) : \mu^i\big(\Theta^{i,1}(\sigma)\big) = 1\right\}$.

Definition 3. A profile $\sigma \in \Sigma$ is a *depth-1 rationalizable Berk–Nash equilibrium* if $\mathcal{B}^{i,1}(\sigma) \neq \varnothing$ for every player i. Let $\mathcal{E}^1$ denote the set of these equilibria.

Since $\mathcal{B}^{i,1}(\sigma) \subseteq \mathcal{B}^{i,0}(\sigma)$ for every player i, we have $\mathcal{E}^1 \subseteq \mathcal{E}^0$.

Remark 3 (Possible failure of existence). Fix a Berk–Nash equilibrium $\sigma \in \mathcal{E}^0$. By Lemma 1, every hypothetical problem (π^i, σ^i) has at least one Berk–Nash response. This does not guarantee that there is a parameter $(\pi^i, \widehat{\sigma}^j) \in \Theta^{i,0}(\sigma)$ for which $\widehat{\sigma}^j \in \mathcal{R}^{j,0}(\pi^i, \sigma^i)$. If no such parameter exists, $\Theta^{i,1}(\sigma)$ is empty. Even when $\Theta^{i,1}(\sigma)$ is nonempty, σ^i need not be optimal under any belief supported on this set. Either failure makes $\mathcal{B}^{i,1}(\sigma)$ empty and excludes σ from $\mathcal{E}^1$. If every Berk–Nash equilibrium has at least one player for whom either failure occurs, then $\mathcal{E}^1 = \varnothing$.

3.3 Rationalizable Berk–Nash equilibrium

The first restriction requires the conjectured strategy $\widehat{\sigma}^j$ to be optimal under a belief over Θ^j supported on $\Theta^j_{\pi^i}(\sigma^i, \widehat{\sigma}^j)$. Parameters in this belief's support may specify strategies for player i that are not Berk–Nash responses in the corresponding hypothetical problems. We now impose that requirement recursively.

[10] Under Assumption 1, $\Theta^{i,1}(\sigma)$ is compact. See Appendix A.4 for the proof.

Definition 4. Starting from $\mathcal{R}^{i,0}$ in Definition 2, define the response correspondences recursively as follows. For every integer $n \geq 0$, every pair of distinct players i and j, every $\rho \in \Delta(\Omega)$, and every $\tau^j \in \Sigma^j$, let

$$\mathcal{R}^{i,n+1}(\rho, \tau^j) = \left\{ \tau^i \in \Sigma^i : \begin{array}{l} \exists \nu^i \in \Delta(\Theta^i) \text{ such that } \tau^i \in \mathrm{BR}^i(\nu^i) \text{ and} \\ \nu^i\big(\{(q, \widehat{\tau}^j) \in \Theta^i_\rho(\tau) : \widehat{\tau}^j \in \mathcal{R}^{j,n}(q, \tau^i)\}\big) = 1 \end{array} \right\}.$$

Under a parameter $(q, \widehat{\tau}^j) \in \Theta^i$, player i conjectures that Nature draws the state according to q and player j uses $\widehat{\tau}^j$. Given player i's candidate strategy τ^i, we therefore test $\widehat{\tau}^j$ in player j's hypothetical problem (q, τ^i). At every depth, wKLD is minimized over the full parameter space; the recursion only restricts which minimizers may belong to the support of player i's belief ν^i. For every integer $n \geq 2$, profile $\sigma \in \Sigma$, and player i, define $\Theta^{i,n}(\sigma) = \{(\pi^i, \widehat{\sigma}^j) \in \Theta^{i,0}(\sigma) : \widehat{\sigma}^j \in \mathcal{R}^{j,n-1}(\pi^i, \sigma^i)\}$, $\mathcal{B}^{i,n}(\sigma) = \big\{\mu^i \in \mathcal{B}^{i,0}(\sigma) : \mu^i\big(\Theta^{i,n}(\sigma)\big) = 1\big\}$.

Definition 5. For every integer $n \geq 2$, a profile $\sigma \in \Sigma$ is a *depth-n rationalizable Berk–Nash equilibrium* if $\mathcal{B}^{i,n}(\sigma) \neq \varnothing$ for every player i. Let $\mathcal{E}^n$ denote the set of these equilibria.

Lemma 2 (Nestedness)**.** *For every integer $n \geq 0$, pair of distinct players i and j, state distribution $\rho \in \Delta(\Omega)$, and opponent strategy $\tau^j \in \Sigma^j$,*

$$\mathcal{R}^{i,n+1}(\rho, \tau^j) \subseteq \mathcal{R}^{i,n}(\rho, \tau^j).$$

For every integer $n \geq 0$, every profile $\sigma \in \Sigma$, and every player i,

$$\Theta^{i,n+1}(\sigma) \subseteq \Theta^{i,n}(\sigma), \qquad \mathcal{B}^{i,n+1}(\sigma) \subseteq \mathcal{B}^{i,n}(\sigma).$$

Moreover, $\mathcal{E}^n \supseteq \mathcal{E}^{n+1}$.

The proof is in Appendix A.3.

Infinite-depth equilibrium. A profile $\sigma \in \Sigma$ is a *rationalizable Berk–Nash equilibrium* (RBNE) if $\sigma \in \mathcal{E}^n$ for every integer $n \geq 0$. The set of these equilibria is $\mathcal{E}^\infty = \bigcap_{n \geq 0} \mathcal{E}^n$. Equivalently,

$$\sigma \in \mathcal{E}^\infty \quad \Longleftrightarrow \quad \sigma^i \in \mathcal{R}^{i,n}(p, \sigma^j) \text{ for every } i \in I \text{ and every integer } n \geq 0.$$

Under Assumption 1, for every profile $\sigma \in \mathcal{E}^\infty$ and player i, there exists a belief $\mu^i \in \Delta(\Theta^i)$ over player i's subjective parameters that supports σ^i at every finite depth simultaneously.[11] The failures of existence described in Remark 3 may also occur at higher depths, so $\mathcal{E}^\infty$ may be empty.

[11] This follows from nestedness and compactness of the supporting-belief sets; see Appendix A.4.

3.4 Compatible explanations

This section characterizes $\mathcal{E}^\infty$ in terms of families of compatible explanations. For each player i, an explanation $e^i = (\rho^i, \tau^i; \widehat{\tau}^j)$ consists of one of her parameters $(\rho^i, \widehat{\tau}^j) \in \Theta^i$ and one of her strategies $\tau^i \in \Sigma^i$; the set of her explanations is

$$E^i = \left\{(\rho^i, \tau^i; \widehat{\tau}^j) : (\rho^i, \widehat{\tau}^j) \in \Theta^i,\ \tau^i \in \Sigma^i\right\}.$$

Under e^i, player i conjectures that player j uses $c(e^i) = \widehat{\tau}^j$ in the hypothetical problem $h(e^i) = (\rho^i, \tau^i)$. A belief $\nu^j \in \Delta(\Theta^j)$ for player j in this problem *justifies* e^i if it makes $\widehat{\tau}^j$ optimal and is supported on parameters that minimize player j's wKLD when she uses $\widehat{\tau}^j$. Thus, a justifying belief exists precisely when $c(e^i) \in \mathcal{R}^{j,0}(h(e^i))$. Each parameter $(\rho^j, \widehat{\tau}^i)$ in the support of a justifying belief ν^j, together with player j's candidate strategy $\widehat{\tau}^j$, generates the explanation $e^j = (\rho^j, \widehat{\tau}^j; \widehat{\tau}^i) \in E^j$. In this explanation, $\widehat{\tau}^j$, previously conjectured by player i, is player j's own candidate strategy. Player j conjectures that Nature draws the state according to ρ^j and player i uses $\widehat{\tau}^i$. Neither $\rho^j = \rho^i$ nor $\widehat{\tau}^i = \tau^i$ is required.

The successive restrictions in Definition 4 can be expressed through beliefs that justify an explanation of one player and generate explanations of the other. At a profile σ, a supporting belief μ^i for player i generates one explanation $e^i = (\pi^i, \sigma^i; \widehat{\sigma}^j)$ for each parameter $(\pi^i, \widehat{\sigma}^j)$ in its support, with $h(e^i) = (\pi^i, \sigma^i)$ and $c(e^i) = \widehat{\sigma}^j$. The depth-1 condition for player i is that some μ^i can be chosen so that each of these e^i has a justifying belief ν^j. The depth-2 condition is that μ^i and the beliefs ν^j can be chosen so that each explanation e^j generated by a ν^j has a justifying belief ν^i; each further depth imposes the same condition on the explanations generated at the previous depth, alternating between the players. Suppose there are sets $\mathcal{C}^i \subseteq E^i$, one for each player, such that every explanation in $\mathcal{C}^i$ has a justifying belief whose generated explanations all belong to $\mathcal{C}^j$. If a supporting belief μ^i generates only explanations in $\mathcal{C}^i$, this condition ensures that those explanations can be justified at every depth. We formalize this property below.

Definition 6. A pair $\mathcal{C} = (\mathcal{C}^1, \mathcal{C}^2)$ of nonempty closed sets $\mathcal{C}^i \subseteq E^i$ is a *family of compatible explanations* if, for each player i and every explanation $(\rho^i, \tau^i; \widehat{\tau}^j) \in \mathcal{C}^i$, there exists a belief $\nu^j \in \Delta(\Theta^j)$ such that

$$\widehat{\tau}^j \in \mathrm{BR}^j(\nu^j), \qquad \operatorname{supp} \nu^j \subseteq \left\{(\rho^j, \widehat{\tau}^i) \in \Theta^j_{\rho^i}(\tau^i, \widehat{\tau}^j) : (\rho^j, \widehat{\tau}^j; \widehat{\tau}^i) \in \mathcal{C}^j\right\}.$$

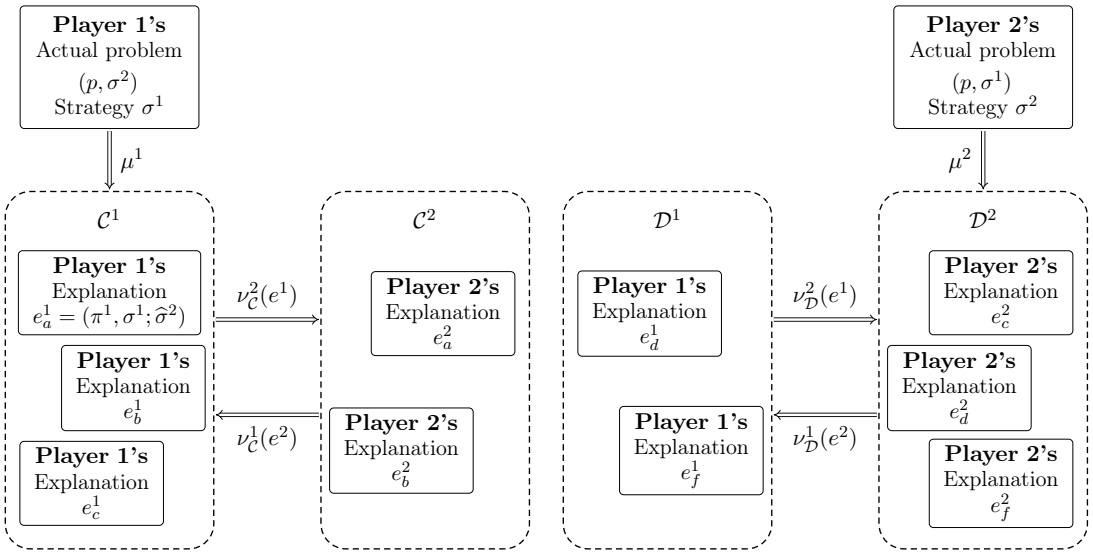
$\mathcal{C}$ is a family of compatible explanations if, for each player i and each explanation $e^i \in \mathcal{C}^i$, we can choose a justifying belief $\nu^j \in \Delta(\Theta^j)$ such that every explanation generated by a parameter in its support belongs to $\mathcal{C}^j$. Each of those explanations must in turn have a justifying belief whose generated explanations all belong to $\mathcal{C}^i$. The choice of a justifying belief may differ across explanations. Compatibility does not require the state

distributions or conjectures about a player's strategy in successive explanations to coincide. A justifying belief may have several or infinitely many parameters in its support, and the sets $\mathcal{C}^1, \mathcal{C}^2$ need not be finite.

