Machine-Learning to Trust*

Ran Spiegler†

May 7, 2025

Abstract

A sequence of agents with one-period recall play an overlapping-generations Prisoner's Dilemma with state-dependent payoffs. players' belief regarding others' behavior is a "coarse fit" of the true population strategy w.r.t a partition of the relevant contingencies. In equilibrium, the partition minimizes the sum of the mean squared prediction error and a complexity penalty on the partition size; and players best-reply to the belief. The scope for cooperation is significantly reduced under this solution concept, relative to symmetric mixed-strategy Nash equilibrium.

*Financial support from UKRI Frontier Research Grant no. EP/Y033361/1 is gratefully acknowledged. I thank Alex Clyde, Philippe Jehiel, Nir Rosenfeld and Yair Weiss for useful conversations and comments.

†Tel Aviv University and University College London

1 Introduction

With the growing ubiquity of machine-learning (ML) algorithms, economist are becoming increasingly interested in how ML affects strategic interactions. Of particular interest are dynamic strategic interactions, where players may use ML to learn how to respond to opponents' history-dependent behavior — which itself may be the product of algorithmic learning. For example, as oligopolists adopt ML algorithms for pricing decisions, a natural question of economic importance is how this tendency impacts the scope for collusive pricing.

ML algorithms broadly fall into two paradigms, based on whether constructing a predictive model of the environment is part of what the algorithm does. Model-free methods such as reinforcement learning are based on a direct association between actions and feedback, without building a model of how actions map into outcomes. In contrast, model-based methods learn a predictive model of the environment's data-generating process. The space of models that the algorithm explores may be highly structured and interpretable (e.g., penalized linear regression or probabilistic graphical models) or loose and uninterpretable (e.g., contemporary LLMs).

So far, the economic literature on dynamic strategic interactions between ML algorithms has focused on model-free ML. Specifically, it has examined the behavior that arises when players use reinforcement-learning algorithms. In contrast, there have been no attempts to analyze models of long-run cooperation when players react to beliefs shaped by model-based ML. This paper is a step in this direction.

Studying dynamic games between players who rely on model-based ML to form beliefs is of interest, not only because of its potential economic relevance, but also because it raises a fundamental conceptual question about how ML performs on endogenous datasets. An integral aspect of ML belief formation is the penalty on complexity to avoid overfitting and thus lower the variance of the algorithm's predictions. In the context of dynamic strategic interactions, this penalty on complexity may take the form of pooling different contingencies (exogenous states of Nature, extensive-game histories) as if they are equivalent. This consideration implies two sources of tension between what the model-based-ML and individual-incentive perspectives regard as important distinctions about opponents' play.

First, for a model-based ML algorithm, a rare event is typically treated as unimportant, and the algorithm will tend to group it with other events to improve prediction. In contrast, from the individual-incentive perspective, an event is important if it affects the player's best-replying action. For instance, in the repeated Prisoner's Dilemma, the threat of defection sustains cooperation when it is a counterfactual (or at least rate) event. Thus, the two perspectives differ in how they link an event's importance to its frequency.

Second, when opponents' behavior is similar in two contingencies, model-based-ML-generated beliefs will tend to treat the two contingencies as equivalent for predicting the opponent's future behavior. However, from the point of view of individual incentives, two contingencies should be classified as equivalent if they induce the same best-replying actions. Two beliefs can be close for prediction purposes yet radically different in terms of the best-reply they induce. Thus, the two perspectives have different notions of what makes two contingencies similar and thus worth pooling.

The challenge, then, is that model-based ML is designed to make successful predictions, not to reach relevant classifications for how to respond to these predictions. This raises the central question: How does this misalignment between the model-based ML and individual-incentive perspectives affect the possibility of sustaining long-run cooperation in dynamic strategic interactions?

I explore this question in the context of a simple discrete-time, infinite-horizon game. At every period t, a distinct player plays a Prisoner's Dilemma with his immediate successor, player $t+1$. The game's payoffs are determined by a random, commonly observed state θ, which specifies the ratio between the cost of cooperating and the benefit from receiving cooperation from one's opponent. Players can condition their actions on the most recent action. This can be viewed as an overlapping-generations trust game, in which players' task is to predict how their immediate successor's rate of cooperation will depend on their own action. Symmetric mixed-strategy Nash equilibrium can sustain arbitrarily high cooperation rates using a probabilistic Tit-for-Tat strategy, such that at every payoff state, players are always indifferent between cooperating and defecting.

I adhere to the equilibrium modeling approach in analyzing the dynamic trust game, while modifying the consistency criterion of equilibrium beliefs, in a way that closely resembles Jehiel and Weber (2024). The modification captures in stylized form a common element of ML, namely explicitly trading off a model's quality of empirical fit against its complexity (e.g., see Hastie et al. (2009)). Specifically, I assume that players' belief is formed according to a partition of all relevant contingencies (i.e., combinations of a payoff state and the recalled history). The belief associated with a partition cell is the average behavior in that cell, as in Jehiel (2005). The average is taken w.r.t the *ergodic distribution* over contingencies that is induced by the population-level mixed strategy. An *ML-optimal* partition is required to *minimize* the sum of two terms: (1) the Mean Squared Prediction Error (MSPE) of the partition-induced belief, calculated according to the ergodic distribution; and (2) the partition's complexity, which is proportional to its size. A partition is strongly ML-optimal when it also satisfies a refinement that essentially imposes continuity on the classification of zero-probability contingencies. When the population-level mixed strategy always assigns best-replies to beliefs induced by (strongly) ML-optimal partitions, we have a *(strong) ML equilibrium*.

The paper's main message is that the equilibrium criterion that penalizes complex beliefs can drastically limit the scope for cooperation in the dynamic trust game, for the two reasons mentioned above. Cooperative behavior relies on the threat to lower the cooperation rate following defection. When the cooperation rate is high, this means that the threat is rarely realized. A preference for simple beliefs can lead to grouping such rare events with more frequent ones, thus destroying the incentive to cooperate. The same preference can also lead players to assign the same belief to different contingencies at which the actual strategy prescribes similar behavior. Yet, distinguishing between these contingencies may be crucial for maintaining equilibrium incentives.

In Section 3, I provide full characterization of the maximal cooperation rate that can be sustained in equilibrium for simple specifications of the model. In particular, I demonstrate that the ML-optimality has more drastic implications than an alternative requirement that beliefs are measurable w.r.t a small partition. Another unusual effect is that lower costs of cooperative behavior can actually make it harder to sustain cooperation in equilibrium. In Section 4, I present necessary conditions on the sustainability of positive cooperation rates in strong ML equilibrium. An immediate corollary is that for a fixed complexity cost, the sustainable rate of cooperation vanishes as the number of payoff states grows. In this sense, trust between ML algorithms is harder to achieve in complex environments.

Related literature

The model in this paper synthesizes ideas from two strands in the behavioral game theory literature, and applies them to the question of sustaining cooperation in long-run interactions.

