Signaling with Private Monitoring*

Gonzalo Cisternas and Aaron Kolb

April 7, 2023

Abstract

A sender signals her private information to a receiver who privately monitors the sender's behavior, while the receiver transmits his private inferences back through an imperfect public signal of his actions. In a linear-quadratic-Gaussian setup in continuous time, we construct linear Markov equilibria where the state variables are the players' beliefs up to the sender's second-order belief. This state is an explicit function of the sender's past play—hence, her private information—which leads to *separation* through the second-order belief channel. We examine the implications of this effect in models of organizations, reputation, and trading. We also provide a fixed-point technique for finding solutions to systems of ordinary differential equations with a mix of initial and terminal conditions, and that can be applied to other dynamic settings.

1 Introduction

This paper introduces a new class of signaling games featuring private signals of behavior. These games can be seen as continuous-time versions of repeated noisy signaling games where a receiver sees imperfect signals of a sender's actions privately, and hence the receiver develops a private belief about the sender's type. In settings where this form of private information is relevant for the sender's behavior, private monitoring forces the players to construct non-trivial *beliefs about each other's beliefs* to determine their best courses of action. We offer a framework where this complex forecasting issue is manageable and develop new methods to study how higher-order uncertainty can affect outcomes in signaling games.

*Cisternas: Federal Reserve Bank of New York, gonzalo.cisternas@ny.frb.org. Kolb: Indiana University Kelley School of Business, 1309 E. Tenth St., Bloomington, IN 47405 kolba@indiana.edu. We thank Alessandro Bonatti, Isa Chavez, Yi Chen, Wouter Dessein, Mehmet Ekmekci, Eduardo Faingold, Robert Gibbons, Nathan Kaplan, Marina Halac, Stephen Morris, Alessandro Pavan, Asani Sarkar, Andy Skrzypacz, Bruno Strulovici, Juuso Toikka, Vish Viswanathan, and numerous seminar audiences for useful conversations and comments. The views expressed in this paper are those of the authors and do not necessarily represent the position of the Federal Reserve Bank of New York or the Federal Reserve System.

Private signals of behavior are a central feature of many economic settings. In organizations, it has long been recognized that individuals subjectively evaluate what others have done (MacLeod, 2003). In the online world, data brokers secretly collect imperfect signals of consumer behavior to quantify unobserved consumer characteristics (Bonatti and Cisternas, 2020). In financial markets, some traders have an advantage in picking up signals of others' trades (Yang and Zhu, 2020). More generally, how do leaders of organizations transmit their knowledge when they do not know how their actions are interpreted? How do agents manage a reputation when they are uncertain about how they are being perceived? How do traders respond to their own actions creating private information available to others?[1]

These questions can be difficult to answer for three main reasons. First, when beliefs beyond the first order are needed, and these evolve over time, the state space may grow without bound as a result of players trying to "forecast the forecasts of others" (Townsend, 1983). Second, these games have the potential to be asymmetric: for instance, a sender of a *fixed* type can interact with a receiver possessing *evolving* private information in the form of a belief. Third, many of these settings can be non-stationary due to endogenous learning effects, stemming from the players signaling their information over time.

In our approach, a forward-looking sender (she) and a myopic receiver (he), both with quadratic preferences, interact over a finite horizon. The sender has a fixed, normally distributed type. Our innovation is to allow the receiver to privately observe a noisy signal of the sender's action; meanwhile, the receiver transmits information back via a public signal of his behavior. Time is continuous and the shocks in the signals are additive and Brownian. In this linear-quadratic-Gaussian (LQG) set up, we construct linear Markov equilibria (LMEs) with the players' beliefs up to the sender's *second-order belief* as the states.

Linear Markov strategies and representation of the second-order belief In dynamic games, players must estimate their rivals' continuation strategies to determine their best action. To see why this is complex if monitoring is private, consider our sender. First, fixing her strategy, she must assign probabilities to the private signals that the receiver could have seen; since these histories grow as play unfolds, this estimation problem gets worse over time. Second, the sender's resulting estimates will vary with her own private histories of behavior: not knowing what the receiver has seen, higher (lower) past actions become indicative of higher (lower) signals observed by the receiver; but this means that the receiver in turn must assign probabilities to the sender's private histories, and so forth.

With incomplete information, however, the players can use strategies that depend on their beliefs emanating from the sender's private type; further, with quadratic preferences,

[1]We discuss the differences from these papers in the literature review section.

it is the means of such posterior beliefs, along with the sender's type itself, that can be used linearly in the strategies. A linear use of Gaussian information then implies that the receiver's posterior mean is a linear aggregate of his private signal's realizations: the sender's aforementioned problem of assessing the receiver's private histories is reduced to estimating a one-dimensional statistic. In turn, the sender's resulting second-order belief—her belief about the receiver's belief—is also a linear aggregate: of the sender's past actions—reflecting that her behavior is used in the forecasting exercise—and of the public signal's history, which carries the receiver's private belief through the latter player's action.

Since the sender's second-order belief explicitly depends on her past actions, this state is also her private information when she signals her type, and hence must be forecasted by the receiver. The first key contribution of this paper corresponds to a novel *representation* of the second-order belief along the path of play of a linear Markov strategy profile (Lemma 1 in Section 3). Specifically, this state variable is a convex combination of the sender's type and the belief about the latter based exclusively on the public signal whenever (i) the sender uses her type, her second-order belief, and the aforementioned public state linearly; while (ii) the receiver uses his private belief and the same public state also in a linear fashion. These belief states constitute our candidate Markov states to be deployed linearly.

Importantly, since deviations are hidden due to the full-support monitoring, the receiver always believes the sender is following her (equilibrium) linear Markov strategy; hence the receiver always believes that the representation holds. The linear aggregation of private histories—in particular, of the sender's actions in her second-order belief—thus provides great tractability: via the representation, forecasting the second-order belief reduces to simply forecasting the type, and hence the players do not need other higher-order beliefs to assess what the other will do. Since the receiver relies on the public belief state in this "third-order" belief forecasting exercise, this state must be included in the players' strategies.

The representation is therefore central to our analysis. Importantly, this result is not merely a "proof of concept" that the state space does not explode: it is central to the new insights that private monitoring brings to signaling games, and it is also key for setting up the sender's best-response problem to determine the coefficients that the players attach to the belief states in their strategies. We discuss these properties in the next two topics.

Applications: the history-inference effect The representation encapsulates a natural idea: different sender types, by having acted differently, necessarily develop different beliefs about the receiver, even when seeing the same signals about the receiver's inferences. This differs from the traditional case in which the signals seen by the receiver are public: there, at any history of these signals, all types would agree on the receiver's (public) belief despite having acted differently in the past. Economically, therefore, the representation reflects that private monitoring opens a new channel for separation: one that operates via the sender's second-order belief. We refer to the signaling implications of this new channel as the *history-inference effect*, which we study in settings where higher-order uncertainty likely matters.

Section 4.1 explores a coordination game in organizations. A leader (sender) and a follower (receiver) form a team. The team's performance depends on the proximity of the leader's action to both a state of the world (the type) and the follower's action. The follower simply tries to coordinate. The issue is that as soon as the leader tries to signal the new state, she loses track of the follower's understanding, and the players need to think about what the other knows to be able to coordinate. Importantly, because leaders facing higher states of the world take higher actions in equilibrium, they believe that their followers also develop higher beliefs—via the coordination motive, therefore, the history-inference effect amplifies separation. Through this channel, a dichotomy between learning and performance in organizations can arise: the leader can transmit more information to the follower compared to when the follower's belief is known, thus improving the receiver's learning; but the team's performance is lower. Indeed, as we show, the follower's more precise acquired knowledge of the environment is a measure of the coordination costs incurred along the way.