Proposition 3 (Characterization by compatible explanations)**.** *A profile $\sigma \in \Sigma$ belongs to $\mathcal{E}^\infty$ if and only if there exist families of compatible explanations $\mathcal{C} = (\mathcal{C}^1, \mathcal{C}^2)$ and $\mathcal{D} = (\mathcal{D}^1, \mathcal{D}^2)$ and beliefs $\mu^1 \in \Delta(\Theta^1)$ and $\mu^2 \in \Delta(\Theta^2)$ satisfying*

$$\begin{aligned} & \sigma^1 \in \mathrm{BR}^1(\mu^1), \qquad \operatorname{supp} \mu^1 \subseteq \left\{(\pi^1, \widehat{\sigma}^2) \in \Theta^{1,0}(\sigma) : (\pi^1, \sigma^1; \widehat{\sigma}^2) \in \mathcal{C}^1\right\}, \\ \textit{and} \qquad & \sigma^2 \in \mathrm{BR}^2(\mu^2), \qquad \operatorname{supp} \mu^2 \subseteq \left\{(\pi^2, \widehat{\sigma}^1) \in \Theta^{2,0}(\sigma) : (\pi^2, \sigma^2; \widehat{\sigma}^1) \in \mathcal{D}^2\right\}. \end{aligned}$$

For sufficiency, the conditions in Proposition 3 ensure that each belief μ^i makes σ^i optimal and is supported on $\Theta^{i,0}(\sigma)$, so $\mu^i \in \mathcal{B}^{i,0}(\sigma)$. The explanations generated by μ^1 belong to $\mathcal{C}^1$, and those generated by μ^2 belong to $\mathcal{D}^2$. Within each family, every explanation has a justifying belief whose generated explanations belong to the other player's set. Applying this property successively verifies every additional finite-depth restriction on the explanations generated by μ^i. Thus, $\mu^i \in \mathcal{B}^{i,n}(\sigma)$ for every player i and every integer $n \geq 0$, so $\sigma \in \mathcal{E}^\infty$. The proof is in Appendix A.5. Note that for each player i, the sets $\mathcal{C}^i$ and $\mathcal{D}^i$ may be disjoint, overlap, or coincide.



(a) Player 1's reasoning (b) Player 2's reasoning

Figure 1: Compatible families supporting $\sigma \in \mathcal{E}^\infty$.

4 Existence

Proposition 3 characterizes RBNE by requiring, for each player, a supporting belief whose generated explanations belong to a family of compatible explanations. For a given payoff

structure, this section studies how subjective models and feedback determine whether such beliefs exist at some strategy profile. For this purpose, it is useful to distinguish whether a model contains the true state distribution from whether it can reproduce the consequence distributions generated at a given profile.

Definition 7 (Specification). (i) Player i's subjective model *contains the truth* if $p \in \Pi^i$ and *excludes the truth* otherwise.

(ii) Player i's subjective model *reproduces player i's consequence distributions* at $\sigma \in \Sigma$ if there exists $\theta^i \in \Theta^i$ such that $Q^i_{\theta^i}(\cdot \mid x^i) = Q^i_\sigma(\cdot \mid x^i)$ for every $x^i \in X^i$.[12]

(iii) Player i's subjective model *reproduces player i's observations* at $\sigma \in \Sigma$ if there exists $\theta^i \in \Theta^i$ such that $Q^i_{\theta^i}(\cdot \mid x^i) = Q^i_\sigma(\cdot \mid x^i)$ for every $x^i \in \operatorname{supp} \sigma^i$.

If $p \in \Pi^i$, then at every profile σ, the parameter (p, σ^j) reproduces player i's consequence distributions after every own action. Reproduction of consequence distributions at a profile implies reproduction of observations at that profile. Under Assumption 1, the model reproduces player i's observations at σ if and only if $\min_{\theta^i \in \Theta^i} K^i(\sigma, \theta^i) = 0$. A model may nevertheless reproduce consequence distributions after every own action at every profile without containing the truth.

4.1 Truth and reproduction of observations

Suppose the players' subjective models contain a common state distribution $\pi \in \Pi^1 \cap \Pi^2$. Consider a profile σ at which each player's strategy is a best response to the opponent's strategy when expected payoffs are evaluated under π. If, for each player i, the parameter (π, σ^j) reproduces her observations at σ under the true distribution p, then σ is a Berk–Nash equilibrium. The following proposition states that it is also an RBNE, providing a first benchmark for existence. Let $NE(\pi)$ denote the set of Nash equilibria with expected payoffs evaluated under π.

Proposition 4. *Let $\pi \in \Pi^1 \cap \Pi^2$ and $\sigma \in NE(\pi)$. If, for each player i, the parameter (π, σ^j) reproduces her observations at σ, then $\sigma \in \mathcal{E}^\infty$.*

For each player i, the belief $\delta_{(\pi,\sigma^j)}$ generates player j's hypothetical problem (π, σ^i). In this problem, (π, σ^i) reproduces player j's observations, and σ^j is optimal under $\delta_{(\pi,\sigma^i)}$ because $\sigma \in NE(\pi)$. The belief $\delta_{(\pi,\sigma^i)}$ then generates player i's hypothetical problem (π, σ^j), where (π, σ^j) reproduces her observations and σ^i is optimal under $\delta_{(\pi,\sigma^j)}$. These two beliefs therefore generate a two-step cycle of compatible explanations.

[12] Reproduction of consequence distributions is equivalent to correct specification given σ in the reduced-form subjective models of Esponda and Pouzo (2016, Definition 2). Our terminology distinguishes this distributional property from the model containing the true structural parameter (p, σ^j), which here is equivalent to $p \in \Pi^i$. An incorrect state distribution combined with an incorrect conjecture about the opponent's strategy may reproduce the true consequence distributions after every own action even when $p \notin \Pi^i$.

Corollary 5 (Existence when both models contain the truth)**.** *If* $p \in \Pi^1 \cap \Pi^2$*, then*

$$NE(p) \subseteq \mathcal{E}^\infty, \qquad \mathcal{E}^\infty \neq \varnothing.$$

The proposition and its corollary are derived in Appendix B.1. The corollary implies that nonexistence requires at least one player's subjective model to exclude the true state distribution. This condition is necessary but not sufficient. The preceding proposition can establish existence under a common state distribution even when both models exclude p. We therefore ask whether reproducing observations can guarantee existence even when the models exclude the truth. If each player's strategy is optimal under a belief supported on parameters that reproduce her observations, the profile is a Berk–Nash equilibrium. However, these parameters may specify opponent strategies that cannot be justified in the corresponding hypothetical problems. In Proposition 4, the parameter (π, σ^j) reproduces player i's observations. Under this parameter, she correctly conjectures that player j uses σ^j. In Example 3, both models exclude the truth and reproduce consequence distributions after every action at every profile. A Berk–Nash equilibrium exists, but no RBNE does.

Remark 4 (Robustness of nonexistence)*.* Under Assumption 1, nonexistence persists under sufficiently small changes in the objective state distribution and payoff functions. Fix the finite state, action, and consequence spaces, the feedback functions, and the subjective models. If $\mathcal{E}^\infty(p, u) = \varnothing$, then $\mathcal{E}^\infty(p', u') = \varnothing$ for every sufficiently close pair (p', u') with p' of full support. The proof is in Appendix B.5.

Example 3 (Reproduction of consequence distributions without existence)**.** There are two states, $\Omega = \{H, L\}$, and each player chooses whether to play, P, or stay out, O. Staying out pays zero. Playing pays one if the state is H and the opponent plays, or if the state is L and the opponent stays out; otherwise it pays zero. Each player observes only her own payoff. The true distribution satisfies $p(H) = 1/2$, while both players use the same model $\Pi^1 = \Pi^2 = \{q\}$, where $q(H) = 3/4$.

Under a parameter $(q, \widehat{\sigma}^j)$ with $\widehat{\sigma}^j(P) = r$, the expected payoff from playing is $\frac{3}{4}r + \frac{1}{4}(1 - r) = \frac{1}{4} + \frac{1}{2}r > 0$. Thus playing is strictly optimal under every belief over subjective parameters. Under the true distribution, playing yields payoff one with probability $1/2$ regardless of the opponent's strategy. The parameter with $r = 1/2$ reproduces each player's consequence distributions after both actions at every profile. At (P, P), each player's unique wKLD-minimizing parameter has $r = 1/2$. The point belief on this parameter makes playing strictly optimal, so $\mathcal{E}^0 = \{(P, P)\}$.

Under this parameter, each player conjectures that her opponent stays out with probability $1/2$. In the opponent's hypothetical problem (q, P), however, staying out remains strictly suboptimal under every belief over her parameters. The conjectured mixture is therefore not a Berk–Nash response. Both players' unique minimizing parameters are excluded at the first restriction, giving $\mathcal{E}^1 = \mathcal{E}^\infty = \varnothing$. A common

model can reproduce consequence distributions after every action without supporting an RBNE.

4.2 Existence by Nash equilibrium

Section 4.1 establishes existence through a two-step cycle of compatible explanations supporting a Nash equilibrium under a common state distribution. Such a cycle can also involve different state distributions for each player. Following Dekel et al. (2004), we consider Nash equilibria with possibly diverse priors: each player's strategy is optimal against the opponent's strategy when expected payoffs are evaluated under her own state distribution.