First, Jehiel and Mohlin (2024) and Jehiel and Weber (2024) take Jehiel's (2005) notion of Analogy-Based Expectations Equilibrium, which captures strategic behavior under coarse beliefs, as a starting point. They then apply basic ideas from the ML literature to endogenize the analogy partitions that underlie players' coarse equilibrium beliefs. In Jehiel and Mohlin (2024), partition cells are shaped by an exogenous notion of similarity between contingencies. However, they also respond to the equilibrium frequency of contingencies — in the spirit of the bias-variance trade-off that is fundamental to the ML literature — such that infrequent contingencies are more likely to be grouped together.[1] In Jehiel and Weber (2024), which is the closest precedent for MLEQ, partition size is fixed at some K, and stability of partitions is determined by (local or global) minimization of MSPE. The global version of the solution concept in Jehiel and Weber (2024) can be viewed as a variant of MLEQ, in which the cost of a partition is 0 when its size is weakly below K, and ∞ above it.

Second, Spiegler (2002,2004,2005), Eliaz (2003) and Maenner (2008) introduced the idea that players use simplicity as a belief-selection criterion in dynamic games. The solution concepts defined in these papers capture the idea that an Occam's Razor principle may rule out off-path threats as part of the explanation of opponents' behavior. Unlike the present paper, these concepts involved deterministic beliefs in strategies that take the form of finite automata, and employed complexity measures that rely on this representation, following the tradition of Rubinstein (1986).

As mentioned earlier, the question of how the use of ML algorithms

[1] Mohlin (2014) studies single-agent decision-making with endogenous formation of coarse beliefs based on MSPE minimization.

affects collusive behavior in dynamic games has received much attention recently. This rapidly growing literature has entirely focused on model free, reinforcement-learning-based algorithms; in this sense, it is orthogonal to the present paper. Therefore, here I make do with mentioning a handful of papers. Calvano et al. (2020) used numerical experiments to demonstrate that a repeated oligopolistic pricing game leads to collusive behavior by players who follow a reinforcement-learning model known as Q-learning. Hansen et al. (2021), Banchio and Mantegazza (2022) and Brown and Mackay (2023) extended the numerical mode of analysis to other games, and also made progress in terms of analytical characterizations. Waizmann (2024) studied an interaction between a long-run player who obeys Q-learning and a sequence of rational short-run players.[2]

Finally, Danenberg and Spiegler (2024) studied the dynamic trust game to illustrate a solution concept according to which players form beliefs by extrapolating naively from representative finite samples drawn from the equilibrium distribution.

2 A Model

Time is discrete and infinite: $t = 0, 1, 2, 3, ...$ At each period t, a distinct player (also denoted t) chooses an action $a_t \in \{0, 1\}$. Player 0 is a dummy whose action is exogenously random. For every $t > 1$, player t's payoff only depends on his action and the action taken by the subsequent player $t + 1$. Specifically, the payoff function is $u_t(a_t, a_{t+1}) = a_{t+1} - \theta a_t$, where $\theta \in (0, 1)$ is publicly observed and drawn uniformly from the set $\Theta = \{\theta_1, ..., \theta_n\}$ at period 0. Thus, for any given θ, the infinite-horizon game is an overlapping-generations Prisoner's Dilemma with complete information. Each player $t > 1$ only observes θ and a_{t-1} prior to taking his action. I refer to a_{t-1} as the *observed history* at period t, and use h_t to denote it. Let $H = \{0, 1\}$ denote the set of possible observed histories. I refer to a pair (θ, h) as a *contingency.*

I will be interested in symmetric strategy profiles, where all players $t > 1$ obey the same mixed strategy σ, such that $\sigma(a \mid \theta, h_t)$ is the probability that

[2] An older literature in evolutionary game theory examined repeated games when players use reinforcement learning to adapt their actions over time — e.g., see Bendor et al. (2001).

each player $t > 1$ plays a in the contingency (θ, h_t). I adopt a population interpretation of σ. I often use the shorthand notation $\sigma(\theta, h)$ for $\sigma(1 \mid \theta, h)$. Under the belief that player $t+1$ follows a strategy $\hat{\sigma}$, player t's expected payoff from playing a is $\hat{\sigma}(\theta, a) - \theta a$. Therefore, $a = 1$ (0) is a best-reply to $\hat{\sigma}$ if $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0)$ is weakly above (below) θ.

For every payoff state θ, a strategy σ induces a two-state Markov process over observed histories, where the probability of transition from h_t to h_{t+1} is $\sigma(h_{t+1} \mid \theta, h_t)$. The long-run frequency of $a = 1$ is the invariant probability of $h = 1$ induced by the Markov process. The joint long-run distribution $p_\sigma \in \Delta(\Theta \times H)$ induced by σ is:

$$p_\sigma(\theta, 1) = \frac{\sigma(\theta, 0)}{n[\sigma(\theta, 0) + 1 - \sigma(\theta, 1)]}$$

and $p_\sigma(\theta, 0) = \frac{1}{n} - p_\sigma(\theta, 1)$. We say that σ induces a *positive cooperation rate* in θ if $p_\sigma(\theta, 1) > 0$. The *overall cooperation rate* under σ is $\sum_\theta p_\sigma(\theta, 1)$.

Benchmark: Symmetric Nash equilibrium

As usual, there is a Nash equilibrium in which players never exhibit trust: $\sigma(\theta, h) = 0$ for every (θ, h). Let us explore other symmetric equilibria. Fix the strategy σ. If $\sigma(\theta, 1) - \sigma(\theta, 0) > \theta$ $(< \theta)$ for some θ, then any player's unique best-reply at θ is $a = 1$ $(a = 0)$, regardless of his observed history — but this contradicts the optimality of $\sigma(\theta, 0) < 1$ $(\sigma(\theta, 1) > 0)$. However, if $\sigma(\theta, 1) - \sigma(\theta, 0) = \theta$, then players are always indifferent between the two actions, such that adhering to σ is consistent with best-replying.

It follows that any σ that satisfies $\sigma(\theta, 1) - \sigma(\theta, 0) = \theta$ for every θ is a symmetric Nash equilibrium strategy. In particular, we can set $\sigma(\theta, 1) = 1$ and $\sigma(\theta, 0) = 1 - \theta$ for every θ, such that the induced long-run distribution p_σ satisfies $p_\sigma(\theta, 1) = 1/n$ for every θ — i.e., players fully cooperate in equilibrium. ■

Let us now introduce the novel model of equilibrium belief formation, which draws inspiration from the idea that beliefs are extrapolated from historical data using ML methods. Fix the strategy σ and its induced long-run distribution p_σ. Let Π be a partition of $\Theta \times H$. Let $\pi(\theta, h)$ denote the partition cell that includes (θ, h). A cell $\pi \in \Pi$ is non-null if $p_\sigma(\theta, h) > 0$ for some $(\theta, h) \in \pi$. Denote

$$p_\sigma(\pi) = \sum_{(\theta,h)\in\pi} p_\sigma(\theta, h)$$

The *representative strategy* of a non-null cell $\pi \in \Pi$ is

$$\hat{\sigma}(\pi) = \sum_{(\theta,h)\in\pi} \frac{p_\sigma(\theta, h)}{p_\sigma(\pi)} \sigma(\theta, h)$$

This is the expected strategy conditional on being in π, where the expectation is taken w.r.t p_σ. When Π is fixed and there is no risk of confusion, I will use the notation $\hat{\sigma}(\theta, h)$ as a shorthand for $\hat{\sigma}(\pi(\theta, h))$.