Section 4.2 examines a reputation game. The sender is an individual with a privately known *bias*—e.g., a politician with a stance on a relevant issue—with the prior mean capturing the unbiased type. The sender finds it costly to take actions away from her type, but wants to appear as unbiased in the eyes of a relevant receiver at the end of the game (e.g., for a reappointment): the sender suffers a quadratic loss in the distance between the receiver's terminal belief and the prior mean. As higher types take higher actions due to their larger biases, they expect their receivers to develop more extreme beliefs, which is reputationally costly. Via the history-inference effect, therefore, higher types correct their actions more aggressively than low-type counterparts, thereby *reducing* separation relative to the case when the sender knows her reputation. A subtle tradeoff emerges: if the sender is uncertain about her reputation, she may be unable to take the best actions to manage it, but she may also reveal less about her bias in the first place. As we show, the latter effect can dominate, and the sender can be better off by not knowing exactly how she is being perceived.

Section 4.3 then explores a trading game in an extension of our methods that allows both players to affect the public signal. The sender is a trader who knows the true value of an asset, while the receiver is a second trader who only sees a noisy leakage of the sender's orders. The public signal is the total order flow, which is used to set the asset's price. As higher sender types buy more shares, they expect the receiver to be more optimistic about the asset. Anticipating an upward drift in future prices, the history-inference effect induces high types to buy even more shares today. This effect amplifies separation and, coupled with the information conveyed by the receiver's trades, leads to an extra layer of price impact emerging. The sender responds by slowing down her trades relative to a world in which no leakage takes place, yet price impact can be higher through the receiver's trades gradually correlating more with the type—a form of history-inference effect linked to the receiver's signaling. Low degrees of insider trading can then coexist with highly responsive prices.

Existence of LME and technical contribution. Having encoded the signals' histories into the belief states, the coefficients attached to them must be time-dependent to allow time-horizon and learning effects that influence the sender's behavior. These coefficients are found via a best-response problem for the sender, where the representation again plays a key role. The starting point is that the receiver must account for the total signaling done by the sender—history-inference effect included—to correctly learn from his private signal. The representation then shapes the receiver's posterior variance, which measures the extent of this player's learning. With Gaussian updating, however, this variance affects the posterior mean's responsiveness to new information, and hence it shapes the sender's ability to influence the receiver's belief—thus, this variance also affects the sender's second-order belief as a proxy for the receiver's counterpart. But this means that a representation under linear strategies is necessary to set up the sender's best-response problem. In this problem (Section 3.3), the second-order belief is used to evaluate the profitability of any strategy.

In equilibrium, of course, the receiver must perfectly anticipate the strategy followed by the sender. The latter requirement implies that (i) the receiver takes an optimal action given his beliefs; and (ii) his beliefs are correct at all times. While the receiver's myopia simplifies (i)—because his own coefficients become a simple function of the sender's contemporaneous ones—this is not a major advantage.[2] The real hurdle is the feedback behind (ii): the sender's signaling coefficients affect the variance-representation pair, which in turn affects the evolution of the belief states, and thus ultimately the choice of coefficients themselves.

Such a feedback loop is obviously present in any dynamic signaling game; the novelty is how it plays out here. Concretely, in Section 5 we show that the LQG structure means that finding an LME boils down to solving a system of ordinary differential equations (ODEs) with a mix of initial and terminal conditions. Such a *boundary value problem* (BVP) consists of ODEs for the coefficients in the sender's strategy, but also two extra ODEs: one for receiver's posterior variance and another for the weight attached to the type in the representation, which captures the sender's learning about the receiver's inferences. These "learning ODEs" are fully coupled with the former "behavior ODEs," in a reflection of the feedback at play.

This problem is complex because of the presence of multiple ODEs in both directions:

[2] In Section 6, we show that the same LMEs arise if a forward-looking receiver faces a *prediction problem*—as in our coordination and reputation applications—and how our methods apply beyond this case.

the learning ODEs are traced forward in time from their initial conditions, while the behavior ODEs are traced backward from the end game via backward induction. Existing work—discussed shortly—has dealt with settings in which only one learning ODE arises due to the symmetry of the environments studied: all the players signal and learn at the same rate. When this occurs, a traditional one-dimensional shooting argument invoking the intermediate-value theorem can be used to find a solution to the BVP. If signaling is asymmetric, and hence the shooting problem multidimensional, this approach does not apply.

Our contribution is to introduce a natural *fixed-point approach* to this issue. Indeed, given functions that proxy for solutions to our learning ODEs, we obtain candidate equilibrium coefficients by solving their respective ODEs backwards. Equipped with these, we obtain solutions to the learning ODEs by solving them forward. With this two-step shooting method, we construct an infinite-dimensional fixed-point problem over candidate learning functions, to which Schauder's theorem applies. Via this approach, Theorem 1 in Section 5 shows the existence of LMEs for horizon lengths up to a threshold time that is inversely proportional to the prior variance about the sender's type, *for all discount rates.* We discuss this method extensively in Section 6: in particular, how it is a major step forward in the literature and, by virtue of handling multiple ODEs in both directions, how it can be implemented in other LQG games or other settings in which a similar feedback loop via ODEs is at play.

Related literature With private monitoring, it can be challenging to compute distributions over rivals' histories, as these grow over time. Further, since such beliefs vary with a player's own past behavior, the game's structure differs between on- and off-path histories from any player's perspective (Kandori, 2002). These issues are usually absent if signals are public, such as with imperfect public monitoring (Abreu et al., 1990): if actions depend on commonly observed signals, everyone knows what a rival should do at all times; and given any public history, a player's best response remains such regardless of her past actions.

Past work has dealt with these issues in games with multi-sided private monitoring, absent incomplete information. Ely and Välimäki (2002) look for mixed-strategy equilibria where, by construction, beliefs about histories are irrelevant. Belief-dependent equilibria instead arise in Mailath and Morris (2002), who examine strategies represented by finite automata; players then form beliefs about a finite set of states, but those beliefs can depend on their own private histories. Building on this, Phelan and Skrzypacz (2012) show how to find equilibria by only looking at extreme beliefs of such states. Our LME are also belief-dependent and based on a reduction of the inference problem (to a finite set of real-valued, evolving states), but we pin down the sender's incentives *at all possible values of her second-order belief.* This latter state varies with the sender's past behavior, and the fact that it is spanned by the rest of the states only along the path of play reflects that the game changes after deviations.

Our model is one of dynamic noisy signaling in which the players' beliefs are private at all times due to the interplay between unobserved actions and private information.[3] LQG models have proven useful in this area, provided the environment has sufficient public information and/or symmetry. For instance, Foster and Viswanathan (1996), Back et al. (2000), and Bonatti et al. (2017) examine symmetric multi-sided incomplete information when everyone learns from an imperfect public signal of behavior; while first-order beliefs are private, the public signal structure eliminates the need for higher-order beliefs. Bonatti and Cisternas (2020) in turn examine two-sided signaling when firms price discriminate based on observing private signals of a consumer's past behavior; the prices firms set, however, fully reveal their beliefs. In none of these papers are higher-order beliefs needed as states; nor does a muldimensional BVP arise, as any non-trivial learning is either symmetric or one-sided.[4]

Regarding our applications, our coordination game is reminiscent of the team theory of Marschak and Radner (1972), where players' incentives are aligned to study information frictions in organizations; see Dessein and Santos (2006) and Rantakari (2008) for static models using quadratic preferences. In turn, private (hence, subjective) evaluations of performance have been studied in principal-agent models such as MacLeod (2003), albeit with complete information. On reputation, Bouvard and Lévy (2019) study a model with quadratic payoffs and symmetric Gaussian uncertainty in which beliefs are public in the linear equilibrium studied. And on trading, Yang and Zhu (2020) find that mixed-strategy equilibria can arise if there is leakage of an informed trader's behavior; with only two rounds of trading, the problem of how a player's own histories are aggregated to forecast a rival's belief is absent.

Finally, this paper belongs to a literature studying incentives using continuous-time methods. Sannikov (2007) examines games with imperfect public monitoring; Faingold and Sannikov (2011) reputation effects with behavioral types; Cisternas (2018) games of symmetric incomplete information; and Bergemann and Strack (2015) dynamic revenue maximization.