Definition 8. Fix $q^i \in \Delta(\Omega)$ for each player i. A strategy profile $\sigma \in \Sigma$ is a Nash equilibrium under (q^1, q^2) if, for each player i, $\sigma^i \in \mathrm{BR}^i\big(\delta_{(q^i,\sigma^j)}\big)$.[13]

Let $NE(q^1, q^2)$ denote the set of these equilibria, and write $NE(q, q) = NE(q)$. Since the game with expected payoffs evaluated under (q^1, q^2) is finite, $NE(q^1, q^2) \neq \varnothing$.

Proposition 6 (Nash cycle)**.** *Let $\pi^1 \in \Pi^1$, $\pi^2 \in \Pi^2$, and $\sigma \in NE(\pi^1, \pi^2)$. If, for each pair of distinct players $i, j \in \{1, 2\}$,*

$$(\pi^i, \sigma^j) \in \Theta^{i,0}(\sigma) \cap \Theta^i_{\pi^j}(\sigma),$$

then $\sigma \in \mathcal{E}^\infty$.

The proof is in Appendix B.1.

For each player i, the Nash equilibrium condition makes σ^i optimal under $\delta_{(\pi^i,\sigma^j)}$, while $(\pi^i, \sigma^j) \in \Theta^{i,0}(\sigma)$ ensures that this belief is supported on parameters minimizing her wKLD in the actual problem. Together, the two beliefs support σ as a Berk–Nash equilibrium. The additional condition $(\pi^i, \sigma^j) \in \Theta^i_{\pi^j}(\sigma)$ ensures that the same parameter minimizes player i's wKLD in the hypothetical problem (π^j, σ^j) specified by player j's explanation. The two beliefs therefore generate a two-step cycle of compatible explanations, establishing $\sigma \in \mathcal{E}^\infty$.

Figure 2 illustrates this construction, with $\theta^i = (\pi^i, \sigma^j)$ for each player i.

Remark 5 (Beyond Nash cycles). The construction in Proposition 6 uses the same belief $\delta_{(\pi^i,\sigma^j)}$ in player i's actual problem and in the hypothetical problem generated by player j. The definition of RBNE does not require these beliefs to coincide: player i's belief in the hypothetical problem generated by player j may differ from her supporting belief at the actual profile. Nor must compatible explanations form a two-step cycle. They may

[13] Here we extend BR^i to beliefs $\mu^i \in \Delta(\Delta(\Omega) \times \Sigma^j)$, evaluating expected payoffs under $Q^i_{q^i,\widehat{\sigma}^j}$ and averaging over $(q^i, \widehat{\sigma}^j)$ according to μ^i. In particular, $\mathrm{BR}^i(\delta_{(q^i,\sigma^j)})$ is defined even when $q^i \notin \Pi^i$.

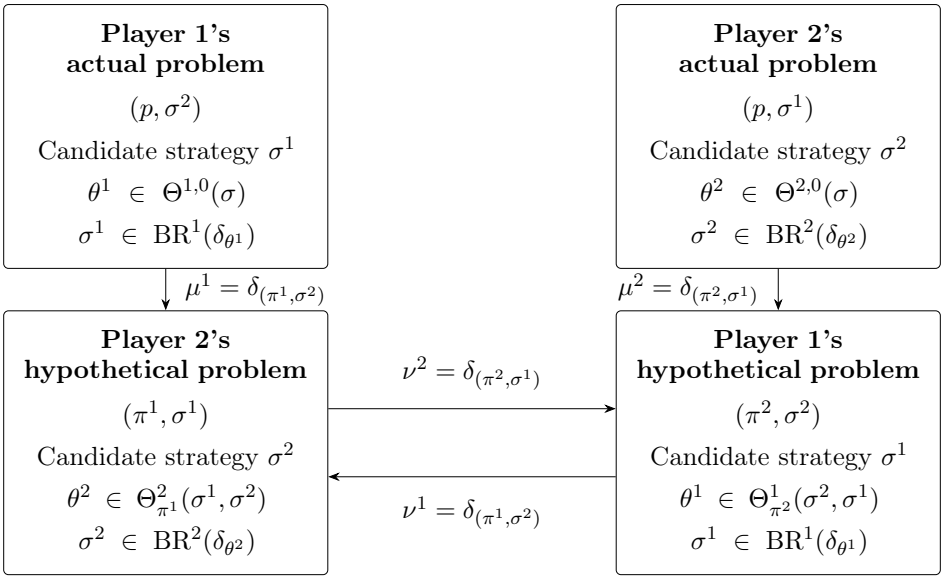


Figure 2: A two-step Nash cycle supporting σ as an RBNE.

involve longer cycles, beliefs supported on several parameters, or infinite families. The two players may also rely on different families of compatible explanations, as allowed by Proposition 3.

4.3 All-or-nothing feedback

We now consider the following restrictions on feedback under which wKLD requirements simplify.

Definition 9. Player i's feedback function f^i is *all-or-nothing* if it satisfies both of the following conditions:

(i) For every own action $x^i \in X^i$, either (a) $f^i(x^i, x^j, \omega) \neq f^i(x^i, \tilde{x}^j, \omega')$ for all distinct $x^j, \tilde{x}^j \in X^j$ and all $\omega, \omega' \in \Omega$, or (b) $f^i(x^i, x^j, \omega) = f^i(x^i, \tilde{x}^j, \omega)$ for all $x^j, \tilde{x}^j \in X^j$ and all $\omega \in \Omega$.

(ii) For every action profile $(x^i, x^j) \in X^i \times X^j$, the map $\omega \mapsto f^i(x^i, x^j, \omega)$ is either injective or constant.

When f^1, f^2 are all-or-nothing, a strategy profile $\sigma \in \Sigma$ is *informative* if, for each player i, there is an own action $x^i \in \operatorname{supp} \sigma^i$ satisfying (i)(a), and there exists an action profile $(\tilde{x}^i, \tilde{x}^j) \in \operatorname{supp} \sigma^i \times \operatorname{supp} \sigma^j$ at which $\omega \mapsto f^i(\tilde{x}^i, \tilde{x}^j, \omega)$ is injective.

Condition (i) requires that, after each own action, player i either observes player j's action regardless of the state or cannot distinguish between any two of player j's actions from her feedback when the state is held fixed. Condition (ii) requires that, holding the action profile fixed, player i either observes the state or cannot distinguish between any

two states from her feedback. Under these conditions, wKLD decomposes into a term concerning opponent strategies and a term concerning state distributions.

Lemma 7 (Decomposition of wKLD)**.** *Suppose f^i is all-or-nothing. For every $\rho \in \Delta(\Omega)$, $\tau \in \Sigma$, and $(\widehat{\rho}, \widehat{\tau}^j) \in \Theta^i$,*

$$K^i_\rho(\tau; (\widehat{\rho}, \widehat{\tau}^j)) = \alpha^i \sum_{x^j \in X^j} \tau^j(x^j) \ln \frac{\tau^j(x^j)}{\widehat{\tau}^j(x^j)} + \beta^i \sum_{\omega \in \Omega} \rho(\omega) \ln \frac{\rho(\omega)}{\widehat{\rho}(\omega)},$$

where

$$\alpha^i = \sum_{x^i} \tau^i(x^i) \mathbf{1}\{f^i(x^i, \cdot, \cdot) \text{ reveals the opponent's action}\},$$
$$\beta^i = \sum_{x^i, x^j} \tau^i(x^i)\tau^j(x^j) \mathbf{1}\{\omega \mapsto f^i(x^i, x^j, \omega) \text{ is injective}\}.$$

The proof is in Appendix B.2.

The two summation terms are Kullback–Leibler (KL) divergences.[14] Since the coefficients α^i and β^i do not depend on $(\widehat{\rho}, \widehat{\tau}^j)$ and $\Theta^i = \Pi^i \times \Sigma^j$, the two terms can be minimized separately.

The opponent-strategy term is minimized by the correct conjecture $\widehat{\tau}^j = \tau^j$. For the state-distribution term, denote the KL divergence between ρ and $\widehat{\rho}$ by $D(\rho \| \widehat{\rho}) = \sum_{\omega \in \Omega} \rho(\omega) \ln \frac{\rho(\omega)}{\widehat{\rho}(\omega)}$ and define the set of KL projections of ρ onto Π^i by

$$\mathsf{P}_i(\rho) = \arg\min_{\widehat{\rho} \in \Pi^i} D(\rho || \widehat{\rho}).$$

Thus, for every problem (ρ, τ^j) and candidate strategy τ^i, writing $\tau = (\tau^1, \tau^2)$, $\mathsf{P}_i(\rho) \times \{\tau^j\} \subseteq \Theta^i_\rho(\tau)$. These need not be the only minimizing parameters when α^i or β^i is zero. The following corollary simplifies Proposition 6 using the projections.

Corollary 8. *Suppose f^1, f^2 are all-or-nothing. Let $\pi^i \in \Pi^i$ for $i = 1, 2$. If for each pair of distinct players i and j,*

$$\pi^i \in \mathsf{P}_i(p) \cap \mathsf{P}_i(\pi^j)$$

then

$$NE(\pi^1, \pi^2) \subseteq \mathcal{E}^\infty, \qquad \mathcal{E}^\infty \neq \varnothing.$$

For any $\sigma \in NE(\pi^1, \pi^2)$, Lemma 7 and $\pi^i \in \mathsf{P}_i(p)$ imply $(\pi^i, \sigma^j) \in \Theta^{i,0}(\sigma)$. Similarly, $\pi^i \in \mathsf{P}_i(\pi^j)$ implies $(\pi^i, \sigma^j) \in \Theta^i_{\pi^j}(\sigma)$, so the same parameter minimizes player i's wKLD in the hypothetical problem (π^j, σ^j) generated by player j. The Nash condition makes σ^i optimal under $\delta_{(\pi^i, \sigma^j)}$. Applying Proposition 6 therefore gives $NE(\pi^1, \pi^2) \subseteq \mathcal{E}^\infty$.

Corollary 9 (Existence under a common model)**.** *Suppose f^1, f^2 are all-or-nothing and $\Pi^1 = \Pi^2 = \Pi$. Write $\mathsf{P} = \mathsf{P}_1 = \mathsf{P}_2$. Then $NE(\pi) \subseteq \mathcal{E}^\infty$ for every $\pi \in \mathsf{P}(p)$, so $\mathcal{E}^\infty \neq \varnothing$.*

[14] We use the conventions $0 \ln(0/b) = 0$ for every $b \geq 0$ and $a \ln(a/0) = +\infty$ for every $a > 0$. Terms with zero coefficients are omitted.

Appendix B.2 derives the two projection corollaries and the necessary condition below.

Under all-or-nothing feedback, suppose $\pi \in \mathsf{P}_1(p) \cap \mathsf{P}_2(p)$. Then π belongs to both subjective models, and $D(\pi\|\pi) = 0$ implies $\pi \in \mathsf{P}_i(\pi)$ for each player i. Taking $\pi^1 = \pi^2 = \pi$ therefore satisfies the conditions of Corollary 8, giving $NE(\pi) \subseteq \mathcal{E}^\infty$. Since $NE(\pi) \neq \varnothing$, this guarantees existence. The common-model corollary follows because both players then have the same nonempty projection set $\mathsf{P}(p)$. More generally, this argument shows that nonexistence under all-or-nothing feedback requires $\mathsf{P}_1(p) \cap \mathsf{P}_2(p) = \varnothing$.