Define the function

$$V_{c,\sigma}(\Pi) = c \cdot |\Pi| + \sum_{(\theta,h)} p_\sigma(\theta, h)[\hat{\sigma}(\pi(\theta, h)) - \sigma(\theta, h)]^2$$

where $c > 0$ is a constant capturing the cost of belief complexity.

In the spirit of machine-learning classification algorithms, $V_{c,\sigma}$ trades off two quantities: (1) A classification's predictive accuracy, represented by the second term in the objective function, which is simply the *mean squared prediction error* (**MSPE**) of the representative strategy induced by the partition; and (2) the classification's complexity, measured by the partition size.

Definition 1 (ML-optimality) *A partition Π is ML-optimal w.r.t p_σ if it minimizes $V_{c,\sigma}(\Pi)$.*

Definition 2 (ML Equilibrium) *A strategy-partition pair (σ, Π) is an ML equilibrium if: (i) Π is ML-optimal w.r.t p_σ; and (ii) if $\sigma(a \mid \theta, h) > 0$, then a is a best-reply to $\hat{\sigma}(\pi(\theta, h))$.*

Thus, in ML Equilibrium (**MLEQ** in short), players' strategy prescribes best-replies to a belief $\hat{\sigma}$, which in turn is induced by an ML-optimal partition w.r.t the long-run distribution induced by the players' strategy.

It is clear that if we assumed $c = 0$, then an ML-optimal Π would be maximally fine, such that $\hat{\sigma}(\pi(\theta, h)) = \sigma(\theta, h)$ for every (θ, h), and MLEQ would collapse to symmetric mixed-strategy Nash equilibrium. We will see that when $c > 0$, MLEQ departs from Nash equilibrium in significant ways.

Clearly, the zero-trust strategy $\sigma(\theta, h) = 0$ for every (θ, h) is consistent with MLEQ. To see why, note that under this σ, players never vary their behavior with the contingency. As a result, a degenerate partition of size 1 induces zero MSPE, hence it is trivially ML-optimal. This partition induces $\hat{\sigma}(\pi(\theta, h)) = 0$ for every (θ, h), such that players' unique best-reply is $a = 0$, as postulated. Our main problem will be to explore the possibility of SMLEQ with positive cooperation rates.

The following simple property of ML-optimal partitions will be applied repeatedly in the sequel.

Lemma 1 *Suppose* Π *is ML-optimal w.r.t* p_σ*. Then, the following inequality holds for every two cells* $\pi, \pi' \in \Pi$*:*

$$\frac{p_\sigma(\pi) p_\sigma(\pi')}{p_\sigma(\pi) + p_\sigma(\pi')} \left(\hat{\sigma}(\pi) - \hat{\sigma}(\pi')\right)^2 \geq c \tag{1}$$

The L.H.S of (1) represents the MSPE increase when we deviate from Π to a new partition that merges the cells π and π' into a single cell. The formula's derivation appears in texts on clustering algorithms — e.g., see Kaufman and Rousseeuw (1990, pp. 230-231).[3] The R.H.S of (1) is the complexity-cost reduction that merging the two cells brings. ML-optimality requires the former to be weakly above the latter.

Corollary 1 *Suppose* Π *is ML-optimal w.r.t* p_σ*. Then,* $\hat{\sigma}(\pi) \neq \hat{\sigma}(\pi')$ *for every distinct* $\pi, \pi' \in \Pi$*.*

This corollary immediately follows from Lemma 1. If $\hat{\sigma}(\pi) = \hat{\sigma}(\pi')$, then we can merge the cells π and π' into one cell, thus lowering the partition's complexity without changing its MSPE.

[3] It is easy to derive it using the variance decomposition formula.

Optimal assignment and Jehiel and Weber (2024)

The definition of MLEQ is closely related to the notion of Clustered Analogy-Based Expectations Equilibrium (Jehiel and Weber (2024)). In both cases, players' equilibrium beliefs are extrapolated from the objective distribution via an ML-inspired procedure. There are two differences. First, Jehiel and Weber examine static games, where the objective distribution over contingences is exogenous. In contrast, the present game is dynamic, and so the objective distribution over $\Theta \times H$ is endogenously induced by the equilibrium strategy. Second, Jehiel and Weber fix the partition size, whereas partition size in the present model is variable and traded off against the prediction error the partition induces. The significance of these two differences will be clarified by the examples in the next section.

Nevertheless, we will be able to make use of a simple observation due to Jehiel and Weber (2024).

Definition 3 (Optimal assignment) *A contingency* (θ, h) *is optimally assigned w.r.t* (σ, Π) *if* $\pi(\theta, h) \in \arg\min_{\pi \in \Pi} |\hat{\sigma}(\pi) - \sigma(\theta, h)|$.

Remark 2 *Suppose that* Π *is ML-optimal w.r.t* p_σ*. If* $p_\sigma(\theta, h) > 0$*, then is optimally assigned w.r.t* (σ, Π)*.*

This result is adapted from Lemma 1 in Jehiel and Weber (2024, Appendix B). It means that under an ML-optimal partition, every contingency in the support of p_σ is assigned to a partition cell having the nearest representative strategy. This is not a trivial observation, given that reassigning contingency from one cell to another may alter the cells' representative strategies. Jehiel and Weber's result is adapted to the present setting. First, partition size is variable in the present paper, whereas it is fixed in Jehiel and Weber (2024). However, it is clear that if a partition is ML-optimal, then it also minimizes the mean squared prediction error among all partitions that share its size. Second, while Jehiel and Weber can assume w.l.o.g that all contingencies have positive probability, this is not guaranteed in the present context because p_σ is endogenous. Motivated by this observation, I present the following definition, which extends the optimal assignment property to zero-probability contingencies.

Definition 4 (Strong ML-optimality) *A partition* Π *is strongly ML-optimal w.r.t* p_σ *if it is ML-optimal w.r.t* p_σ*, and if every contingency is optimally assigned w.r.t* (Π, σ)*.*

Definition 5 (Strong MLEQ) *A MLEQ* (σ, Π) *is strong if* Π *is strongly ML-optimal w.r.t* p_σ*.*

In what follows, I use the abbreviation **SMLEQ** to describe a strong MLEQ.