2 Model

In this section, we lay out our baseline model for examining two-player *dynamic noisy signaling* games when the ex ante informed player does not directly observe the signals of her actions. The framework is general enough to constitute a class of games, but it does not

[3] Noisy signaling with public beliefs has been extensively studied: in classic static settings (e.g., Matthews and Mirman, 1983; Carlsson and Dasgupta, 1997), the receiver's (prior) belief is common knowledge when the sender acts; and this also occurs in dynamic settings with observable actions and exogenous public signals (e.g., Kremer and Skrzypacz, 2007; Daley and Green, 2012; Kolb, 2019; Gryglewicz and Kolb, 2021).

[4] Private beliefs can also arise with *exogenous* private signals of a sender's type (Feltovich et al., 2002; Cetemen and Margaria, 2020; Kolb et al., 2021), or if types exhibit correlation (e.g, Cetemen et al., 2023).

exhaust the realm of settings that we can analyze. To make these points, we develop two applications under the umbrella of this model, and a third one based on an extension of it.

Model An informed agent—the "sender" (she)—interacts with an ex ante uninformed agent—the "receiver" (he)—continuously over a time interval $[0, T]$, $T < \infty$. The sender has payoff-relevant private information (her *type*) denoted by $\theta \in \mathbb{R}$. This type is normally distributed with mean $\mu \in \mathbb{R}$ and variance $\gamma^o > 0$, where (μ, γ^o) are model parameters.

We denote the sender's chosen action at time t by a_t, while the receiver's analog is denoted by $\hat{a}_t$, $t \in [0, T]$. Both actions take values over the real line. The sender is forward looking, and given realized action paths $(a_t)_{t\in[0,T]}$ and $(\hat{a}_t)_{t\in[0,T]}$, her ex post payoff is given by

$$\int_0^T e^{-rt} u(a_t, \hat{a}_t, \theta)\, dt + e^{-rT}\psi(\hat{a}_T), \tag{1}$$

where $u : \mathbb{R}^3 \to \mathbb{R}$ is a quadratic function and $r \geq 0$ is her discount rate. The terminal payoff ψ in turn exhibits a dependence on the receiver's endgame action, thus resembling the traditional sequentiality of classic one-shot signaling games. Given the nature of the applications that we study, we assume that $\psi : \mathbb{R} \to \mathbb{R}$ is a concave quadratic, which includes the case where ψ is linear or identically zero.

The receiver is assumed to be myopic, and thus concerned only about maximizing his flow utility at all instants of time. This assumption is useful for isolating how the sender's incentives vary due to the presence of higher-order uncertainty, and we discuss its relaxation in Section 6. Given realized actions a_t and $\hat{a}_t$, this player's ex post time-t payoff is denoted

$$\hat{u}(a_t, \hat{a}_t, \theta) \tag{2}$$

with $\hat{u} : \mathbb{R}^3 \to \mathbb{R}$ also a quadratic function. We will be interested in the case where u and $\hat{u}$ are strictly concave in a and $\hat{a}$, respectively; i.e., taking actions is costly for each player according to a quadratic function. For simplicity, we set $\partial^2 u/\partial a^2 = \partial^2 \hat{u}/\partial \hat{a}^2 = -1$; with quadratic preferences, this simply amounts to a normalization of the players' payoffs.

As argued, the sender knows θ at the outset, while the receiver only knows its distribution $\theta \sim \mathcal{N}(\mu, \gamma^o)$ (and this is common knowledge). There are also two one-dimensional noisy signals of the players' actions. In this baseline model, these signals have a *product structure*:

$$dX_t = \hat{a}_t dt + \sigma_X dZ_t^X \quad \text{and} \quad dY_t = a_t dt + \sigma_Y dZ_t^Y, \tag{3}$$

where Z^X and Z^Y are orthogonal Brownian motions, while σ_X and σ_Y are strictly positive volatility parameters. Our key innovation is to make Y—which carries information about the sender's actions—*privately observed by the receiver*; meanwhile, X carrying the receiver's action remains *public*. This mixed private-public information structure is important for our construction, but it is also appropriate for two reasons: it makes the departure from the existing literature minimal, and it suits the applications we study.[5]

The signals in (3) have full support, so the players cannot observe each other's actions. As the sender conditions her actions on her type, the receiver will then rely on his private signal Y to update his belief about θ. Our focus is on the cases in which the sender needs to forecast the resulting private belief for her best response.[6] The next assumption narrows the analysis to those non-trivial cases; subscripts in utility functions denote partial derivatives.

Assumption 1. *(i)* $u_{a\theta} \neq 0$*; (ii)* $|\hat{u}_{\hat{a}\theta}| + |\hat{u}_{a\hat{a}}| \neq 0$*; (iii)* $|u_{a\hat{a}}| + |u_{\hat{a}\hat{a}}| + |\psi_{\hat{a}\hat{a}}| \neq 0$.

By part (i), the sender's action is sensitive to her type, so there is scope for information transmission. Part (ii) is needed for the receiver's action to be sensitive to his private belief; this happens when he cares about the type directly ($\hat{u}_{\hat{a}\theta}$ term) or when he does so indirectly through the sender's action ($\hat{u}_{a\hat{a}}$ term). Part (iii) in turn guarantees the use of a second-order belief: in the sender's utility, either a non-trivial strategic interaction term ($u_{a\hat{a}} \neq 0$), or a nonlinearity stemming from the receiver's action ($|u_{\hat{a}\hat{a}}| + |\psi_{\hat{a}\hat{a}}| \neq 0$) will force the sender to forecast the receiver's belief to determine her optimal course of action. In Section 5, we complement these conditions with minimal technical ones that ensure the existence of equilibria in which there is separation through the second-order belief channel.

Examples Let us briefly make our model concrete by illustrating three examples that we explore in Section 4. The first two are specific instances of our baseline model, while the third is based on an extension of our methods presented in the Supplementary Appendix.[7] (The multiplicative factors are simply used to conform to our normalization of payoffs.)

1. A coordination game. Suppose that the players' payoffs are given by

$$\text{sender: } \frac{1}{4}\int_0^T e^{-rt}\{-(a_t-\theta)^2-(a_t-\hat{a}_t)^2\}dt; \quad \text{receiver: } \hat{u}(a_t,\hat{a}_t,\theta) = -\frac{1}{2}(\hat{a}_t-a_t)^2.$$

Consider a team/organization: a leader (the sender) tries to adapt her organization to new economic conditions—the state of the world, θ—but successful adaptation requires coordination between her actions and those of the rest of the organization (the receiver). For simplicity, the receiver only wants to coordinate. In reality, one expects that when leadership

[5]Other papers displaying a public "flavor" in their information structures are Bhaskar and Obara (2002), studying almost-perfect monitoring, and Mailath and Morris (2002), examining almost-public monitoring.

[6]The public signal X will be used in this forecasting exercise, but it will not be the sole input.

[7]Specifically, both signals' drifts are allowed to depend on both players' actions in an additive way.

introduces important changes, organizations' inferences are likely subjective, so both parties need to think about each other's understanding to coordinate. We model this situation as a signaling game in which the receiver develops a private belief about θ via observing Y; in turn, X can be seen as a traditional imperfect public signal of the receiver's performance.[8]

2. A reputation game. For notational simplicity, set the prior mean μ to zero and consider:

$$\text{sender: } \frac{1}{2}\left[-\int_0^T e^{-rt}(a_t-\theta)^2dt - e^{-rT}\psi\hat{a}_T^2\right]; \quad \text{receiver: } \hat{u}(a_t,\hat{a}_t,\theta) = -\frac{1}{2}(\hat{a}_t-\theta)^2,$$

where $\psi \in \mathbb{R}_+$. Suppose now that θ represents a privately known *bias* on a relevant issue, with the prior mean $\mu = 0$ capturing an unbiased type. The sender (e.g., a politician or expert), finds it costly to take actions away from her type ($-(a_t-\theta)^2$ term) but she benefits from appearing as unbiased at a terminal time T; this is because the receiver's action is, at all times, his current best estimate of the sender's type, and $\hat{a}_T = \mu = 0$ fully eliminates the terminal loss. For interpretation, the receiver could be a news outlet that gets private signals Y of the sender's behavior and that reports its perception of the bias;[9] the reporting process X is imperfect, but fair on average (the shocks have zero mean)—and naturally public.