Convex models Suppose both players' feedback functions are all-or-nothing and each Π^i is convex. Consider an informative profile σ. For any $\rho \in \Delta(\Omega)$ with full support, its KL projection onto each Π^i has finite divergence and therefore also has full support. Strict convexity of $D(\rho\|\widehat{\rho})$ in $\widehat{\rho}$, together with convexity of Π^i, implies $\mathsf{P}_i(\rho) = \{\mathsf{p}_i(\rho)\}$ for a unique state distribution $\mathsf{p}_i(\rho)$. The following proposition states that an informative profile is an RBNE if and only if it is a Nash equilibrium under every pair of state distributions in the projection sequence defined below.

Proposition 10 (Characterization under convex models)**.** *Suppose both players' feedback functions are all-or-nothing and each Π^i is convex. Let σ be informative. Define*

$$\pi_0^i = \mathsf{p}_i(p), \qquad \pi_{k+1}^i = \mathsf{p}_i(\pi_k^j), \qquad i \neq j, \quad k \geq 0.$$

Then

$$\sigma \in \mathcal{E}^\infty \iff \sigma \in NE(\pi_k^1, \pi_k^2) \quad \textit{for every } k \geq 0.$$

The proof, including the corresponding characterization at every finite depth, is in Appendix B.3.

At an informative profile, both coefficients in the wKLD decomposition are positive, so wKLD minimization identifies the opponent's strategy as σ^j, while convexity makes the state projection unique. In each successive hypothetical problem, player i's unique minimizing parameter specifies player j's strategy as σ^j, while its state distribution follows the projection sequence. The profile σ is therefore an RBNE if and only if, for every player i and every integer $k \geq 0$, each action in $\operatorname{supp}\sigma^i$ maximizes expected payoff against σ^j under the state distribution π_k^i.

Example 4 (Inspection with convex models)**.** A firm chooses compliance C or violation V, and an inspector chooses no inspection N or inspection I. The state determines the scale of activity: $r(H) = 3$ and $r(L) = 1$, with $p(H) = 1/2$. Violation earns r, inspection costs r, and an inspected violator pays the inspector a fine of 4. Payoffs are

	N	I
C	$(0, 0)$	$(0, -r)$
V	$(r, 0)$	$(r-4, 4-r)$

.

Each player observes only her own payoff. Compliance and no inspection always give zero. Violation and inspection each produce four distinct possible payoffs, $-3, -1, 1, 3$, identifying both the state and the opponent's action. Feedback is therefore all-or-nothing.

The firm overestimates the scale of activity, while the inspector underestimates it:

$$\Pi^1 = \{q \in \Delta(\Omega) : q(H) \geq 3/4\}, \qquad \Pi^2 = \{q \in \Delta(\Omega) : q(H) \leq 1/4\}.$$

These models are compact and convex. Let $\pi^1(H) = 3/4$ and $\pi^2(H) = 1/4$. The KL projections satisfy

$$\mathsf{p}_1(p) = \mathsf{p}_1(\pi^2) = \pi^1, \qquad \mathsf{p}_2(p) = \mathsf{p}_2(\pi^1) = \pi^2.$$

Consider the mixed profile

$$\sigma^1(V) = \frac{3}{8}, \qquad \sigma^2(I) = \frac{5}{8}.$$

Under π^1, the firm's expected payoff difference between violation and compliance is $5/2 - 4\sigma^2(I) = 0$. Under π^2, the inspector's expected payoff difference between inspection and no inspection is $4\sigma^1(V) - 3/2 = 0$. Because σ is informative and both models are convex, each player i has the unique wKLD-minimizing parameter (π^i, σ^j) at σ. The payoff calculations show that σ^i is optimal under the point belief $\delta_{(\pi^i,\sigma^j)}$. For each player i, π^i is also the KL projection of π^j onto Π^i, so the same point belief $\delta_{(\pi^i,\sigma^j)}$ justifies σ^i in the hypothetical problem (π^j, σ^j). Corollary 8 therefore gives $\sigma \in \mathcal{E}^\infty$.

The complete equilibrium sets are

$$\mathcal{E}^0 = \{(C, N), (V, N), \sigma\}, \qquad \mathcal{E}^1 = \{(V, N), \sigma\}, \qquad \mathcal{E}^2 = \mathcal{E}^\infty = \{\sigma\}.$$

At (C, N), compliance can be optimal only if the firm expects inspection with positive probability. In each hypothetical problem (q, C) for the inspector, with $q \in \Pi^1$, observations following inspection reveal compliance. If inspection is played with positive probability, all parameters minimizing the inspector's wKLD therefore specify pure compliance, making inspection strictly suboptimal under every belief supported on those parameters. The first restriction thus leaves the firm only parameters specifying no inspection, under which violation is strictly preferable.

At (V, N), every firm strategy surviving the first restriction in a hypothetical problem (q, N), with $q \in \Pi^2$, is pure violation. If a candidate strategy uses violation with positive probability, observations following violation reveal no inspection, so violation is strictly optimal under every belief supported on the firm's wKLD minimizers. Pure compliance fails the first restriction by the preceding argument. The second restriction therefore leaves the inspector only parameters specifying violation, under which inspection is strictly

profitable. No belief supported on these parameters makes no inspection optimal.

Appendix B.4 verifies the equilibrium sets and the minimum wKLD in the actual and hypothetical problems.

5 Relationship to other solution concepts

Rationalizable conjectural equilibrium. Restricting equilibrium beliefs through introspection about opponents' behavior is familiar from Rubinstein and Wolinsky (1994) and Esponda (2013). In games without payoff uncertainty, Rubinstein and Wolinsky (1994) consider pure action profiles. Each player's action must be optimal under a conjecture consistent with her observations, and each opponent action receiving positive probability must itself be justified by optimality and consistency with observations. These requirements apply at every depth. Esponda (2013) extends RCE to games with payoff uncertainty. His framework specifies a common space of fundamentals containing the true fundamental and player-specific feedback partitions. A fundamental θ describes uncertain features of the environment, such as a payoff parameter, a distribution over payoff states, or the precision of private information. He defines an RCE as a strategy profile supported by beliefs under which players are rational and consistent with their information, with common belief of both properties.

To map Esponda's framework into our setting, take the fundamental to be a state distribution q, with true fundamental p, while keeping the game structure and feedback fixed. Each player's action in this representation is a complete mixed strategy. Suppose $p \in \Pi^1 \cap \Pi^2$, and take $\Pi = \Pi^1 \cap \Pi^2$ as the common model for the RCE benchmark. In this specification, each player knows her own strategy and observes the conditional distribution of consequences after each action she plays with positive probability.

In an actual or hypothetical problem with state distribution ρ and candidate profile τ, consistency requires a belief $\mu^i \in \Delta(\Pi \times \Sigma^j)$ satisfying $\mu^i(\{\theta^i : K^i_\rho(\tau; \theta^i) = 0\}) = 1$. Optimality additionally requires $\tau^i \in \mathrm{BR}^i(\mu^i)$. Esponda's characterization imposes both requirements at every depth. At the actual profile, $\rho = p$ and $\tau = \sigma$.

Let $\mathcal{E}^{\mathrm{RCE}}_\Pi(p)$ denote the corresponding set of RCE profiles. Under $\Pi = \Pi^1 \cap \Pi^2 \ni p$, this set is contained in the RBNE set:

$$\mathcal{E}^{\mathrm{RCE}}_\Pi(p) \subseteq \mathcal{E}^\infty.$$

For each player i, every parameter in the support of her RCE belief belongs to Θ^i and has zero wKLD in the corresponding problem. These parameters therefore still minimize wKLD over Θ^i. The same beliefs continue to make the candidate strategies optimal and to justify the conjectured strategies in the successive hypothetical problems. When

$\Pi^1 = \Pi^2 = \Pi$, the two equilibrium sets coincide. Appendix C.1 proves the common-model equivalence and then derives the inclusion when the subjective models differ.

ζ–rationalizability Battigalli and Siniscalchi (2003) introduce Δ-rationalizability, where Δ specifies restrictions on each player's beliefs about opponents' payoff types and strategies. A player's strategy must be a best response to a belief satisfying the restrictions specified for her by Δ. Each conjectured opponent strategy must likewise be a best response for the corresponding payoff type to a belief satisfying the restrictions that Δ specifies for that opponent and type. In their statistical specification, ζ-rationalizability, the restrictions come from public statistics about payoff types and observed play. Beliefs must reproduce the opponent-type and action frequencies implied by these statistics. The same public statistics constrain the beliefs used to justify behavior at every depth.

RBNE restricts the parameters used in supporting beliefs through the hypothetical problems they describe. Consider a wKLD-minimizing parameter $(\pi^i, \widehat{\sigma}^j)$ in the support of a belief making player i's strategy σ^i optimal. Whether or not this explanation is correct, it specifies player j's hypothetical problem (π^i, σ^i), with consequence distributions $Q^j_{\pi^i,\sigma^i}$. Player j must have a belief supported on parameters minimizing her wKLD in (π^i, σ^i) when she uses $\widehat{\sigma}^j$. If no such belief makes $\widehat{\sigma}^j$ optimal, the parameter $(\pi^i, \widehat{\sigma}^j)$ is ruled out. A parameter can minimize player i's wKLD yet fail to justify the opponent's conjectured strategy. The same requirement applies to every conjecture used in the opponent's justifying belief. If no belief supported on the surviving original minimizers makes σ^i optimal, the profile cannot be an RBNE. Thus, each player's strategy must be optimal under a belief supported on parameters minimizing her wKLD. These statistical and optimality requirements must also hold for the beliefs justifying the conjectured strategies in every subsequent hypothetical problem.

Both concepts impose statistical restrictions on the beliefs used to justify conjectured strategies at every depth. In ζ-rationalizability, these restrictions refer to the same public statistics throughout. In RBNE, each explanation determines the opponent's hypothetical observations, and her subjective model determines which parameters minimize wKLD for those observations. Player j's justifying belief need not be supported on parameters whose state distribution equals π^i, and her minimum wKLD may be positive. RBNE therefore asks which explanations of a player's observations remain compatible with the opponent's inference and optimal choice in the hypothetical problems they generate.

References

Pierpaolo Battigalli. Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali. Unpublished thesis, Università Bocconi, 1987.

Pierpaolo Battigalli and Danilo Guaitoli. Conjectural Equilibria and Rationalizability in a Game with Incomplete Information. In Pierpaolo Battigalli, Aldo Montesano, and Fausto Panunzi, editors, *Decisions, Games and Markets*, pages 97–124. Kluwer Academic Publishers, Dordrecht, 1997. doi: 10.1007/978-1-4615-6337-2_4.

Pierpaolo Battigalli and Marciano Siniscalchi. Rationalization and Incomplete Information. *Advances in Theoretical Economics*, 3(1), 2003. Article 3.