In Jehiel and Weber (2024), the optimal assignment property is related to the variant on their solution concept, in which partitions are not required to minimize mean squared prediction error, but instead contingencies are required to be optimally assigned. Partitions that satisfy this criterion can be obtained via a simple iterative procedure, known as Lloyd's algorithm (Lloyd (1975)). By contrast, finding an ML-optimal partition (even when we hold the partition size fixed) is a computationally hard problem.

Discussion

I conclude the section with a discussion of two aspects of the model. First, mixed strategies are crucial for the dynamic trust game. As we saw, cooperation cannot be sustained in symmetric pure-strategy Nash equilibrium. This is due to the combination of the game's sequential-move and bounded-recall aspects. However, with mixed strategies, any strategy σ that satisfies the indifference condition $\sigma(\theta, 1) - \sigma(\theta, 0) = \theta$ for every θ is consistent with Nash equilibrium. Thus, with mixed strategies, Nash equilibrium poses no restriction on the ability to sustain long-run cooperation. As we will see, indifference conditions continue to play a key role in the analysis of MLEQ.

Second, the notion of ML-optimality imposes a penalty on complex beliefs even though the MSPE is calculated w.r.t the actual long-distribution p_σ over contingencies. In practice, the underlying rationale for ML methods that explicitly penalize complexity is that empirical fit is calculated against a finite sample, and therefore penalizing complexity is required to mitigate over-fitting. When the sample is infinite (which is implicitly the case in our model), this rationale vanishes.

Therefore, the appropriate way to interpret the preference for simple beliefs in MLEQ is that it is a tractability-motivated reduced form of a more elaborate model in which players form their beliefs on the basis of finite samples drawn from the endogenous distribution. Danenberg and Spiegler (2024) is an example of such a more elaborate model, albeit without a simplicity-seeking component. Note, however, that incorporating finite samples into the present model would not only make it less tractable, but also clash with the central role of the indifference condition that characterizes players' beliefs. When players form beliefs according to a finite sample, their beliefs will almost surely generate a strict preference for one action or another, and this will destroy the incentive to play a history-dependent strategy — which is weak and relies on indifferences in our model. Thus, combining the incentive structure of the dynamic trust game with a belief-formation model that relies on finite samples is challenging.

3 Examples

In this section I present two examples that illustrate MLEQ and demonstrate how the penalty on complex beliefs, inherent in the notion of ML-optimality, constrains the ability to sustain trust in MLEQ.

3.1 Example I: $n = 1$

This is the simplest specification of the model. When there is only one cost value $\theta \in (0, 1)$, there are only two possible contingencies, hence a partition size of 2 induces rational expectations, as the partition isolates each contingency in a separate cell. In contrast, when the partition size is 1, both contingencies are in the same cell, hence $\hat{\sigma}(\theta, 1) = \hat{\sigma}(\theta, 0)$. But this means that $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0) < \theta$, such that the only best-reply to $\hat{\sigma}$ is $a = 0$.

It follows that the only way to sustain trust in MLEQ is with the fine partition of size 2. In this case, $\hat{\sigma}(\theta, h) = \sigma(\theta, h)$ for every h, such that in MLEQ, equilibrium, $\sigma(\theta, 1) - \sigma(\theta, 0) = \theta$, as in Nash equilibrium. The only additional requirement is that the fine partition is ML-optimal. This requirement holds if and only if the inequality (1) is satisfied. Since $p_\sigma(\theta, 1) + p_\sigma(\theta, 0) = 1$, this inequality is reduced to

$$p_\sigma(\theta, 1)p_\sigma(\theta, 0) \cdot (\sigma(\theta, 1) - \sigma(\theta, 0))^2 \geq c$$

Plugging the formula for $p_\sigma(\theta, h)$ and the equilibrium condition $\sigma(\theta, 1) - \sigma(\theta, 0) = \theta$, we obtain the following condition for σ to be consistent with equilibrium:

$$\frac{(\sigma(\theta,1)-\theta)(1-\sigma(\theta,1))}{(1-\theta)^2}\theta^2 \geq c \tag{2}$$

It can be easily checked that if $c > \theta^2/4$, there exists no solution to (2). Thus, a necessary and sufficient condition for the existence of MLEQ that exhibit trust is $c \leq \theta^2/4$. When the condition holds, the maximally cooperative equilibrium strategy is given by the value of $\sigma(\theta, 1) \geq (1 + \theta)/2$ that solves (2) bindingly (and then $\sigma(\theta, 0) = \sigma(\theta, 1) - \theta$). Importantly, this value is bounded away from 1. In particular, when $c = \theta^2/4$, $\sigma(\theta, 1) = (1+\theta)/2$ such that $p_\sigma(1) = \frac{1}{2}$. Thus, the penalty on complex beliefs implies a limit on the amount of trust that can be sustained in MLEQ

3.2 Example II: $n = 2$

Suppose now that $\Theta = \{\theta_1, \theta_2\}$. The following result characterizes the maximal amount of cooperation that can be sustained in MLEQ under certain restrictions on the parameter values.

Proposition 1 *Let $\theta_2, \theta_1 > \frac{1}{2}$ and $c \in (\frac{1}{8}, \frac{1}{4})$. Then, any MLEQ that exhibits a positive cooperation rate in some state θ must satisfy $|\Pi| = 2$; $p_\sigma(\theta, 1) \leq \theta^2/(1 + \theta^2)$; and $p_\sigma(\theta', 1) = 0$ in $\theta' \neq \theta$. If $\theta^2 < \sqrt{c}/(1 - \sqrt{c})$, cooperation in state θ is unsustainable. If $\theta^2 \geq \sqrt{c}/(1 - \sqrt{c})$, the upper bound on $p_\sigma(\theta, 1))$ is sustainable.*

Proof. As before, a SMLEQ (σ, Π) can sustain trust with positive probability only if $|\Pi| > 1$. Let us distinguish between two cases.

Case 1: $|\Pi| \geq 3$.
W.l.o.g, assume $\hat{\sigma}(\theta_1, 1) - \hat{\sigma}(\theta_1, 0) = \theta_1$. Since $\theta_1, \theta_2 > \frac{1}{2}$, $\hat{\sigma}(\theta_1, 0) < \frac{1}{2}$. If players exhibit trust with positive probability in θ_2, then $\hat{\sigma}(\theta_2, 0) < \frac{1}{2}$ as well; otherwise, $\hat{\sigma}(\theta_2, 0) = 0$. In either case, $|\hat{\sigma}(\theta_1, 0) - \hat{\sigma}(\theta_2, 0)| < \frac{1}{2}$.