3. A trading game. Consider a public signal $dX_t = (a_t+\hat{a}_t)dt + \sigma_X dZ_t^X$ and payoffs

$$\text{sender: } \int_0^T \left[(\theta - \mathbb{E}[\theta|\mathcal{F}_t^X])a_t - \frac{a_t^2}{2}\right]dt; \quad \text{receiver: } (\theta - \mathbb{E}[\theta|\mathcal{F}_t^X])\hat{a}_t - \frac{\hat{a}_t^2}{2}.$$

The sender is an informed trader who knows the fundamental value θ of an asset, while the receiver is an ex ante uninformed investor who sees a leakage Y of the sender's actions.[10] The term $\mathbb{E}[\theta|\mathcal{F}_t^X]$ corresponds to the asset's price at time t, based on the public *total order flow* X, as in Kyle (1985). For the sender (and analogously, for the receiver) $(\theta - \mathbb{E}[\theta|\mathcal{F}_t^X])a_t\, dt$ represents her trading gains over $[t, t+dt)$ if $a_t\, dt$ units are bought/sold in that instant; in turn, the convex costs $a_t^2/2$ capture other types of transaction costs.[11] The game departs from our baseline model because (i) the public signal carries the informed player's action too; and (ii) there is a "third action"—the price—based exclusively on the public information.

[8]Obviously, communication renders the problem trivial; the interpretation is of a situation where, what should be done (here, θ, of dimension 1), is considerably more complex than the richness of the communication channel available (completely shut down here; hence, of dimension zero).

[9]Actions such as voting, contributions, favors, statements to groups of influence, etc. often have a private nature, and hence are likely to be leaked with error, justifying the noise in Y.

[10]Yang and Zhu (2020) argue that, by handling retail order flow (proxy for noise trading), proprietary trading firms can construct private signals of institutional investors' (proxy for informed traders) behavior.

[11]E.g., taxes from trades (Subrahmanyam, 1998). Additional costs from large "long" positions also arise from limited resources within a fund; and on the "short" side, due to the use of brokers for borrowing shares.

Strategies and Equilibrium Concept. Since the full-support monitoring prevents each player from seeing what the other has done, deviations by the counterparty go undetected. This means that, from the perspective of any player, the only off-path histories are those in which that same player himself/herself has deviated. We can therefore use the Nash equilibrium concept for defining the equilibrium of the game, leaving off-path behavior unspecified for now; indeed, as is well-known in games with unobserved actions, imposing full sequential rationality—i.e., specifying optimal behavior also after deviations—does not further refine the set of equilibrium outcomes.[12]

From this perspective, a (pure) strategy for the sender corresponds to any square-integrable real-valued process $(a_t)_{t\in[0,T]}$ that is progressively measurable with respect to the filtration generated by (θ, X). For the receiver, the measurability restriction is with respect to (X, Y), with the same integrability condition at play.[13] Let $\mathbb{E}_t[\cdot]$ and $\hat{\mathbb{E}}_t[\cdot]$, $t \in [0,T]$, denote the sender's and receiver's expectation operators, respectively.

Definition 1 (Nash equilibrium). *A pair of strategies $(a_t, \hat{a}_t)_{t\geq 0}$ is a Nash equilibrium if: (i) the process $(a_t)_{t\in[0,T]}$ maximizes $\mathbb{E}_0\left[\int_0^T e^{-rt}u(a_t,\hat{a}_t,\theta)\,dt + e^{-rT}\psi(\hat{a}_T)\right]$; and (ii) for each $t\in[0,T]$, $\hat{a}_t$ maximizes $\hat{\mathbb{E}}_t[\hat{u}(a_t,\hat{a}_t,\theta)]$ when $(\hat{a}_s)_{s<t}$ has been followed.*

The LQG structure suggests looking for Nash equilibria in strategies that are *linear* functions of the signals observed by each player. While this is a simple task in static settings, it is far more challenging in dynamic environments. Indeed, recall that evaluating the candidacy of an equilibrium profile necessarily requires assessing the profitability of deviations; but with incomplete information and unobserved actions, the sender will find it optimal to condition on more information than (θ, X) after she deviates, in a reflection that the game's structure changes after deviations. The next section formalizes these ideas. Specifically, we will develop a method for finding Nash equilibria that relies on imposing full sequential rationality for the sender under a richer set of strategies than above. In this equilibrium, the players linearly aggregate their relevant histories as play unfolds, and the sender's actions will effectively be a function of (θ, X) along the path of play (i.e., when the strategies prescribed by the equilibrium are followed).

[12]See Mailath and Samuelson (2006), pp. 395-396. With hidden actions, a Nash equilibrium fails to be sequentially rational only if it dictates suboptimal behavior for a player after her own deviation. Since such off-path histories are not reached, the same outcome arises if optimal behavior is specified after the deviation.

[13]Square integrability refers to $\int_0^T a_t^2\,dt$ and $\int_0^T \hat{a}_t^2\,dt$ being finite in expectation. Coupled with progressive measurability for the dependence of actions on information (Karatzas and Shreve, 1991, Ch. 1), it ensures that a (strong) solution to (3) exists, i.e., that the outcome of the game is well-defined. These conditions are standard in continuous-time optimization; see Ch. 1.3 and 3.2 in Pham (2009) for decision problems.

3 Equilibrium Analysis: Linear Markov Equilibria

Let us begin by offering a high-level overview of the idea behind our construction.

3.1 Belief States: An Overview

The presence of incomplete information opens the possibility of our players employing strategies that depend on the beliefs that originate from the sender's type. With quadratic preferences, one first expects the *means* of such posterior beliefs—henceforth, *beliefs*—to become the key states, and second that these states are used linearly. Specifically, in addition to the sender's type θ, our construction employs the following belief states:

$$\hat{M}_t := \hat{\mathbb{E}}_t[\theta], \quad M_t := \mathbb{E}_t[\hat{M}_t], \quad \text{and} \quad L_t := \mathbb{E}[\theta|\mathcal{F}_t^X]. \tag{4}$$

Here, $\hat{M}_t$ is the receiver's first-order belief, i.e., his belief about the sender's type; M_t is the sender's second-order belief, i.e., her belief about the receiver's first-order belief; and L_t is the belief about θ using the public information exclusively, $t \in [0, T]$. These states will encode how the players aggregate their own histories (the signals they have observed, and possibly their past actions taken), to forecast what their counterparty knows and hence might do.

Equipped with these states, we characterize equilibria in which, on and off the path of play, the sender and receiver take actions according to *linear Markov strategies* of the form

$$a_t = \beta_{0t} + \beta_{1t}M_t + \beta_{2t}L_t + \beta_{3t}\theta \tag{5}$$

$$\hat{a}_t = \delta_{0t} + \delta_{1t}\hat{M}_t + \delta_{2t}L_t. \tag{6}$$

The coefficients β_{it}, $i = 0, 1, 2, 3$, and δ_{jt}, $j = 0, 1, 2$, will be differentiable functions of time. Intuitively, having encoded the players' histories into their belief states, the weights attached to them must be allowed to be time-dependent to capture how behavior changes due to end-game and learning effects that vary without abrupt changes over time.

For illustration, consider our coordination game, where $u(a, \hat{a}, \theta) \propto -(a - \theta)^2 - (a - \hat{a})^2$ and $\hat{u}(a, \hat{a}, \theta) \propto -(a - \hat{a})^2$. First, the type is obviously a relevant state for the sender because she cares about her action's proximity to it. The players' coordination motive then kicks in: because the receiver wants to match the sender's action, $\hat{M}$ is relevant for the receiver; as a result, the sender is forced to forecast $\hat{M}$, and the second-order belief M appears. Naturally, the sender would like $\hat{M}$ to always coincide with θ; thus, from the sender's perspective, discrepancies between M—which proxies for $\hat{M}$—and θ matter for her payoffs. But the sender's incentives to correct such discrepancies depend both on how much time is left in the game—with longer horizons creating stronger incentives—and on her ability to move the receiver's belief—which depends on the latter's precision/variance, which will be deterministic. Thus, the coefficients in the strategy must be time-dependent.