B. Douglas Bernheim. Rationalizable Strategic Behavior. *Econometrica*, 52(4):1007–1028, 1984.

Eddie Dekel, Drew Fudenberg, and David K. Levine. Learning to Play Bayesian Games. *Games and Economic Behavior*, 46(2):282–303, 2004.

Ignacio Esponda. Rationalizable Conjectural Equilibrium: A Framework for Robust Predictions. *Theoretical Economics*, 8(2):467–501, 2013.

Ignacio Esponda and Demian Pouzo. Berk–Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models. *Econometrica*, 84(3):1093–1130, 2016.

Mira Frick, Ryota Iijima, and Yuhta Ishii. Dispersed Behavior and Perceptions in Assortative Societies. *American Economic Review*, 112(9):3063–3105, 2022. doi: 10.1257/aer.20190486.

Drew Fudenberg and Yuichiro Kamada. Rationalizable Partition-Confirmed Equilibrium. *Theoretical Economics*, 10(3):775–806, 2015.

Drew Fudenberg and Yuichiro Kamada. Rationalizable Partition-Confirmed Equilibrium with Heterogeneous Beliefs. *Games and Economic Behavior*, 109:364–381, 2018. doi: 10.1016/j.geb.2017.12.021.

Drew Fudenberg and David K. Levine. Self-Confirming Equilibrium. *Econometrica*, 61(3): 523–545, 1993.

Kevin He and Jonathan Libgober. Higher-Order Beliefs and (Mis)learning from Prices. Working paper, 2025.

Paul Heidhues, Botond Kőszegi, and Philipp Strack. Misinterpreting Yourself. Working paper, 2025.

Philippe Jehiel. Analogy-Based Expectation Equilibrium. *Journal of Economic Theory*, 123(2):81–104, 2005. doi: 10.1016/j.jet.2003.12.003.

Philippe Jehiel and Frédéric Koessler. Revisiting Games of Incomplete Information with Analogy-Based Expectations. *Games and Economic Behavior*, 62(2):533–557, 2008. doi: 10.1016/j.geb.2007.06.006.

Philippe Jehiel and Giacomo Weber. Endogenous Clustering and Analogy-Based Expectation Equilibrium. *Review of Economic Studies*, 93(2):1077–1102, 2026. doi: 10.1093/restud/rdaf054.

David G. Pearce. Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica*, 52(4):1029–1050, 1984.

Ariel Rubinstein and Asher Wolinsky. Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability. *Games and Economic Behavior*, 6(2):299–311, 1994.

A Proofs and foundations

This appendix proves the results in Sections 3, 4, and 5. Throughout, when the strategies τ^i and τ^j are specified, $\tau = (\tau^1, \tau^2)$ denotes their profile in fixed player order; the arguments of $K^i_\rho(\tau;\cdot)$ and $\Theta^i_\rho(\tau)$ use that order.

A.1 wKLD minimizers

Lemma 11 (wKLD minimizers)**.** *Under Assumption 1, for every pair of distinct players i and j and every $(\rho, \tau^i, \tau^j) \in \Delta(\Omega) \times \Sigma^i \times \Sigma^j$, the function $K^j_\rho(\tau;\cdot)$ takes at least one finite value and attains a finite minimum on Θ^j. Consequently, $\Theta^j_\rho(\tau)$ is nonempty and compact. The correspondence $(\rho, \tau^i, \tau^j) \mapsto \Theta^j_\rho(\tau)$ has a closed, hence compact, graph.*

Proof. Apply the argument of Esponda and Pouzo (2016, Lemma 1(ii) and Appendix A, Claim A(i)–(iii)) to the frequencies $P^j_{\rho,\tau}(x^j, y^j) = \tau^j(x^j)Q^j_{\rho,\tau^i}(y^j \mid x^j)$. Their result fixes the objective state distribution. Here, the frequencies vary continuously with both ρ and τ. Assumption 1(ii) provides approximating parameters whose subjective distributions assign positive probability to every feasible consequence following each action. The same approximating parameters work for every ρ, including distributions without full support. Their arguments for a finite minimum and a closed graph therefore apply jointly in (ρ, τ). Compactness of the parameter and strategy spaces and the simplex of state distributions gives the remaining conclusions. □

A.2 Existence of Berk–Nash responses

Proof of Lemma 1. Fix (ρ, τ^i) and consider player j's one-player problem with state space $\mathrm{supp}(\rho) \times \mathrm{supp}(\tau^i)$, full-support distribution $\rho \otimes \tau^i$, consequence function $f^j(x^j, x^i, \omega)$, payoff function u^j, and original subjective model. Its consequence distributions are Q^j_{ρ,τ^i} and its wKLD is $K^j_\rho(\tau;\cdot)$. Assumption 1 implies Esponda and Pouzo (2016, Assumption 1) for this problem. Their Theorem 1 gives a strategy optimal under a belief over player j's subjective parameters supported on $\Theta^j_\rho(\tau)$, which is precisely a strategy in $\mathcal{R}^{j,0}(\rho, \tau^i)$. □

A.3 Nestedness of the recursive restrictions

Proof of Lemma 2. The depth-1 requirement restricts the support of a belief to a subset of the original wKLD minimizers, so $\mathcal{R}^{i,1} \subseteq \mathcal{R}^{i,0}$. If $\mathcal{R}^{j,n+1} \subseteq \mathcal{R}^{j,n}$ for both players, every belief satisfying the depth-$(n+2)$ requirement also satisfies the depth-$(n+1)$ requirement. Simultaneous induction proves nestedness of all response correspondences. Substitution into the definitions of $\Theta^{i,n}$, $\mathcal{B}^{i,n}$, and $\mathcal{E}^n$ gives their respective inclusions. □

A.4 Compactness and infinite-depth support

Define $\mathcal{R}^{i,\infty}(\rho,\tau^j)=\bigcap_{n\geq 0}\mathcal{R}^{i,n}(\rho,\tau^j)$, and use weak convergence for beliefs.

Proposition 12 (Compactness and infinite-depth support)**.** *Under Assumption 1, for every player i and finite integer $n\geq 0$, the graph of $\mathcal{R}^{i,n}$ is compact. For every profile σ, the sets $\Theta^{i,n}(\sigma)$ and $\mathcal{B}^{i,n}(\sigma)$ are compact, and the set $\mathcal{E}^n$ is compact.*

For every problem (ρ,τ^j) and candidate strategy τ^i, we have $\tau^i\in\mathcal{R}^{i,\infty}(\rho,\tau^j)$ if and only if there is a belief $\nu^i\in\Delta(\Theta^i)$ such that $\tau^i\in \mathrm{BR}^i(\nu^i)$ and

$$\nu^i\big(\{(q,\widehat{\tau}^j)\in\Theta^i_\rho(\tau):\widehat{\tau}^j\in\mathcal{R}^{j,\infty}(q,\tau^i)\}\big)=1.$$

In particular, if $\sigma\in\mathcal{E}^\infty$, then $\bigcap_{n\geq 0}\mathcal{B}^{i,n}(\sigma)\neq\varnothing$ for each player i.

Proof. By the argument in Esponda (2013, proof of Lemma 3), if a parameter correspondence on compact spaces has a closed graph, the correspondence of beliefs supported on its values also has a closed graph. Lemma 11 and continuity of expected payoffs then imply that the set of problems, candidate strategies, and beliefs satisfying the depth-zero requirements is compact. Its projection onto the problem and strategy coordinates is the graph of $\mathcal{R}^{i,0}$, which is therefore compact.

For $a=(\rho,\tau^j,\tau^i)$, define

$$\begin{aligned}F^i_n(a)&=\{(q,\widehat{\tau}^j)\in\Theta^i_\rho(\tau):\widehat{\tau}^j\in\mathcal{R}^{j,n}(q,\tau^i)\},\\ W^i_n(a)&=\{\nu^i\in\Delta(\Theta^i):\tau^i\in\mathrm{BR}^i(\nu^i),\ \nu^i(F^i_n(a))=1\}.\end{aligned}$$

If both depth-n response graphs are compact, F^i_n has a closed graph. The same argument for limits of supporting beliefs and continuity of expected payoffs make the graph of W^i_n compact. Its projection onto the problem and strategy coordinates is the graph of $\mathcal{R}^{i,n+1}$, proving compactness by induction. The definitions then make $\Theta^{i,n}(\sigma)$ and $\mathcal{B}^{i,n}(\sigma)$ compact. Since $\sigma\in\mathcal{E}^n$ exactly when $\sigma^i\in\mathcal{R}^{i,n}(p,\sigma^j)$ for both players, $\mathcal{E}^n$ is compact as well.

If $\tau^i\in\mathcal{R}^{i,\infty}(\rho,\tau^j)$, the sets $W^i_n(a)$ are nonempty, compact, and nested. Choose ν^i in their intersection. Continuity of probability on descending sets gives $\nu^i(\bigcap_n F^i_n(a))=1$, which is the stated infinite-depth support condition. Conversely, a belief satisfying that condition supports τ^i at every finite depth. At an RBNE profile, the sets $\mathcal{B}^{i,n}(\sigma)$ are likewise nonempty, compact, and nested, so their intersection is nonempty. □

A.5 Compatible explanations

Proof of Proposition 3. For any family $\mathcal{C}$ in Definition 6, we first show that $c(e^i)\in\mathcal{R}^{j,n}(h(e^i))$ for every $e^i\in\mathcal{C}^i$ and every finite n. At depth zero this follows from the justifying belief. Suppose it holds at depth n for all explanations in both sets. A belief

justifying e^i generates only explanations in $\mathcal{C}^j$, whose conjectured strategies satisfy the depth-n requirement by induction. That same belief therefore satisfies the conditions for $c(e^i)$ at depth $n+1$. If the actual supporting beliefs in the proposition generate explanations in $\mathcal{C}^1$ and $\mathcal{D}^2$, respectively, this argument applies to each family and gives $\mu^i \in \mathcal{B}^{i,n}(\sigma)$ for both players and every n. Hence $\sigma \in \mathcal{E}^\infty$.

Conversely, suppose $\sigma \in \mathcal{E}^\infty$ and define

$$\mathcal{C}^i = \{e^i \in E^i : c(e^i) \in \mathcal{R}^{j,\infty}(h(e^i))\}.$$

These sets are closed by Proposition 12. For every $e^i \in \mathcal{C}^i$, that proposition gives a justifying belief supported on the opponent's wKLD minimizers and generating only explanations in $\mathcal{C}^j$. It also gives beliefs $\mu^i \in \bigcap_{n\geq 0} \mathcal{B}^{i,n}(\sigma)$ supporting each player's strategy in her actual problem. Each such belief assigns probability one to the closed set

$$\bigcap_{n\geq 0} \Theta^{i,n}(\sigma) = \{(\pi^i, \widehat{\sigma}^j) \in \Theta^{i,0}(\sigma) : \widehat{\sigma}^j \in \mathcal{R}^{j,\infty}(\pi^i, \sigma^i)\}.$$

Every parameter in its support therefore generates an explanation in $\mathcal{C}^i$, so both sets $\mathcal{C}^i$ are nonempty. Thus $\mathcal{C}$ is a family of compatible explanations, and taking $\mathcal{D} = \mathcal{C}$ proves necessity. The sufficiency argument allows the players to use different families, as in the statement. □

B Proofs for existence and nonexistence

The Nash-cycle proposition follows from the characterization by compatible explanations. It implies the existence results for a common state distribution reproducing observations and for models containing the truth. Together with the wKLD decomposition, it also implies the projection corollaries.