Let us consider two sub-cases. First, suppose $\hat{\sigma}(\theta_1, 0) \neq \hat{\sigma}(\theta_2, 0)$, such that $(\theta_1, 0)$ and $(\theta_2, 0)$ belong to different partition cells. Note that $\alpha\beta/(\alpha +$

$\beta) < \frac{1}{2}$ for every $\alpha, \beta \in [0,1]$. It follows that

$$\frac{p_\sigma(\pi(\theta_1,0))p_\sigma(\pi(\theta_2,0))}{p_\sigma(\pi(\theta_1,0)) + p_\sigma(\pi(\theta_2,0))}(\hat{\sigma}(\theta_1,0) - \hat{\sigma}(\theta_2,0))^2 < \frac{1}{8} < c$$

in violation of (1). Second, suppose $\hat{\sigma}(\theta_1,0) = \hat{\sigma}(\theta_2,0)$. Then,

$$\begin{aligned} |\hat{\sigma}(\theta_2,1) - \hat{\sigma}(\theta_1,1)| &= |\hat{\sigma}(\theta_2,0) + \theta_2 - \hat{\sigma}(\theta_1,0) - \theta_1| \\ &= |\theta_2 - \theta_1| \\ &< \frac{1}{2} \end{aligned}$$

such that

$$\frac{p_\sigma(\pi(\theta_1,1))p_\sigma(\pi(\theta_2,1))}{p_\sigma(\pi(\theta_1,1)) + p_\sigma(\pi(\theta_2,1))}(\hat{\sigma}(\theta_1,1) - \hat{\sigma}(\theta_2,1))^2 < \frac{1}{8} < c$$

in violation of (1). □

Case 2: $|\Pi| = 2$.
We will first rule out partitions that consist of two equally sized cells:

(i) Suppose Π consists of the cells $\{(\theta_1,0),(\theta_2,0)\}$ and $\{(\theta_1,1),(\theta_2,1)\}$. Then, $\hat{\sigma}(\theta_1,h) = \hat{\sigma}(\theta_2,h)$ for both $h = 0,1$. However, this is inconsistent with equilibrium. To see why, note first that if $\hat{\sigma}(\theta,1) - \hat{\sigma}(\theta,0) > \theta$ for some θ, then we must have in equilibrium that $\sigma(\theta,1) = \sigma(\theta,0) = 1$. But then, by Remark 2, both $(\theta,1)$ and $(\theta,0)$ should be assigned to the cell with the highest $\hat{\sigma}$, a contradiction. In the same manner, we can rule out the possibility that $\hat{\sigma}(\theta,1) - \hat{\sigma}(\theta,0) < \theta$ for some θ. It follows that $\hat{\sigma}(\theta,1) - \hat{\sigma}(\theta,0) = \theta$ for both θ, a contradiction. Thus, we can rule out this particular partition.

(ii) Now suppose Π consists of the cells $\{(\theta_1,1),(\theta_1,0)\}$ and $\{(\theta_2,1),(\theta_2,0)\}$. Then, $\hat{\sigma}(\theta,1) = \hat{\sigma}(\theta,0)$ for both θ. The unique best-reply against $\hat{\sigma}$ is to play $a = 0$, contradicting the assumption that the equilibrium sustains trust.

(iii) Finally, the partition that consists of $\{(\theta_1,1),(\theta_2,0)\}$ and $\{(\theta_2,1),(\theta_1,0)\}$ implies $\hat{\sigma}(\theta,0) > \hat{\sigma}(\theta,1)$ for some θ, thus making $a = 0$ the unique best-reply to $\hat{\sigma}$ at θ. But then $\sigma(\theta,1) = \sigma(\theta,0) = 0$, contradicting the assignment of $(\theta,1)$ and $(\theta,0)$ to different cells.

It follows that if Π is consistent with MLEQ, one of its two cells must be a singleton $\{(\theta, h)\}$. This means that for $\theta' \neq \theta$, $(\theta', 1)$ and $(\theta', 0)$ are assigned to the same cell, which (as we have seen) means that in equilibrium, $\sigma(\theta', 1) = \sigma(\theta', 0) = 0$. Suppose that this cell also contains $(\theta, 1)$, such that the other, singleton cell is $\{(\theta, 0)\}$. It follows that $\hat{\sigma}(\theta, 0) = \sigma(\theta, 0)$. Equilibrium requires that $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0) = \theta$, hence $\hat{\sigma}(\theta, 1) > \hat{\sigma}(\theta, 0)$. By Remark 2, ML-optimality requires that since $\sigma(\theta', 1) = \sigma(\theta', 0) = 0 < \hat{\sigma}(\theta, 0)$, $(\theta', 0)$ and $(\theta', 1)$ should be bundled together with $(\theta, 0)$ rather than $(\theta, 1)$, a contradiction.

The only remaining option is that Π consists of two cells that take the form $\{(\theta', 1), (\theta', 0), (\theta, 0)\}$ and $\{(\theta, 1)\}$, such that

$$\begin{aligned} \hat{\sigma}(\theta, 1) &= \sigma(\theta, 1) \\ \hat{\sigma}(\theta, 0) &= \frac{p_\sigma(\theta, 0)}{\frac{1}{2} + p_\sigma(\theta, 0)} \sigma(\theta, 0) \end{aligned}$$

Equilibrium requires $\sigma(\theta, 1) - \hat{\sigma}(\theta, 0) = \theta$. ML-optimality requires that if we move $(\theta, 0)$ out of its cell and into the same cell as $(\theta, 1)$, the mean squared error will not decrease:

$$\frac{\frac{1}{2} \cdot p_\sigma(\theta, 0)}{\frac{1}{2} + p_\sigma(\theta, 0)} (\sigma(\theta, 0) - 0)^2 \leq \frac{p_\sigma(\theta, 0) \cdot p_\sigma(\theta, 1)}{p_\sigma(\theta, 0) + p_\sigma(\theta, 1)} (\sigma(\theta, 1) - \sigma(\theta, 0))^2 \tag{3}$$

Plugging the expressions for $p_\sigma(\theta, h)$ and the equilibrium requirement $\sigma(\theta, 1) - \hat{\sigma}(\theta, 0) = \theta$ into (3), and with a bit of algebra, we can derive the following tight upper bound:

$$p_\sigma(\theta, 1) \leq \frac{\theta^2}{1 + \theta^2}$$

This inequality is binding when (3) is binding.

The only remaining constraint is that the putative equilibrium is robust to deviating from the two-cell partition to the degenerate partition:

$$\frac{p_\sigma(\pi(\theta, 1)) p_\sigma(\pi(\theta, 0))}{p_\sigma(\pi(\theta, 1)) + p_\sigma(\pi(\theta, 1))} (\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0))^2 = p_\sigma(\theta, 1)(1 - p_\sigma(\theta, 1))\theta^2 \geq c$$

Since $p_\sigma(\theta, 1) < \frac{1}{2}$, the L.H.S of this inequality holds only if it holds under the upper bound on $p_\sigma(\theta, 1)$, which happens if and only if

$$c \leq \frac{\theta^4}{(1+\theta^2)^2} \tag{4}$$

This inequality is equivalent to $\theta^2 \geq \sqrt{c}/(1-\sqrt{c})$. ■

The derivation does not take a stand on whether θ is θ_1 or θ_2. Since $\theta_2 > \theta_1$, we can see that the largest possible equilibrium probability of $h = 1$ is $\theta_2^2/(1+\theta_2^2)$. Thus, a lower cost of cooperation leads to a lower upper bound on the long-run probability of cooperative behavior, and also makes it harder to sustain cooperation at all.