More generally, Assumption 1 part (iii) ensures that M is needed as a proxy for $\hat{M}$ for the sender to determine her best course of action in every instance of our class of games. This state is central to our construction; let us preview two of its properties that we establish later. First, after all private histories of the sender, M_t is an explicit linear function of her past actions $(a_s)_{s<t}$ and past realizations of the public signal $(X_s)_{s<t}$ (Remark 1). The dependence on X is clear given that this signal carries the receiver's action, and hence ultimately $\hat{M}$. Due to the private monitoring, however, the sender also relies on her *past play*: higher past actions are indicative of higher realizations of Y observed by the receiver, so M is higher for any fixed public history of X; by contrast, if Y were public, past behavior would be irrelevant, as the receiver's belief would be fully determined by the realizations of Y.

The upshot is that, because of this dependence on past play, M is also the sender's private information, and must be forecasted by the receiver. This brings us to the second property: along the path of (5)–(6), M_t is a convex combination of θ and L_t, where the dependence on the type stems from M_t conditioning on the sender's past actions (Lemma 1). There are two consequences of this *representation of the sender's second-order belief*. First, when it is used by the receiver to forecast M, the public state L becomes payoff-relevant, and so it must be added to the strategies. Second, this private-public structure of the representation ensures that the players do not need additional higher-order beliefs to forecast their behavior—i.e., the state space does not explode. The next section formally presents this result, which is instrumental to our analysis—even beyond ensuring that the state space is bounded.

3.2 Representation of the Second-Order Belief

Representation Suppose that the players follow the linear-Markov strategies (5)–(6). Given the LQG structure, it is natural to expect a representation of the form

$$M_t = \chi_t \theta + (1 - \chi_t) L_t, \tag{7}$$

where $L_t := \mathbb{E}[\theta | \mathcal{F}_t^X]$ (or by the law of iterated expectations, $\mathbb{E}[\hat{M}_t | \mathcal{F}_t^X]$) and $(\chi_t)_{t \in [0,T]}$ is deterministic. Intuitively, to forecast the receiver's belief, the sender takes the information in the public signal and adjusts it based on her additional private information—her own past actions, carrying θ under (5).[14] The weight χ captures how this balancing changes over time:

[14]For intuition, note that with pure strategies, the *outcome* of the game should be a function of the signals available to the players. Thus, M must be a function of θ and X exclusively.

early on, past actions have little forecasting value, so χ should be low; but as more signaling takes place and the sender expects the receiver to learn the type, θ is given a higher weight.

All this is intuitive, but in equilibrium the players need to know the *exact* dependence of (χ, L) on the associated linear Markov strategy profile. This is where our first contribution lies: we establish laws of motion for χ and L as a function of the coefficients in (5)–(6). Let us briefly outline our constructive approach: we begin by assuming that (7) holds with $(L_t)_{t\in[0,T]}$ a generic process depending only on the public information. Inserting (7) into (5) then delivers the sender's actions along the path of play of a strategy profile (5)–(6):

$$a_t = \alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}\theta, \tag{8}$$

$$\text{where} \quad \alpha_{0t} := \beta_{0t}, \quad \alpha_{2t} := \beta_{2t} + \beta_{1t}(1-\chi_t), \quad \text{and} \quad \alpha_{3t} := \beta_{3t} + \beta_{1t}\chi_t. \tag{9}$$

In particular, note that information transmission is guided by the total weight on the type, α_{3t}, which carries χ_t—this means that the receiver needs to know the exact form of χ partly because he must anticipate the precise informational content in his private signal Y.

Assuming that the sender's actions satisfy (8), the receiver's problem of learning θ from Y is (conditionally) Gaussian (Liptser and Shiryaev, 1977, Theorems 12.6 and 12.7). The receiver's belief is characterized by a stochastic mean $(\hat{M}_t)_{t\in[0,T]}$ and a deterministic variance

$$\gamma_t := \hat{\mathbb{E}}_t[(\theta - \hat{M}_t)^2],\ t \in [0,T],$$

where we have omitted the hat symbol for notational convenience.[15] Importantly, the linearity of the signal structure renders the sender's problem of filtering $\hat{M}$ using X under (6) (conditionally) Gaussian again: a second mean-variance pair emerges, where the mean M depends explicitly on the sender's past actions (i.e., for any given history of the public signal, changes in actions imply a shift in the mean of her belief). Imposing that the second-order belief M coincides with (7) when (5) is followed then delivers differential equations for (χ, L).

Lemma 1. *Suppose that (X, Y) is driven by (5)–(6) and the receiver believes that (7) holds, with $(L_t)_{t\in[0,T]}$ a process that depends only on the public information.*[16] *Then (7) holds at all times if and only if $L_t = \mathbb{E}[\theta|\mathcal{F}_t^X]$ and $\chi_t = \mathbb{E}_t[(\hat{M}_t - M_t)^2]/\gamma_t$, where*

$$\dot{\gamma}_t = -\frac{\gamma_t^2(\beta_{3t} + \beta_{1t}\chi_t)^2}{\sigma_Y^2}, \qquad \gamma_0 = \gamma^o, \tag{10}$$

$$\dot{\chi}_t = \frac{\gamma_t(\beta_{3t} + \beta_{1t}\chi_t)^2(1-\chi_t)}{\sigma_Y^2} - \frac{\gamma_t\chi_t^2\delta_{1t}^2}{\sigma_X^2}, \qquad \chi_0 = 0, \tag{11}$$

[15]The law of motion of $\hat{M}$ is presented in (A.1) in the appendix. It responds to changes in Y in a linear way, with a sensitivity that is proportional to γ (i.e., more precise beliefs are less responsive to news).

[16]Formally, $(L_t)_{t\in[0,T]}$ can be any square-integrable process progressively measurable w.r.t. $(\mathcal{F}_t^X)_{t\in[0,T]}$.

$$dL_t \;=\; (l_{0t} + l_{1t}L_t)dt + B_t dX_t, \qquad\qquad L_0 = \mu, \tag{12}$$

with (l_{0t}, l_{1t}, B_t) *deterministic and given in* (A.10)*. Also,* $0 < \gamma_t \leq \gamma^o$ *and* $0 \leq \chi_t < 1$ *for all* $t \in [0, T]$*, with strict inequalities over* $(0, T]$ *if* $\beta_{3,0} \neq 0$*.*

From the lemma, the public state in the representation must coincide with $\mathbb{E}[\theta|\mathcal{F}_t^X]$; its linearity in the history of X—virtue of the Gaussian learning—is clear from (12) being linear in L and in the increment dX. The lemma also characterizes the weight χ_t attached to the type in (7) as a ratio of the players' posterior variances. This is not surprising, as the players' learning is necessarily connected: if the sender has signaled her type more aggressively, she will expect the receiver to be more certain about it, so lower values of γ are likely associated with higher values of χ in the representation. These observations are clear from the system of ODEs (10)–(11) for (γ, χ): the system is fully coupled (i.e., χ affects $\dot{\gamma}$ while γ affects $\dot{\chi}$), thus reflecting that learning is interconnected; and higher values of the signaling coefficient $\alpha_3 = \beta_3 + \beta_1\chi$ prompt both a faster decay of γ and a faster growth of χ.[17]

The advantage of this ODE system, which we will leverage later in the paper, is that it explicitly tells us how these *learning coefficients* γ and χ depend on the coefficients in the player's strategies linked to information transmission: β_3 and β_1 for the sender, and δ_1 for the receiver. The initial conditions in (10)–(12) simply reflect the absence of higher-order uncertainty in the beginning of the game: $M = \hat{M} = L = \mu$ at $t = 0$, and so $\chi_0 = 0$ in the representation. In turn, the bounds $\gamma_t < \gamma^o$ and $0 < \chi_t$ when $\beta_{3,0} \neq 0$ capture that some learning must take place if signaling occurs at $t = 0$, while $0 < \gamma_t$ and $\chi_t < 1$ capture that, with finite signaling coefficients, this learning is never complete. Finally, the last term in (11) reflects how the informativeness of the public signal affects the weight on the type in the representation: as the signal-to-noise ratio of X, δ_{1t}^2/σ_X^2, grows, more downward pressure is put on the growth of χ. In other, words, as the public signal becomes more informative, the sender increasingly favors this source of information over his own past history of play.