B.1 Nash cycles and common distributions

Proof of Proposition 6. For each player i, let $\mu^i = \delta_{(\pi^i,\sigma^j)}$ and $e^i = (\pi^i, \sigma^i; \sigma^j)$. The Nash condition makes σ^i optimal under μ^i. Since $(\pi^i, \sigma^j) \in \Theta^{i,0}(\sigma)$, this belief assigns probability one to a parameter minimizing wKLD in her actual problem. It therefore supports her strategy in that problem. Under e^i, player j faces (π^i, σ^i) and is conjectured to use σ^j. The same belief μ^j justifies e^i: it makes σ^j optimal and, by $(\pi^j, \sigma^i) \in \Theta^j_{\pi^i}(\sigma)$, is supported on her wKLD minimizers in this problem. It generates e^j, which μ^i justifies in turn. Thus $\mathcal{C}^i = \{e^i\}$ defines a family of compatible explanations containing those generated by both actual supporting beliefs. Proposition 3 gives $\sigma \in \mathcal{E}^\infty$. □

A common distribution reproducing observations. For Proposition 4, take $\pi^1 = \pi^2 = \pi \in \Pi^1 \cap \Pi^2$ and $\sigma \in NE(\pi)$. Reproduction gives $K_p^i(\sigma; (\pi, \sigma^j)) = 0$, while the same parameter gives $K_\pi^i(\sigma; (\pi, \sigma^j)) = 0$ in the hypothetical problem. By nonnegativity of wKLD (Esponda and Pouzo, 2016, Lemma 1(i)), it minimizes both expressions. Proposition 6 therefore applies.

Corollary 5 follows by additionally taking $\pi = p$. When $p \in \Pi^1 \cap \Pi^2$, each (p, σ^j) reproduces player i's observations, so every $\sigma \in NE(p)$ satisfies the preceding proposition. Since the objective game is finite, $NE(p) \neq \varnothing$. Hence nonexistence requires at least one model to exclude p.

B.2 Proof of Lemma 7

Proof. Fix player i, a state distribution ρ, a profile τ, and a parameter $(q, \widehat{\tau}^j) \in \Theta^i$, where $q = \widehat{\rho}$ in the statement. Fix an own action x^i. For each opponent action $b \in X^j$, let R_b and T_b be the distributions of $f^i(x^i, b, \omega)$ when the state is drawn according to ρ and q, respectively.

Suppose first that feedback after x^i reveals the opponent's action. The sets $\{f^i(x^i, b, \omega) : \omega \in \Omega\}$ are disjoint across b. Consequently, each observed consequence identifies the opponent's action, and the chain rule for relative entropy gives

$$D_{\mathrm{KL}}\big(Q^i_{\rho,\tau^j}(\cdot \mid x^i) \,\big\|\, Q^i_{q,\widehat{\tau}^j}(\cdot \mid x^i)\big) = D_{\mathrm{KL}}(\tau^j \| \widehat{\tau}^j) + \sum_{b \in X^j} \tau^j(b) D_{\mathrm{KL}}(R_b \| T_b).$$

For every b, condition (ii) states that the map from states to consequences is either injective or constant. In the injective case, relabeling states by their consequences gives $D_{\mathrm{KL}}(R_b \| T_b) = D_{\mathrm{KL}}(\rho \| q)$. In the constant case, the two consequence distributions are the same point mass, so their divergence is zero. Thus the contribution following x^i is

$$D_{\mathrm{KL}}(\tau^j \| \widehat{\tau}^j) + \left(\sum_{b \in X^j} \tau^j(b) \mathbf{1}\{\omega \mapsto f^i(x^i, b, \omega) \text{ is injective}\} \right) D_{\mathrm{KL}}(\rho \| q).$$

Suppose instead that feedback after x^i does not reveal the opponent's action. Condition (i) then makes $f^i(x^i, b, \omega)$ independent of b, so both consequence distributions are independent of the opponent's strategy. The common map from states to consequences is injective or constant by condition (ii). The divergence between the two consequence distributions is therefore $D_{\mathrm{KL}}(\rho \| q)$ in the injective case and zero in the constant case. Because this map is the same for every b, the coefficient of the state divergence equals the sum over b in the preceding formula, while the opponent-strategy term is absent.

The formulas above also hold when one or more divergences are infinite. Terms with zero probability under the distribution in the first argument of the divergence are omitted. If $\tau^j(b) > 0$ but $\widehat{\tau}^j(b) = 0$ and feedback after x^i reveals the opponent's action, both

the divergence between consequence distributions following x^i and the opponent-strategy divergence are infinite. If the opponent-strategy divergence is finite but, for some b with $\tau^j(b) > 0$, R_b assigns positive probability to a consequence assigned zero probability by T_b, both the corresponding conditional term and the divergence following x^i are infinite. In all other cases the chain rule follows by expanding the finite sums. As in the definition of wKLD, a term with zero coefficient contributes zero even when its divergence is infinite.

Multiplying the divergence following each own action by $\tau^i(x^i)$ and summing over x^i gives the coefficients α^i and β^i in the statement. The alternatives in the definition can overlap only when a set is a singleton: if X^j is a singleton, the opponent-strategy divergence is zero; if Ω is a singleton, the state divergence is zero. Hence the stated formula also holds in those cases. □

The decomposition implies

$$\mathsf{P}_i(\rho) \times \{\tau^j\} \subseteq \Theta^i_\rho(\tau) \qquad \text{for every } \rho \in \Delta(\Omega) \text{ and } \tau \in \Sigma. \tag{1}$$

To prove the inclusion, note that $q \mapsto D_{\mathrm{KL}}(\rho\|q)$ is lower semicontinuous on the nonempty compact set Π^i, so $\mathsf{P}_i(\rho)$ is nonempty. This includes the case in which the divergence is infinite for every $q \in \Pi^i$: then $\mathsf{P}_i(\rho) = \Pi^i$. If $\beta^i = 0$, choosing the correct opponent strategy gives zero wKLD for every $q \in \Pi^i$. If $\beta^i > 0$, Lemma 11 guarantees a parameter with finite wKLD. The decomposition then implies that $D_{\mathrm{KL}}(\rho\|q)$ is finite for at least one $q \in \Pi^i$. Its minimum is therefore finite. The parameter formed by any $q \in \mathsf{P}_i(\rho)$ and $\widehat{\tau}^j = \tau^j$ attains $\beta^i \min_{r\in\Pi^i} D_{\mathrm{KL}}(\rho\|r)$, a lower bound on wKLD over all parameters. This proves the inclusion, including the cases in which either coefficient is zero.

Reciprocal projections. For Corollary 8, suppose both feedback functions are all-or-nothing and $\pi^i \in \mathsf{P}_i(p) \cap \mathsf{P}_i(\pi^j)$ for each player. For any $\sigma \in NE(\pi^1, \pi^2)$, inclusion (1) with state distributions p and π^j gives

$$(\pi^i, \sigma^j) \in \Theta^{i,0}(\sigma) \cap \Theta^i_{\pi^j}(\sigma).$$

Proposition 6 implies $NE(\pi^1, \pi^2) \subseteq \mathcal{E}^\infty$, and the Nash set is nonempty because the game is finite. Neither convexity nor informativeness is required: the projection inclusion also covers zero coefficients in the decomposition.

A common model. Corollary 9 adds $\Pi^1 = \Pi^2 = \Pi$. Choose any $\pi \in \mathsf{P}(p)$, which exists because the projection set is nonempty. Since $D(\pi\|\pi) = 0$, taking $\pi^1 = \pi^2 = \pi$ satisfies the reciprocal-projection conditions and yields $NE(\pi) \subseteq \mathcal{E}^\infty \neq \varnothing$.

The necessary condition following that corollary does not require identical models. If $\pi \in \mathsf{P}_1(p) \cap \mathsf{P}_2(p)$, then $\pi \in \Pi^1 \cap \Pi^2$ and $\pi \in \mathsf{P}_i(\pi)$ for both players, so the same choice gives existence. Consequently, under all-or-nothing feedback, $\mathcal{E}^\infty = \varnothing$ requires $\mathsf{P}_1(p) \cap \mathsf{P}_2(p) = \varnothing$. Example 4 shows that this condition is not sufficient.

B.3 Proof of Proposition 10

Proof. Fix an informative profile σ. We first show that, for every player i and every full-support state distribution ρ, the projection $\mathsf{p}_i(\rho)$ is well defined, has full support, and determines the unique minimizing parameter at σ. Informativeness makes both coefficients α^i and β^i in Lemma 7 strictly positive. Lemma 11 guarantees a parameter with finite wKLD at (ρ, σ), so the decomposition implies that $D(\rho\|q) < \infty$ for some $q \in \Pi^i$. Lower semicontinuity and compactness of Π^i therefore give a finite attained minimum of $D(\rho\|\cdot)$. Since ρ has full support, every minimizer has full support. On the set of full-support distributions, $D(\rho\|\cdot)$ is strictly convex; convexity of Π^i thus makes the minimizer unique. Denote it by $\mathsf{p}_i(\rho)$.

The opponent-strategy divergence has the unique minimizer $\widehat{\sigma}^j = \sigma^j$. The two coefficients are positive and the state-distribution minimum is finite, so minimizing the two terms separately gives

$$\Theta^i_\rho(\sigma) = \{(\mathsf{p}_i(\rho), \sigma^j)\}. \tag{2}$$

Consequently, the only belief supported on the original minimizing parameters is $\delta_{(\mathsf{p}_i(\rho),\sigma^j)}$. At higher depths this parameter either satisfies the recursive restrictions or is excluded; parameters with larger wKLD cannot replace it. The response definitions therefore give

$$\sigma^i \in \mathcal{R}^{i,0}(\rho, \sigma^j) \iff \sigma^i \in \mathrm{BR}^i\big(\delta_{(\mathsf{p}_i(\rho),\sigma^j)}\big)\,, \tag{3}$$

$$\sigma^i \in \mathcal{R}^{i,n+1}(\rho, \sigma^j) \iff \begin{cases} \sigma^i \in \mathrm{BR}^i\big(\delta_{(\mathsf{p}_i(\rho),\sigma^j)}\big)\,, \\ \sigma^j \in \mathcal{R}^{j,n}(\mathsf{p}_i(\rho), \sigma^i). \end{cases} \tag{4}$$

These equivalences hold for every full-support ρ and every integer $n \geq 0$. Since each projection preserves full support, they apply successively in the hypothetical problems generated by these point beliefs.