In the $n = 1$ example, the only possible deviation in the Π dimension was merging the contingencies $h = 1$ and $h = 0$ into one cell, thus destroying the incentive to cooperate. When $n = 2$, there is an additional possible deviation, which merges the contingencies (θ_1, h) and (θ_2, h') into one cell, thus potentially destroying a distinction between the two payoff states that is necessary to maintain the incentive to cooperate. Proposition 1 reflects this additional force. As a result, the scope for cooperation is lower than in the $n = 1$ case, where it is possible to sustain $p_\sigma(1) \geq \frac{1}{4}$ whenever $c \leq \theta^2/4$. Condition (4), which is necessary for sustaining cooperation when $n = 2$, is more stringent.

Under the restrictions on (θ_1, θ_2, c), the partition size in cooperation-sustaining MLEQ is 2. However, *by itself*, this restriction on $|\Pi|$ implies a *weaker* constraint on cooperation rates. For example, suppose $\theta_2 \gtrapprox \theta_1 > \frac{1}{2}$, and consider the partition $\Pi = \{\{(\theta_1, 0), (\theta_2, 0)\}, \{(\theta_1, 1), (\theta_2, 1)\}\}$. Construct the strategy $\sigma(\theta_2, 1) = \sigma(\theta_2, 0) = \sigma(\theta_1, 0) = 0$, $\sigma(\theta_1, 1) = \theta_1$. Then, $p_\sigma(\theta_1, 1) = p_\sigma(\theta_2, 0) = \frac{1}{2}$, such $\hat{\sigma}(\theta_1, 1) = \hat{\sigma}(\theta_2, 1) = \theta_1$ while $\hat{\sigma}(\theta_1, 0) = \hat{\sigma}(\theta_2, 0) = 0$. Players' behavior is a best-reply to this belief. The strategy induces an overall rate of cooperation of $\theta_1/2$, which is above the MLEQ upper bound of $\theta_2^2/(1+\theta_2^2)$ because $\theta_2 \gtrapprox \theta_1$.

4 Results

The first result establishes a necessary condition for SMLEQ-sustaining positive cooperation rates in m payoff states. For a fixed complexity cost, there is an upper bound on the number of payoff states in which cooperation can be sustained in SMLEQ. The bound is not tight, because it is calculated without taking into account the endogeneity of p_σ.

Proposition 2 *Suppose* $2cm^3 > 1$. *Then, for generic* Θ*, there exists no SMLEQ that induces positive cooperation rates in* m *payoff states.*

Proof. Consider an SMLEQ (σ, Π) in which $p_\sigma(\theta, 1) > 0$ for m payoff states θ. The proof proceeds stepwise.

Step 1: *If* $p_\sigma(\theta, 1) > 0$*, then* $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0) = \theta$.

Proof: Consider a payoff state θ for which $p_\sigma(\theta, 1) > 0$. Suppose $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0) < \theta$. Then, the unique best-reply at θ is $a = 0$. Hence, $\sigma(\theta, 1) = \sigma(\theta, 0) = 0$ in equilibrium, contradicting the assumption that $p_\sigma(\theta, 1) > 0$. Now suppose $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0) > \theta$. Then, the unique best-reply at θ is $a = 1$. Hence, $\sigma(\theta, 1) = \sigma(\theta, 0) = 1$ in equilibrium. By the optimal assignment property of SMLEQ, $(\theta, 1)$ and $(\theta, 0)$ must both be assigned to $\arg\max_{\pi \in \Pi} \hat{\sigma}(\pi)$, such that $\hat{\sigma}(\theta, 1) = \hat{\sigma}(\theta, 0)$, a contradiction. □

Step 2: *For generic* Θ*, if* $p_\sigma(\theta, 1) > 0$ *for* m *payoff states* θ*, then* $|\Pi| > m$.

Proof: Consider a payoff state θ for which $p_\sigma(\theta, 1) > 0$. By Step 1, $\hat{\sigma}(\theta, 1) \neq \hat{\sigma}(\theta, 0)$, which by Corollary 1 means that $\pi(\theta, 1) \neq \pi(\theta, 0)$. Construct a *non-directed graph* whose nodes correspond to the cells in Π, such that π and π' are linked if there is θ for which $p_\sigma(\theta, 1) > 0$ such that $\pi(\theta, h) = \pi$ and $\pi(\theta, 1 - h) = \pi'$ for some h. Note that by definition, the graph has m edges.

Suppose $\pi(\theta, 1)$ and $\pi(\theta, 0)$ are linked. Then, by Step 1, $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0) = \theta$. Suppose the graph contains an additional, *indirect* path between $\pi(\theta, 1)$ and $\pi(\theta, 0)$. By Step 1, this means that there is a sequence of payoff states $\theta^1, ..., \theta^K$, such that $\theta = \sum_{k=1}^{K} \theta^k = \theta$. For generic Θ, this requirement fails to hold. It follows that if two graph nodes are linked, there is no additional indirect path between them (in other words, the link is a bridge). It follows that the graph must be a forest (i.e., every connected graph component is a tree), hence it has at least $m + 1$ nodes. It follows that $|\Pi| > m$. □

Step 3: *Formulating an auxiliary max-min problem*

Proof: We have established that Π consists of $K \geq m+1$ cells, each with its own distinct $\hat{\sigma}$. Enumerate the partition cells as $\pi_1, ..., \pi_K$. Denote $p_i = p_\sigma(\pi_i)$ and $\hat{\sigma}_i = \hat{\sigma}(\pi_i)$. By (1), (σ, Π) is an MLEQ only if

$$c \leq \min_{i \neq j} \frac{p_i p_j}{p_i + p_j} (\hat{\sigma}_i - \hat{\sigma}_j)^2$$

for every distinct $i, j \in \{1, ..., K\}$. By definition, the R.H.S. of this inequality is bounded from above by

$$\max_{p \in \Delta\{1,...,K\},\ \hat{\sigma} \in [0,1]^K} \ \min_{i \neq j} \frac{p_i p_j}{p_i + p_j} (\hat{\sigma}_i - \hat{\sigma}_j)^2 \tag{5}$$

Without loss of generality, let $0 \leq \hat{\sigma}_1 < \hat{\sigma}_2 < \cdots < \hat{\sigma}_K \leq 1$. For every $k = 1, ..., K-1$, denote $q_k = \hat{\sigma}_{k+1} - \hat{\sigma}_k$. Denote $p = (p_k)_{k=1,...,K}$ and $q = (q_k)_{k=1,...,K-1}$. By definition, (5) is weakly below

$$\max_{p,q} \ \min_{k=1,...,K-1} \frac{p_k p_{k+1}}{p_k + p_{k+1}} q_k^2 \tag{6}$$

By definition, $p \in \Delta\{1, ..., K\}$ is a probability n-vector, whereas $q_k > 0$ for every $k = 1, ..., K-1$ and $\sum_k q_k \leq 1$. Since (6) is increasing in q, we can take (as far as the solution of this max-min problem is concerned) the latter constraint to be binding, such that $q \in \Delta\{1, ..., K-1\}$. □

Step 4: *The value of (6) is strictly below* $1/2(K-1)^3$.