Equipped with this characterization, we can establish two important observations.

The belief states are sufficient statistics Observe that the receiver always believes that the representation holds. Indeed, the receiver always assumes that the sender uses (5) because deviations are undetected; by construction, the representation holds from his perspective if he follows (6). But the same occurs if he deviates from (6): since his own deviations are hidden, the receiver expects the sender to believe that (i) he always uses (6) and that (ii) the representation holds from his perspective. From the receiver's standpoint, therefore, the sender always constructs her second-order belief as if (5)–(6) is being followed.

[17] If $\sigma_X = +\infty$ (the public signal is uninformative), the solution to (10)–(11) satisfies $\chi = 1 - \frac{\gamma_t}{\gamma^o}$, so γ and χ are inversely related—see Lemma B.1 in the appendix, and Section 6 for a generalization.

As a result, the players do not need additional higher-order beliefs to infer their counterparties' states if these are being used linearly: in fact, the receiver's *third-order belief*, via the representation, combines $\hat{M}$ and L linearly; so the sender's *fourth-order belief* is also a linear combination between M and L; which means that the receiver resorts to the representation again for his *fifth-order belief*, and so forth. Thus, the state space does not explode.

Signaling: history-inference effect The representation captures the new effects that private monitoring brings to signaling games: if different types take different actions in equilibrium, their reliance on past play to assess the continuation game will necessarily lead them to hold different beliefs M, even after seeing the same public information. In the sender's strategy (5), this leads to M becoming an additional channel for separation on top of the direct contribution that the type has on behavior. The signaling implications of this new channel are captured by $\beta_{1t}\chi_t$ in the signaling coefficient $\alpha_3 = \beta_1\chi + \beta_3$—we refer to it as the *history-inference effect* on signaling, which we study in Section 4. Importantly, this effect is absent if the environment is public: after observing a history of signal realizations of Y, all sender types would agree on the value that the receiver's belief takes.

3.3 The Long-Run Player's Best-Response Problem

The representation reveals that, on the path of play of the linear Markov profile (5)–(6), the sender's actions depend only on the pair (θ, L) via $a_t = \alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}\theta$ (see (8)–(9)). To determine these coefficients, however, the sender must evaluate deviations from the previous action path. The need for another state is obvious, as both θ and L are unaffected by the sender's actions (the latter because her action does not influence the public signal). That additional state is our second-order belief M. The next result presents laws of motion for M and L for arbitrary strategies of the sender (up to technical conditions specified shortly).

Lemma 2 (Controlled dynamics)**.** *From the sender's perspective, if she follows $(a'_t)_{t\in[0,T]}$,*

$$dM_t = \frac{\gamma_t\alpha_{3t}}{\sigma_Y^2}(a'_t - [\alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}M_t])dt + \frac{\gamma_t\chi_t\delta_{1t}}{\sigma_X}dZ_t \tag{13}$$

$$dL_t = \frac{\gamma_t^X\chi_t\delta_{1t}}{\sigma_X^2}[\delta_{1t}(M_t - L_t)dt + \sigma_X dZ_t], \tag{14}$$

where $Z_t := \frac{1}{\sigma_X}[X_t - \int_0^t(\delta_{0s} + \delta_{1s}M_s + \delta_{2s}L_s)ds]$ is a Brownian motion, and $\gamma_t^X := \frac{\gamma_t}{(1-\chi_t)} = \mathbb{E}[(\theta - L_t)^2|\mathcal{F}_t^X]$. Also, $\mathbb{E}_t[(\hat{M}_t - M_t)^2] = \gamma_t\chi_t$ for any such $(a'_t)_{t\in[0,T]}$.

When the sender deviates from (5), the representation need not hold (as it assumes that the sender follows a linear Markov strategy) and hence the sender must keep track of M and

L separately. The appearance of an additional payoff-relevant state demonstrates that the game's structure changes after deviations: the sender will behave differently there because her past behavior will lead her to perceive a different continuation game, an issue that we will explore in more depth in our applications. Note the importance of the second-order belief state in this respect: since the sender's actions only affect M in (13)–(14), any approach for finding an equilibrium must be linked to establishing optimality with respect to M. It is for this reason that we start with an extended strategy (5) that involves both M and L.

The law of motion of M, (13), encapsulates how the sender expects the receiver's private belief $\hat{M}$ to evolve in response to different continuation strategies by the sender. On the other hand, changes in L matter for the sender's incentives because the receiver uses this state to predict M in his third-order belief exercise. To understand why M feeds into L in (14), suppose that the sender has taken "high" actions in the past: expecting high values of the receiver's belief, the sender predicts a steep growth in L through the channel of the receiver's actions influencing X. Our applications will shed light on these "prediction channels."

But the sender's ability to influence $\hat{M}$ naturally depends on the extent of the receiver's learning: in (13), the sender's action has a slope that is proportional to γ—the receiver's posterior variance—which falls as more learning has taken place. Importantly, this variance depends on χ through the signaling that occurs through the second-order belief channel. In other words, having upfront knowledge of a representation is necessary for setting up a best-response problem, because the receiver's learning must account for the total signaling done by the sender in equilibrium. Much of the complexity of the fixed point at play in these games in fact operates through this "variance" channel: the pair (γ, χ) depends on the signaling coefficients (5)–(6); but by also shaping the responsiveness of (M, L) to the sender's actions, the same pair affects the choice of coefficients in the strategies themselves.[18]

Remark 1. *To see why M is an explicit function of the past actions of the sender and past realizations of X, we can first insert the definition of Z_t into the law of motion* (13) *of M. This yields a dynamic that is linear in M, from which the solution M_t is a linear function of $(a_s, L_s, X_s)_{s<t}$; but the same procedure applied to* (12) *shows that L_s is a function of $(X_\tau)_{\tau<s}$ (because* (12) *holds on- and off-path from each player's perspective).*

We can now state the sender's best-response problem. By the last part of Lemma 2, the

[18] The responsiveness of (M, L) also depends on the (perceived) strength of the players' signaling: α_3 influences M, while δ_1 affects L. In the latter state, the variance with respect to the public information, γ^X, plays the role that γ plays in M: the appearance of χ multiplying $\delta_1\gamma_t^X$ stems from the covariance between θ and dX_t conditional on the public information taking the form $\gamma_t^X\chi_t\delta_{1t}$ after using the representation and that $\mathbb{E}_t[\theta|\mathcal{F}_t^X] = \mathbb{E}_t[M|\mathcal{F}_t^X]$. (And by Lemma 1, $\chi < 1$, so the law of motion of L is always well-defined.) Finally, the drift of M reflects that the sender expects $\hat{M}$ to be revised upward only when $a_t' > \mathbb{E}_t[\alpha_{0t}+\alpha_{2t}L_t+\alpha_{3t}\hat{M}_t]$, i.e., when she expects to beat the receiver's expectation of her behavior.

sender's posterior variance satisfies $\mathbb{E}_t[(\hat{M}_t - M_t)^2] = \gamma_t\chi_t$, i.e., it has a value invariant to deviations by the sender; intuitively, linear signals coupled with Gaussian noise imply that changes in the sender's actions simply shift the receiver's belief. Using this fact and that payoffs are quadratic, we obtain $\mathbb{E}_t[u(a_t, \delta_{0,t} + \delta_{1,t}\hat{M}_t + \delta_{2t}L_t, \theta)] = \mathbb{E}_t[u(a_t, \delta_{0t} + \delta_{1,t}M_t + \delta_{2t}L_t, \theta)] + \frac{1}{2}u_{\hat{a}\hat{a}}\delta_{1t}^2\gamma_t\chi_t$, and likewise for the terminal payoff ψ in place of u at $t = T$. In other words, no *higher moments* are needed as states—i.e., our states do capture the payoff-relevant aspects of the players' *own* histories. We conclude that, up to an additive constant in the sender's objective, her best-response problem consists of maximizing

$$\mathbb{E}_0\left[\int_0^T e^{-rt}u(a_t, \delta_{0t} + \delta_{1t}M_t + \delta_{2t}L_t, \theta)dt + e^{-rT}\psi(\delta_{0T} + \delta_{1T}M_T + \delta_{2T}L_T)\right] \tag{15}$$

subject to the dynamics (13)–(14) of (M, L), and where (γ, χ) follow the ODEs (10)–(11).[19] The space of admissible strategies for this problem is the set of $\mathbb{R}$-valued square-integrable processes $(a_t)_{t\in[0,T]}$ that are (θ, M, L)-progressively measurable. This space is richer than that used in the Nash equilibrium concept due to the explicit conditioning on past behavior via M, and it is the traditional strategy space in continuous-time optimization.[20]