For the finite-depth characterization, fix a starting player i and a full-support state distribution ρ. Set $i_0 = i$, $i_{\ell+1} = 3 - i_\ell$, $q_{-1} = \rho$, and $q_\ell = \mathsf{p}_{i_\ell}(q_{\ell-1})$ for $\ell \geq 0$. We claim that

$$\sigma^i \in \mathcal{R}^{i,n}(\rho, \sigma^j) \iff \sigma^{i_\ell} \in \mathrm{BR}^{i_\ell}\Big(\delta_{(q_\ell,\sigma^{i_{\ell+1}})}\Big) \quad \text{for every } \ell = 0, \ldots, n. \tag{5}$$

For $n = 0$, this is (3). Suppose the equivalence holds at depth n for either starting player and every full-support state distribution. Equation (4) says that depth $n+1$ requires the optimality condition in (5) at $\ell = 0$ and a depth-n response by player i_1 at state distribution q_0. Applying the induction hypothesis to that problem gives exactly the conditions at $\ell = 1, \ldots, n+1$. Conversely, those conditions yield the opponent's depth-n response by the same induction hypothesis, and the condition at $\ell = 0$ then yields the original depth-$n+1$ response. This proves (5) in both directions.

Now take $\rho = p$ and consider the paths starting from each of the two players. Along either path, $q_\ell = \pi_\ell^{i_\ell}$: this holds at $\ell = 0$ by the definition of π_0^i, and the induction step follows from $\pi_\ell^{i_\ell} = \mathsf{p}_{i_\ell}(\pi_{\ell-1}^{i_{\ell-1}})$. At each depth ℓ, the two paths together impose optimality for both players under (π_ℓ^1, π_ℓ^2). Because $\sigma \in \mathcal{E}^n$ is equivalent to $\sigma^i \in \mathcal{R}^{i,n}(p, \sigma^j)$ for both players, (5) gives

$$\sigma \in \mathcal{E}^n \quad \Longleftrightarrow \quad \sigma \in NE(\pi_k^1, \pi_k^2) \quad \text{for every } k = 0, \dots, n. \tag{6}$$

Taking the intersection over all $n \geq 0$ establishes the proposition. The index k counts successive hypothetical problems. The strategies remain fixed throughout this argument, and the projected state distributions need not converge. □

B.4 Verification of Example 4

Proof. Let ρ be the state distribution in the problem and q a subjective state distribution. Write $s = \rho(H)$, $t = q(H)$, $v = \tau^1(V)$, and $a = \tau^2(I)$. For a parameter $(q, \widehat{\tau}^j)$, the payoff differences are

$$U^1(V;(q,\widehat{\tau}^2)) - U^1(C;(q,\widehat{\tau}^2)) = 1 + 2t - 4\widehat{\tau}^2(I), \qquad U^2(I;(q,\widehat{\tau}^1)) - U^2(N;(q,\widehat{\tau}^1)) = 4\widehat{\tau}^1(V) - 1 - 2t.$$

Let $d(s\|t)$ denote the KL divergence between Bernoulli distributions with probabilities s and t. Playing V or I reveals both the state and the opponent's action, whereas C and N always yield zero. Lemma 7 therefore gives

$$K_\rho^1(\tau;(q,\widehat{\tau}^2)) = v\big[d(s\|t) + d(a\|\widehat{\tau}^2(I))\big], \qquad K_\rho^2(\tau;(q,\widehat{\tau}^1)) = a\big[d(s\|t) + d(v\|\widehat{\tau}^1(V))\big].$$

A term with zero coefficient is omitted. If the player uses her informative action with positive probability, every wKLD-minimizing parameter specifies the opponent's strategy correctly and uses a state distribution minimizing KL divergence. If only the uninformative action is played, every parameter minimizes wKLD. The unique state projections are

$$\mathsf{p}_1(\rho)(H) = \max\{s, 3/4\}, \qquad \mathsf{p}_2(\rho)(H) = \min\{s, 1/4\}.$$

Indeed, for $t \in (0,1)$, the derivative of $d(s\|t)$ is $(t-s)/(t(1-t))$, so minimization over the two closed intervals gives these projection formulas, including their endpoints by continuity and the conventions for infinite divergence.

Berk–Nash equilibria. At $p(H) = 1/2$, wKLD minimization following the informative actions gives the unique subjective state distributions $\pi^1(H) = 3/4$ and $\pi^2(H) = 1/4$. First consider profiles in which at least one player uses a pure strategy. If $v = 0 < a$, observations following inspection identify compliance, giving the inspector a payoff difference of $-3/2 < 0$, contradicting $a > 0$. If $v > 0 = a$, observations following violation

identify no inspection, giving the firm a payoff difference of $5/2 > 0$, so $v = 1$. Both (C, N) and (V, N) are Berk–Nash equilibria. At (C, N), the firm can use $\delta_{(\pi^1, I)}$ and the inspector $\delta_{(\pi^2, C)}$; both players observe payoff zero regardless of the state and the opponent's action, so every parameter minimizes wKLD. Under these point beliefs, compliance and no inspection, respectively, are strictly optimal. At (V, N), the firm instead uses the point belief $\delta_{(\pi^1, N)}$, its only belief supported on wKLD-minimizing parameters. The inspector again uses $\delta_{(\pi^2, C)}$.

Now suppose $v > 0$ and $a > 0$. If $a = 1$, violation is strictly suboptimal under the firm's unique wKLD-minimizing parameter, since $5/2 - 4 < 0$. If $v = 1$, inspection is strictly optimal under the inspector's unique wKLD-minimizing parameter, since $4 - 3/2 > 0$, implying the already excluded case $a = 1$. Consequently both probabilities are interior, and the supporting beliefs must make both players indifferent. The equations $5/2 - 4a = 0$ and $4v - 3/2 = 0$ give the unique such profile σ, with $(v, a) = (3/8, 5/8)$. Thus

$$\mathcal{E}^0 = \{(C, N), (V, N), \sigma\}.$$

The first restriction. For every state distribution ρ, the inspector's only Berk–Nash response against compliance is N. If the inspector uses I with positive probability, her observations reveal compliance. Every wKLD-minimizing parameter therefore specifies compliance as the firm's strategy, against which inspection is strictly costly. The candidate N is supported by $\delta_{(\pi^2, C)}$. Therefore $\mathcal{R}^{2,0}(\rho, C) = \{N\}$. When the firm attempts to justify C against N at depth one, every parameter satisfying the first restriction must consequently specify N as the inspector's strategy. Against N, violation is strictly profitable under every distribution in Π^1, so no belief on those parameters makes compliance optimal. We have proved $C \notin \mathcal{R}^{1,1}(\rho, N)$ for every ρ, and in particular $(C, N) \notin \mathcal{E}^1$.

The profile (V, N) satisfies the first restriction. The firm's actual supporting belief is $\delta_{(\pi^1, N)}$. Its conjecture N is a Berk–Nash response in the inspector's problem (π^1, V): the uninformative action N is supported by $\delta_{(\pi^2, C)}$. The inspector's actual supporting belief is $\delta_{(\pi^2, C)}$. Its conjecture C is a Berk–Nash response in the firm's problem (π^2, N): the uninformative action C is supported by $\delta_{(\pi^1, I)}$. Each of these beliefs assigns probability one to a wKLD-minimizing parameter in the corresponding problem and makes the stated pure strategy strictly optimal. Thus $(V, N) \in \mathcal{E}^1$.

The second restriction. Against N, if the firm uses V with positive probability, its observations following V reveal that the inspector uses N. Violation is then strictly optimal under every minimizing parameter, so a Berk–Nash response assigning positive probability to violation must be V itself. The preceding argument excludes the only other possibility, C, at depth one. Moreover, V remains a depth-one response: its minimizing parameter specifies N as the inspector's strategy, which is a Berk–Nash response in the inspector's generated problem, supported by a conjecture of C. Hence $\mathcal{R}^{1,1}(\rho, N) = \{V\}$

for every state distribution ρ. At (V, N), the inspector's depth-two supporting belief would therefore have to assign probability one to parameters conjecturing V. All such parameters give inspection payoff difference $4 - 1 - 2t \geq 5/2 > 0$, since $t \leq 1/4$. No such belief makes N optimal, and $(V, N) \notin \mathcal{E}^2$.

The mixed profile. The projection formulas give $\mathsf{p}_1(p) = \mathsf{p}_1(\pi^2) = \pi^1$ and $\mathsf{p}_2(p) = \mathsf{p}_2(\pi^1) = \pi^2$. The indifference calculations show that $\sigma \in NE(\pi^1, \pi^2)$. Corollary 8 therefore applies, giving $\sigma \in \mathcal{E}^\infty$. Combining this conclusion with nestedness and the preceding exclusions proves

$$\mathcal{E}^1 = \{(V, N), \sigma\}, \qquad \mathcal{E}^2 = \mathcal{E}^\infty = \{\sigma\}.$$

Positive minimum wKLD. At σ, the minimum wKLD values in the actual problems of the firm and inspector are, respectively,

$$\frac{3}{8} d(1/2 \| 3/4) = \frac{3}{16} \ln \frac{4}{3} > 0, \qquad \frac{5}{8} d(1/2 \| 1/4) = \frac{5}{16} \ln \frac{4}{3} > 0.$$

In the firm's hypothetical problem generated by the inspector's supporting parameter, the state distribution is π^2; in the inspector's hypothetical problem generated by the firm's supporting parameter, the state distribution is π^1. The corresponding minimum wKLD values are

$$\frac{3}{8} d(1/4 \| 3/4) = \frac{3}{16} \ln 3 > 0, \qquad \frac{5}{8} d(3/4 \| 1/4) = \frac{5}{16} \ln 3 > 0.$$

The same two hypothetical problems recur at every further depth. Thus the beliefs in this construction satisfy the optimality and recursive requirements even though minimum wKLD is positive in both actual and hypothetical problems. □

B.5 Finite detection and local persistence of nonexistence

For the perturbation statement, fix the finite state, action, and consequence spaces, both feedback functions, and both subjective parameter spaces. Only the objective state distribution p and the payoff functions $u = (u^1, u^2)$ vary. The payoff functions remain functions of own actions and observed consequences. Write $\mathcal{E}^n(p, u)$ for the equilibrium set under state distribution p and payoffs u.

Proposition 13 (Finite detection and local persistence)**.** *Under Assumption 1, if $\mathcal{E}^\infty(p, u) = \varnothing$, then $\mathcal{E}^N(p, u) = \varnothing$ for some finite integer N. Moreover, there is a neighborhood of (p, u) such that $\mathcal{E}^N(p', u') = \varnothing$, and hence $\mathcal{E}^\infty(p', u') = \varnothing$, for every (p', u') in that neighborhood with p' of full support.*

Proof. By Lemma 2 and Proposition 12, the sets $\mathcal{E}^n(p, u)$ are nested and compact. If they were all nonempty, their intersection would be nonempty. Hence $\mathcal{E}^N(p, u) = \varnothing$ for some finite N.