Proof: Let us break the max-min problem into two steps:

$$\max_p \left(\max_q \ \min_{k=1,...,K-1} \frac{p_k p_{k+1}}{p_k + p_{k+1}} q_k^2 \right)$$

As a first step, fix p. For every $k = 1, ..., K-1$, denote

$$A_k = \sqrt{\frac{p_k \, p_{k+1}}{p_k + p_{k+1}}}$$

Since k is selected to minimize $(A_k q_k)^2$, it is clear that the solution to the parenthetical max-min problem $\max_q \min_k (A_k q_k)^2$ equalizes $A_k q_k$ across all

k, such that

$$q_k = \frac{\frac{1}{A_k}}{\sum_{j=1}^{K-1} \frac{1}{A_j}}$$

and the max-min value is

$$\frac{1}{\left(\sum_{j=1}^{K-1} \frac{1}{A_j}\right)^2} \tag{7}$$

In the procedure's second step, choose p to maximize this value. This is equivalent to choosing p to minimize

$$\sum_{k=1}^{K-1} \sqrt{\frac{1}{p_k} + \frac{1}{p_{k+1}}} \tag{8}$$

This expression is strictly convex. It is also symmetric between p_1 and p_K, as well as across all interior components $p_2, ..., p_{K-1}$. Therefore, the expression's unique minimizer has full support and satisfies $p_1 = p_K$. Construct an alternative probability vector $p^* = (p_1, ..., p_{K-1})$ defined as follows:

$$p_k^* = \frac{p_k}{\sum_{j=1}^{K-1} p_j}$$

Then, we can rewrite (8) as

$$\begin{aligned} &\frac{1}{\sqrt{\sum_{j=1}^{K-1} p_j}} \left(\sum_{k=1}^{K-2} \sqrt{\frac{1}{p_k^*} + \frac{1}{p_{k+1}^*}} + \sqrt{\frac{1}{p_{K-1}^*} + \frac{1}{p_1^*}} \right) \\ > &\sum_{k=1}^{K-2} \sqrt{\frac{1}{p_k^*} + \frac{1}{p_{k+1}^*}} + \sqrt{\frac{1}{p_{K-1}^*} + \frac{1}{p_1^*}} \end{aligned}$$

The latter expression is strictly convex and symmetric. Therefore, its unique minimizer is the uniform distribution, yielding a minimal value of $(K - 1)\sqrt{2(K-1)}$. It follows that (7) is strictly below $1/2(K-1)^3$. □

Step 5: *Completing the proof*

It follows from Steps 3 and 4 that (5), and therefore also c, are below $1/2(K-1)^3$. Since $K \geq m+1$, this contradicts the assumption that $2cm^3 > 1$. We can conclude that there is no SMLEQ with positive cooperation rates in m payoff states. ■

The intuition behind the result is as follows. First, in order to sustain cooperation in m payoff states, players' beliefs must be sufficiently complex to create the distinctions that sustain cooperation incentives. Specifically, I show that the equilibrium partition must have at least $m+1$ cells.

Second, as a partition gets larger, the probability of individual cells goes down on average, and so does the average distance between cells' average cooperation probability. These two quantities are inversely proportional to $m+1$ and $(m+1)^2$, respectively. As m grows larger, a deviation in belief space that merges two cells of a putative equilibrium partition becomes more likely.

In this sense, an environment that is more complex in the sense of having a larger n makes it harder to sustain equilibrium beliefs that make the necessary distinctions between contingencies for maintaining the incentive to cooperate.

Corollary 3 *Fix c. As $n \to \infty$, the overall cooperation rate in SMLEQ converges to zero.*

This result immediately follow from Proposition 2. Since the maximal number of payoff states with positive cooperation rates is bounded from above by $\sqrt[3]{c}$, the fraction of these states becomes negligible as n grows larger.

The following result exploits the endogeneity of p_σ to derive a different kind of limit on the ability to sustain cooperation in MLEQ.

Proposition 3 *Fix c, n, and let σ be an SMLEQ strategy. Then, the probability of payoff states θ for which $p_\sigma(\theta, 1) = \frac{1}{n}$ is at most $\frac{1}{2}$.*

Proof. Suppose that $p_\sigma(\theta, 1) = \frac{1}{n}$ for some payoff state θ. Then, we must have $\sigma(\theta, 1) = 1$ and $\hat{\sigma}(\theta, 1) - \hat{\sigma}(\theta, 0) = \theta$ (the latter was established in Step 1 of the proof of Proposition 2). Since $p_\sigma(\theta, 0) = 0$, SMLEQ implies that $(\theta, 0)$ is assigned to a partition cell π that minimizes $|\hat{\sigma}(\pi) - \sigma(\theta, 0)|$. That is, there must be some other contingency (θ', h') such that $\pi(\theta', h') = \pi(\theta, 0)$ and $\hat{\sigma}(\pi(\theta, 1)) - \hat{\sigma}(\pi(\theta, 0)) = \theta$.

Now suppose $p_\sigma(\theta'', 1) = \frac{1}{n}$ for some other payoff state $\theta'' \neq \theta$. Since $\sigma(\theta'', 1) = 1 = \sigma(\theta, 1)$, SLMEQ implies that $(\theta'', 1)$ and $(\theta, 1)$ are assigned to the same partition cell (the one with the highest $\hat{\sigma}$), such that $\hat{\sigma}(\theta'', 1) = \hat{\sigma}(\theta, 1)$. Since we must also have $\hat{\sigma}(\theta'', 1) - \hat{\sigma}(\theta'', 0) = \theta'' \neq \theta$, it follows that $(\theta'', 0)$ and $(\theta, 0)$ are assigned to different partition cells. By the same argument as in the previous paragraph, there must be a contingency (θ''', h''') such that $\pi(\theta''', h''') = \pi(\theta'', 0)$. But since $(\theta'', 0)$ and $(\theta, 0)$ are assigned to different cells, $(\theta''', h''') \neq (\theta', h')$.

It follows that for every payoff state that exhibits full cooperation, there must be a distinct payoff state that exhibits less-than-full cooperation. Therefore, the fraction of payoff states that exhibit full cooperation is at most $\frac{1}{2}$. ■

The logic behind this result is simple. If $\sigma(\theta, 1) = 1$ for some θ, then $p_\sigma(\theta, 0) = 0$. Strong MLEQ then requires the equilibrium partition to bundle $(\theta, 0)$ with other non-zero-probability contingencies, which in turn means that there is some other payoff state for which the cooperation rate is below 1. An additional argument establishes that this pairing must be different for different payoff states, which implies the upper bound of $\frac{1}{2}$ on the fraction of states with full cooperation.