Assuming a myopic receiver simplifies the determination of the coefficients $(\delta_0, \delta_1, \delta_2)$ in his strategy as simple functions of the sender's contemporaneous counterparts (and possibly of (γ, χ) too, via the inferences made). As argued, this is conceptually useful, but it provides only a modest technical advantage, as we explain in Section 6. From this perspective, a tuple $\vec{\beta} := (\beta_0, \beta_1, \beta_2, \beta_3)$ of differentiable functions of time induces a *linear Markov equilibrium* if $\beta_{0t} + \beta_{1t}M + \beta_{2t}L + \beta_{3t}\theta$ is an optimal policy for the sender when the coefficients $(\delta_0, \delta_1, \delta_2)$ satisfy the myopic best reply condition for the receiver:

$$\hat{a}_t := \delta_{0t} + \delta_{1t}\hat{M}_t + \delta_{2t}L_t = \arg\max_{\hat{a}'\in\mathbb{R}} \hat{\mathbb{E}}_t[\hat{u}(\alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}\theta, \hat{a}', \theta)]. \tag{16}$$

This notion of equilibrium is clearly *perfect* in that it specifies optimal behavior by the sender after deviations—and along the path of play of such a policy, $a_t = \alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}\theta$, where $(L_t)_{t\in[0,T]}$ follows (12) in Lemma 1, so a Nash equilibrium in linear strategies ensues.[21]

In the next section, we discuss the equilibrium coefficients that arise in each of our three applications, deferring the question of the existence of LME to Section 5.

[19] The sender's problem is, in practice, one of optimally controlling an *unobserved state* $\hat{M}$. We are allowed to filter first and then optimize due to the *separation principle*. See the proof of Lemma 2.

[20] See Chapter 3.1 in Pham (2009). Obviously, it is not limited to a linear use of the states and so forth.

[21] While deviations by the receiver do affect L, it is clear that no additional states other than $(t, L, \hat{M})$ are needed after deviations. Also, all the payoff-relevant histories for this player are reachable on-path, so the sequential rationality requirement is trivial for this player. All this is true if the receiver is forward-looking.

4 Applications

Our applications showcase how the history-inference effect can be a relevant phenomenon in a variety of settings—existence results for our coordination and reputation games fall under the umbrella of our general existence theorem of Section 5, while an existence result for our trading game is presented in Section S.4 in the Supplementary Appendix.

4.1 Application 1: A coordination game in organizations

From Section 2, up to positive factors, the payoffs take the form

$$\text{sender}: \int_0^T e^{-rt}\{-(a_t-\theta)^2-(a_t-\hat{a}_t)^2\}dt; \quad \text{receiver}: \hat{u}(a_t,\hat{a}_t,\theta)=-(\hat{a}_t-a_t)^2.$$

The interpretation is one of a leader (the sender) trying to introduce changes into an organization in order to adapt it to a new economic environment ($-(a_t-\theta)^2$ term), with those changes requiring a coordinated response by the rest of the organization (the receiver; $-(\hat{a}_t-a_t)^2$ term). Note that if the state of the world were common knowledge, everyone would coordinate on θ, and no losses would take place—all sender types thus have a *revelation* motive, but this is hindered by the imperfect, private, signals seen by the receiver. How does the leader, through her actions, guide the organization towards its new goal?[22]

Proposition 1. *Suppose that $r \geq 0$ and $\sigma_X \in (0,\infty)$. In any LME, the coefficients satisfy $\beta_{0t}=0$, $\beta_{1t}+\beta_{2t}+\beta_{3t}=1$, and $\alpha_{3t}:=\beta_{3t}+\beta_{1t}\chi_t>0$. On the equilibrium path, therefore,*

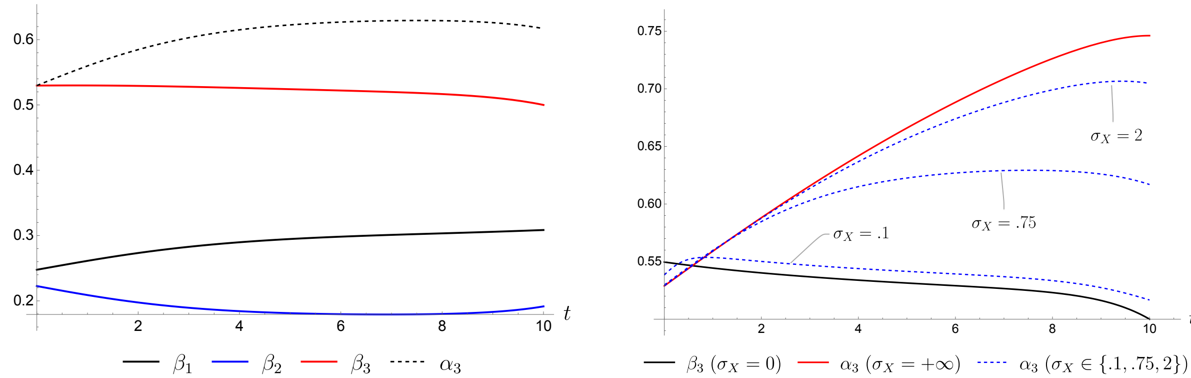
$$a_t=\alpha_{3t}\theta+(1-\alpha_{3t})L_t \quad \textit{and} \quad \hat{a}_t=\alpha_{3t}\hat{M}_t+(1-\alpha_{3t})L_t. \tag{17}$$

Further, if $r>0$, α_{3t} is non-monotonic and decreasing at T. And if the time horizon is not too long, $\beta_{3t}\in[1/2,1)$, $\beta_{1t},\beta_{2t}\in(0,1/2)$, and $\alpha_{3t}\in(0,1)$ can be shown analytically.

Panel (a) in Figure 1 illustrates typical coefficients $(\beta_1,\beta_2,\beta_3,\alpha_3)$: the leader's action is a *convex combination* of her states (θ,M_t,L_t), and also of (θ,L) in equilibrium via (17). Since the signs of these coefficients do not change, they can be explained by the incentives that arise at the endgame T when the players act myopically. The starting point is that the sender's adaptation motive leads to $\beta_3>0$, after which the coordination motive kicks in: since higher types take higher actions, higher receiver types $\hat{M}$ must also take higher actions, which induces the sender to take even higher actions via M ($\beta_1>0$). But this means that

[22] In organizations, information transmission through actions or practice is important due to knowledge often having a "tacit" form: Garicano (2000) describes it as "production know-how is [...] 'embodied' in individuals" (p. 875) such that it is "acquired [...] in the form of learning by doing" (p. 894).

the receiver must reciprocate after observing a higher value of L—due to this state being used to forecast M—in turn inducing the sender to attach a positive weight to L ($\beta_2 > 0$).[23]



(a) Sender's strategy coefficients ($\sigma_X = 0.75$). (b) Sender's signaling coefficient as σ_X varies.

Figure 1: Coefficients in the coordination game. Other parameters: $(\gamma^o, r, \sigma_Y) = (1, 1, 1.5)$.