The proof of Proposition 12 also applies when payoffs vary: expected-payoff comparisons are continuous in u, and the wKLD minimizers do not depend on u. Thus $(p,u) \mapsto \mathcal{E}^N(p,u)$ has a closed graph. If emptiness failed to persist, there would be $(p_k,u_k) \to (p,u)$ and $\sigma_k \in \mathcal{E}^N(p_k,u_k)$. A subsequence converges to a profile σ, which would then belong to $\mathcal{E}^N(p,u)$, a contradiction. Since $\mathcal{E}^\infty(p',u') \subseteq \mathcal{E}^N(p',u')$, the infinite-depth conclusion follows. □

This proposition fixes the subjective models and feedback functions. It does not assert persistence under changes in the parameter spaces or in what players observe. In particular, when feedback is own payoff in an example, changing the utility assigned to an observed consequence while keeping its signal fixed is a different perturbation from changing which consequences are distinguishable.

C Comparisons with other solution concepts

C.1 Esponda's rationalizable conjectural equilibrium

Let $\Pi \subseteq \Delta(\Omega)$ be nonempty and compact, with $p \in \Pi$. To apply Esponda (2013), take Π as the common fundamental space, p as the true fundamental, and Σ^i as player i's strategy space. A strategy in this representation is a mixed strategy of the original game. On $W = \Pi \times \Sigma$, define payoffs and feedback by

$$V^i(\eta;q) = \sum_{x^i \in X^i} \eta^i(x^i) U^i(x^i;(q,\eta^j)),$$

$$P^i(\rho,\tau) = \left\{(q,\eta) \in W : \eta^i = \tau^i,\ Q^i_{q,\eta^j}(\cdot \mid x^i) = Q^i_{\rho,\tau^j}(\cdot \mid x^i) \text{ for every } x^i \in \operatorname{supp} \tau^i\right\}.$$

The partition records the player's own mixed strategy and the consequence distribution after each played action, following Esponda (2013, Section 5.3, footnotes 35–36). The fundamental and the opponent's strategy may be statistically dependent under a player's belief. A single belief must make her mixed strategy optimal. The spaces are compact, payoffs are continuous, and the feedback restrictions are the continuous equalities $\eta^i = \tau^i$ and $\tau^i(x^i)[Q^i_{q,\eta^j}(y^i \mid x^i) - Q^i_{\rho,\tau^j}(y^i \mid x^i)] = 0$. Thus the represented game satisfies Esponda (2013, Definition 6).

For $\lambda^i \in \Delta(W)$, let $\Phi^i(\lambda^i)$ be the strategies maximizing expected V^i under the marginal of this belief over the fundamental and the opponent's strategy. Define

$$W_0 = W, \qquad W_{k+1} = \left\{w = (\rho,\tau) \in W : \forall i\ \exists \lambda^i \in \Delta(W),\ \lambda^i(P^i(w) \cap W_k) = 1,\ \tau^i \in \Phi^i(\lambda^i)\right\}.$$

The sets are nested and compact by monotonicity and Esponda (2013, Lemma 3(ii)). His Lemma 4 establishes the existence of supporting beliefs for states in the intersection, so

his Theorems 1–2, with the compact-game extension in Section 6.1, give

$$\mathcal{E}^{\mathrm{RCE}}_{\Pi}(p) = \left\{ \sigma \in \Sigma : (p,\sigma) \in \bigcap_{k\geq 0} W_k \right\}.$$

Write $\mathcal{R}^{i,n}_{\Pi}$ and $\mathcal{E}^{n}_{\Pi}$ for this paper's response and equilibrium sets when both models equal Π, and set $\mathcal{E}^{\infty}_{\Pi} = \bigcap_n \mathcal{E}^{n}_{\Pi}$. The state distribution in every actual or hypothetical problem generated by the recursion belongs to Π. For any such state distribution ρ and profile τ, the parameter (ρ, τ^j) gives zero wKLD, so the minimizing set is exactly

$$F^i_{\Pi}(\rho,\tau) = \left\{ (q,b) \in \Pi \times \Sigma^j : K^i_{\rho}(\tau;(q,b)) = 0 \right\}.$$

It consists of the parameters reproducing the consequence distributions after each played action. The common-model comparison therefore does not require Assumption 1(ii).

Proposition 14 (Equivalence under a common model)**.** *Let $\Pi \subseteq \Delta(\Omega)$ be nonempty and compact, with $p \in \Pi$, and use the preceding payoff and feedback specification. For every $n \geq 0$,*

$$W_{n+1} = \left\{ (\rho,\tau) \in \Pi \times \Sigma : \tau^i \in \mathcal{R}^{i,n}_{\Pi}(\rho,\tau^j) \text{ for both players } i \right\}.$$

Consequently, if $\Pi^1 = \Pi^2 = \Pi$, then $\mathcal{E}^{\mathrm{RCE}}_{\Pi}(p) = \mathcal{E}^{\infty}_{\Pi}$.

Proof. For $w = (\rho,\tau)$, each parameter $(q,b) \in F^i_{\Pi}(\rho,\tau)$ corresponds to exactly one state in $P^i(w)$: its fundamental is q, its player-i strategy is τ^i, and its player-j strategy is b. The corresponding beliefs over parameters and states give the same expected payoff to every strategy of player i. Thus τ^i is optimal under one belief if and only if it is optimal under the other.

If $w' = (q,\eta) \in P^i(w)$, then $\eta^i = \tau^i$ and $F^i_{\Pi}(q,\eta) = F^i_{\Pi}(\rho,\tau)$. For a parameter (r,b) in this common set, the recursive restriction concerns whether b is justified in player j's problem (r,τ^i), which is also the same at w and w'. Hence, for every $m \geq 0$,

$$\tau^i \in \mathcal{R}^{i,m}_{\Pi}(\rho,\tau^j) \quad \Longleftrightarrow \quad \eta^i \in \mathcal{R}^{i,m}_{\Pi}(q,\eta^j).$$

Let A_n be the set on the right-hand side of the proposition's first equality. The belief correspondence gives $W_1 = A_0$. Suppose $W_{n+1} = A_n$. At $w \in A_{n+1}$, player i has a supporting belief on zero-loss parameters (r,b) with $b \in \mathcal{R}^{j,n}_{\Pi}(r,\tau^i)$. The corresponding states satisfy player j's depth-n requirement. Player i's strategy satisfies the depth-n requirement at w by nestedness. The preceding equivalence shows that it also satisfies that requirement at each of these states. Thus the corresponding belief is supported on $P^i(w) \cap A_n = P^i(w) \cap W_{n+1}$. For both players this gives $w \in W_{n+2}$.

Conversely, a belief supporting player i at $w \in W_{n+2}$ is supported on $P^i(w) \cap A_n$. The corresponding belief over parameters is supported on $F^i_{\Pi}(\rho,\tau)$ and satisfies $b \in \mathcal{R}^{j,n}_{\Pi}(r,\tau^i)$

for every parameter (r, b) in its support. This belief therefore establishes $\tau^i \in \mathcal{R}_\Pi^{i,n+1}(\rho, \tau^j)$. Applying this to both players proves $W_{n+2} = A_{n+1}$. The equilibrium equality follows by setting $\rho = p$ and taking intersections. □

Corollary 15 (Inclusion when the models differ)**.** *Suppose the players' subjective models satisfy Assumption 1 and $p \in \Pi^1 \cap \Pi^2$. For $\Pi = \Pi^1 \cap \Pi^2$, the RCE benchmark defined above satisfies*

$$\mathcal{E}_\Pi^{\mathrm{RCE}}(p) \subseteq \mathcal{E}^\infty.$$

Proof. The intersection Π is compact and contains p, so Proposition 14 applies. At any state distribution $\rho \in \Pi$, a player's supporting belief under the common model assigns probability one to parameters in her original parameter space that have zero wKLD. These parameters also minimize wKLD in the larger models, and the same belief preserves all payoff comparisons. This proves $\mathcal{R}_\Pi^{i,0}(\rho, \tau^j) \subseteq \mathcal{R}^{i,0}(\rho, \tau^j)$. If the inclusion holds at depth n, every opponent strategy justified at that depth under the common model is also justified under the original models. Since the state distribution in each further hypothetical problem belongs to Π, the same belief, supported on parameters with zero wKLD, proves the inclusion at depth $n + 1$. At $\rho = p$, intersecting over depths and applying the proposition gives $\mathcal{E}_\Pi^{\mathrm{RCE}}(p) = \mathcal{E}_\Pi^\infty \subseteq \mathcal{E}^\infty$. The intersection need not inherit Assumption 1(ii): the argument uses compactness of Π and zero loss at every problem generated within it. □

C.2 The inspection comparison

Keep the objective game and feedback in Example 4, but replace the subjective models by $\Pi^1 = \Pi^2 = \Delta(\Omega)$. Write $\mathcal{E}^{n,R}$ for the resulting equilibrium sets and let σ_R have $\sigma_R^1(V) = \sigma_R^2(I) = 1/2$. In this common-model specification, observations following an informative action identify both the opponent's strategy and the state distribution, so the player's minimum wKLD is zero. At $p(H) = 1/2$, a player who uses her informative action with positive probability therefore evaluates the expected scale of activity at 2. The boundary argument in Appendix B.4 again leaves only (C, N) and (V, N), while any profile using both informative actions must be interior and satisfy $2 - 4a = 0$ and $4v - 2 = 0$. Thus $\mathcal{E}^{0,R} = \{(C, N), (V, N), \sigma_R\}$.

The first- and second-depth arguments in Appendix B.4 also apply to the enlarged models. They use only that inspection against compliance is strictly costly, violation against no inspection is strictly profitable, and inspection against violation is strictly profitable. These inequalities hold for every $q \in \Delta(\Omega)$ because $1 \leq \mathbb{E}_q[r] \leq 3 < 4$. Consequently (C, N) fails the first restriction, while (V, N) satisfies the first restriction and fails the second. The point beliefs $\delta_{(p,\sigma_R^j)}$ assign probability one to parameters specifying the true state distribution and the opponent's strategy. They support σ_R in

the actual and hypothetical problems at every depth. Hence

$$\mathcal{E}^{1,R} = \{(V, N), \sigma_R\}, \qquad \mathcal{E}^{2,R} = \mathcal{E}^{\infty,R} = \{\sigma_R\}.$$

The common-model equivalence proved in Appendix C.1 gives $\mathcal{E}^{\mathrm{RCE}}_{\Delta(\Omega)}(p) = \{\sigma_R\}$. The restricted models of the example instead give $\mathcal{E}^{\infty} = \{\sigma\}$, where $(\sigma^1(V), \sigma^2(I)) = (3/8, 5/8)$, with the positive minimum wKLD values in the actual and hypothetical problems calculated in Appendix B.4. The comparison therefore changes the subjective models while keeping the objective game and feedback fixed.