Comment: The objective meaning of c

The model in this paper treats c as a primitive, as if players have an intrinsic taste for simple beliefs. However, under the ML interpretation, we should view ML-optimality — and the role that c plays in it — as a reduced-form formalization of an underlying bias-variance trade-off. This trade-off would arise in a more elaborate model in which players do not observe σ directly, but instead learn about it from a noisy sample.

Specifically, suppose that $n = 1$ and that for every $h = 0, 1$, players observe $x(h) = \sigma(h) + \varepsilon(h)$, where $\varepsilon(h)$ is an independent noise term with mean zero and variance $v/p_\sigma(h)$; $v > 0$ is a constant. The basic model would correspond to the case in which $v = 0$. The assumption that the variance is inversely proportional to $p_\sigma(h)$ is in the spirit of Danenberg and Spiegler (2024): Players obtain a representative finite sample drawn from the ergodic distribution over contingencies, such that the number of observations about a contingency is proportional to its frequency.

Fix a partition Π of $H = \{0, 1\}$. For every cell $\pi \in \Pi$, define $\hat{\sigma}(h)$ as the expected value of x in $\pi(h)$. As before, the MSPE it induces is $E(\hat{\sigma}(h) - \sigma(h))^2$. Then, the MSPE induced by the fine partition is

$$p_\sigma(0) \cdot \frac{v}{p_\sigma(0)} + p_\sigma(1) \cdot \frac{v}{p_\sigma(1)} = 2v$$

whereas the MSPE induced by the degenerate, coarse partition is

$$\begin{aligned} & p_\sigma(0)p_\sigma(1)[\sigma(1) - \sigma(0)]^2 + (p_\sigma(0))^2 \cdot \frac{v}{p_\sigma(0)} + (p_\sigma(1))^2 \cdot \frac{v}{p_\sigma(1)} \\ = \; & p_\sigma(0)p_\sigma(1)[\sigma(1) - \sigma(0)]^2 + v \end{aligned}$$

It follows that even if there is no intrinsic preference for simple beliefs and ML-optimality is entirely based on minimizing MSPE, the fine partition is ML-optimal if

$$p_\sigma(0)p_\sigma(1)[\sigma(1) - \sigma(0)]^2 \geq v$$

This is the same criterion as in the basic model, except that the role of c in the basic model is now played by the noise variance constant v.

I should emphasize that this exercise cannot be turned into a full-fledged "foundation" for MLEQ. Apart from considerations of tractability and generalizability, the fundamental difficulty is that if players form beliefs according to a finite sample with smooth noise, then they will almost always have a strict best-replying action to their sample-based belief. This action is by definition independent of the observed history. Yet, history-dependent behavior is crucial for sustaining cooperation in the dynamic trust game. This is a limitation of this game as a metaphor for the long-run interactions with ML-based beliefs it is meant to capture.

Nevertheless, this little exercise gives a sense of how we may want to interpret the inequality in Proposition 2. It essentially a says that as the number m of payoff states that exhibit cooperation grows, the variance of the noise with which players observe σ should decrease at a rate of approximately $1/m^3$.

5 Conclusion

The ML dilemma between what I described here as model-based vs. model-free methods has occupied AI specialists (e.g., see Sutton and Barto (1998) and Levine et al. (2020)), and involves technical considerations that are far outside my expertise. Nevertheless, to the extent that ML methods are applied to dynamic strategic decision-making — as is the case in such domains as oligopoly pricing or autonomous driving — I hope that this paper has made a valuable contribution to the discussion.

Model-based ML involves reducing the dimensionality of the environment's representation. In game-theoretic contexts, this may take the form of collapsing distinct contingencies into the same equivalence class. The basic qualitative insight of this paper is that when this classification is guided by trading off predictive success against model complexity, it may fail to make the distinctions that sustain individual incentives to act cooperatively in a long-run interaction. Successful long-run cooperation is sustained by counterfactual (or at least rare) threats, yet their very rarity causes ML methods that involve the fit-simplicity trade off to group them together with other contingencies. This same trade-off causes similar beliefs to be grouped together, even when they imply radically different best-replying actions.

This paper used the dynamic trust game to offer various illustrations of these two themes, and demonstrated that they drastically narrow the scope for equilibrium long-run cooperation. Whether this conclusion has broader lessons for the performance of model-based ML in dynamic strategic interactions will hopefully be explored in future research.

References

[1] Banchio, M. and G. Mantegazza (2022). Artificial intelligence and spontaneous collusion. arXiv preprint arXiv:2202.05946.

[2] Bendor, J., D. Mookherjee, and D. Ray (2001). Reinforcement learning in repeated interaction games. The BE Journal of Theoretical Economics 1, 20011004.

[3] Brown, Z. and A. MacKay (2023). Competition in pricing algorithms. American Economic Journal: Microeconomics 15, 109-156.

[4] Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello (2020). Artificial intelligence, algorithmic pricing, and collusion. American Economic Review, 110, 3267-3297.

[5] Danenberg, T. and R. Spiegler (2024). A Representative-Sampling Model of Stochastic Choice. Journal of Political Economy: Microeconoics, forthcoming.

[6] Eliaz, K. (2003). Nash equilibrium when players account for the complexity of their forecasts. Games and Economic Behavior 44, 286-310.

[7] Hansen, K., K. Misra, and M. Pai (2021). Algorithmic collusion: Supra-competitive prices via independent algorithms. Marketing Science, 40, 1-12.

[8] Hastie, T., R. Tibshirani, and J. Friedman (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer.

[9] Jehiel, P. (2005). Analogy-based expectation equilibrium, Journal of Economic theory 123, 81-104.

[10] Jehiel, P. and E. Mohlin (2024). Categorization in games: A bias-variance perspective. Working paper.

[11] Jehiel, P. and G. Weber (2024). Endogenous clustering and analogy-based expectation equilibrium. Working paper.

[12] Kaufman, L. and P. Rousseeuw (1990). Finding groups in data: An introduction to cluster analysis. John Wiley & Sons.

[13] Levine, S., A. Kumar, A., G. Tucker, and J. Fu (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. arXiv preprint arXiv:2005.01643.

[14] Mohlin, E. (2014). Optimal categorization. Journal of Economic Theory 152, 356-381.

[15] Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. Journal of economic theory 39, 83-96.

[16] Spiegler, R. (2002). Equilibrium in justifiable strategies: A model of reason-based choice in extensive-form games. Review of Economic Studies 69, 691-706.

[17] Spiegler, R. (2004). Simplicity of beliefs and delay tactics in a concession game. Games and Economic Behavior 47, 200-220.

[18] Spiegler, R. (2005). Testing threats in repeated games. Journal of Economic Theory 121, 214-235.

[19] Sutton, R and A. Barto (1998). Reinforcement learning: An introduction. MIT press.

[20] Waizmann, S. (2024). AI in Action: Algorithmic Learning with Strategic Consumers. Working paper.