Nonetheless, these coefficients depart from their myopic counterparts due to the dynamics at play. In particular, the signaling coefficient α_3 is *generically non-monotone*—hump-shaped, as the dashed lines in Figure 1. This is non-trivial because a traditional "signaling" logic would suggest a decreasing profile. Specifically, if more time remains, the leader might want to place more weight on her type because stronger signaling steers the follower's belief in the direction of the state of the world—the correction of the discrepancy between $\hat{M}$ and θ discussed in Section 3. In other words, the leader sacrifices coordination today to obtain "coordinated adaptation" tomorrow, which pays off only if there is enough future ahead. Indeed, β_3—the direct weight attached to the type—is strictly decreasing in the same figure.

The explanation lies in the history-inference effect at play. Indeed, since different types take different actions ($\alpha_3 > 0$ in (17)), higher types necessarily develop higher second-order beliefs M as in the representation (7). But the coordination motive then induces high types to take even higher actions, captured by $\beta_1 > 0$ in Figure 1. As a result, the possibility of types separating more over time emerges, reflected in the increasing pattern of $\alpha_3 := \beta_3 + \beta_1\chi$ through the $\beta_1\chi$ channel; this effect grows over time due to both the coordination incentive becoming stronger, and the second-order belief M reflecting the type more and more.

One natural way to incorporate these two "opposing" views—decreasing separation via steering vs. increasing separation via the history-inference effect—into a unified framework is to look at two extreme values of the noise in the public signal. If $\sigma_X = 0$, the public signal perfectly reveals the receiver's actions and potentially, in turn, his belief; the environment

[23]The myopic equilibrium is the fixed point of the system $a = \frac{1}{2}\theta + \frac{1}{2}\mathbb{E}[\hat{a}]$ and $\hat{a} = \hat{\mathbb{E}}[a]$, which prevails at $t = T$. Simple algebra yields $(\beta_0, \beta_1, \beta_2, \beta_3) = \left(0, \frac{1}{2(2-\chi)}, \frac{1-\chi}{2(2-\chi)}, \frac{1}{2}\right)$ and $(\delta_0, \delta_1, \delta_2) = \left(0, \frac{1}{2-\chi}, \frac{1-\chi}{2-\chi}\right)$.

is effectively *public*. At the other extreme, if the leader observes nothing she must rely only on her past play to forecast the receiver's belief. This case is obtained by setting $\sigma_X = +\infty$, which offers the maximum scope for the history-inference effect. The cases are very tractable.

Proposition 2. *Suppose that $\sigma_X \in \{0, +\infty\}$. Then, an LME exists for all $T > 0$ and $r \geq 0$.*

(i) If $\sigma_X = 0$, $a_t = \beta_{3t}\theta + (1-\beta_{3t})\hat{M}_t$, where $\frac{d\beta_{3t}}{dt} < 0$, $\beta_{3t} \in (1/2, 1)$, $t < T$, and $\beta_{3T} = 1/2$.

(ii) If $\sigma_X = +\infty$, $a_t = \alpha_{3t}\theta + (1-\alpha_{3t})\mu$, where $\alpha_{3t} \in (1/2, 1)$, $t \in [0, T]$. Also, $\frac{d\alpha_{3t}}{dt} > 0$ if $r > 0$ (and constant for $r = 0$).

When $\sigma_X = 0$, $\hat{M}$ is publicly observable. The signaling coefficient $\alpha_3 \equiv \beta_3$ is then decreasing, so separation effectively shrinks over time; but β_3 remains above the terminal value $1/2$ due to the steering motive at play. On the other hand, in the absence of a public signal (in which case $L \equiv \mu$), the history-inference effect coupled with the coordination motive reinforce separation over time. The monotone solid lines in Figure 1b illustrate this.

The non-monotone pattern of α_3 is thus the net effect of a decreasing steering incentive and an increasing history-inference effect. Two additional effects help explain why a hump-shaped pattern arises. First, since the history-inference effect increases the responsiveness of the receiver's belief to Y later in the game, the leader does not need to sacrifice as much on coordination relative to the public case: the leader reduces β_3 (to increase $\beta_1 + \beta_2$) in the beginning of the game, thus favoring the growth of $\beta_1\chi$—in Figure 1b, the dashed α_3 coefficients start below the public counterpart, and they exhibit an initial positive slope that is absent in the case $\sigma_X = 0$. Second, relative to the case $\sigma_X = +\infty$, introducing an informative public signal makes L carry more weight in the representation; the receiver's action then becomes more sensitive to L relative to $\hat{M}$, which leads the sender to favor L over M when her need for coordination is strong. As discussed, the latter occurs at the end of the game—see Figure 1a where β_2 has a steeper growth than β_1 close to the endgame. In other words, compared to the $\sigma_X = +\infty$ case, the history-inference effect is tapered off at the end of the game both by the direct effect of a public signal available and the strategic effect of the players coordinating in L (flatter β_1). The decreasing steering motive becomes dominant, and α_3 is decreasing at T, unlike in (ii) of Proposition 2.

If we interpret the signaling coefficient in the sender's action as the magnitude of "change" introduced by a leader in an organization during a pre-specified transition period, the robust prediction is as follows: change starts small; then, it gradually increases as leaders expect their organizations to understand their ultimate goals; but it eventually declines as the sender acquiesces to the organization's understanding due to the increasing importance of coordination as the transition period approaches its end.

Outcomes: a dichotomy between learning and performance Figure 1b illustrates that when the public signal is noisier, the total signaling coefficient increases for most of the game. This suggests that there can be more information transmission in a setting with higher-order uncertainty than if beliefs were public. To make our points, it suffices to compare the extreme cases of $\sigma_X = 0$ and $\sigma_X = \infty$ already introduced—we can think of them as limits of our general model. We do this when $r = 0$, where we obtain analytic solutions.

Proposition 3. *Suppose that $r = 0$. For all $T > 0$, the sender's ex ante payoff is larger if $\sigma_X = 0$ than if $\sigma_X = +\infty$, but the receiver's terminal belief is more precise if $\sigma_X = +\infty$.*

The first part of the result says that a leader is better off when she knows the follower's belief with certainty, which is natural given that this is a coordination game. The second part says that there is always more total information transmission to the sender when the leader is forced to rely on his actions exclusively. Figure 2 in fact confirms that these extreme values maximize the leader's ex ante payoff and the receiver's terminal precision.

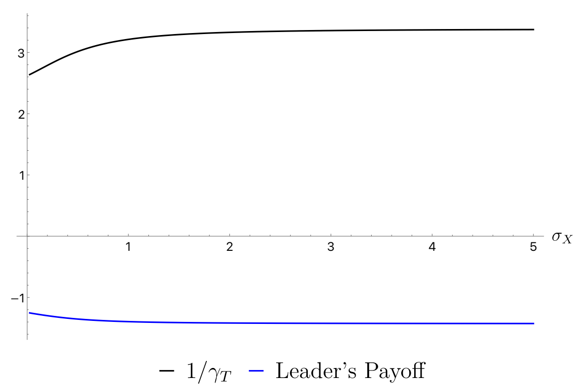


Figure 2: Leader's payoff (unnormalized) and follower's learning: $(T, \gamma^o, r, \sigma_Y) = (4, 1, 0, 1)$.

The leader's payoff is a natural measure of an organization's performance, while the precision of the receiver's terminal belief is a proxy for an organization's attained understanding during a relevant transition period. From this perspective, the proposition uncovers a potential dichotomy between learning and performance: organizations exhibiting a better understanding of the environment can in fact exhibit worse performance. Indeed, with information transmission through actions, learning is a measure of coordination costs.[24] For intuition, consider the public case (e.g., $\sigma_X = 0$): the leader could opt to take the follower's action in any period, thereby eliminating any miscoordination, but this implies that the leader neglects her private information, and hence that no information is transmitted. Our model predicts that the quality of the information fed back to organizations' leaders can be a key explanatory variable behind this tension between learning and performance.

[24]Formally, $\mathbb{E}_0 \int_0^T (a_t - \hat{a}_t)^2\, dt = \int_0^T \alpha_{3t}^2 \mathbb{E}_0(\theta - \hat{M}_t)^2\, dt = \int_0^T -\sigma_Y^2 \frac{\dot{\gamma}_t}{\gamma_t}\, dt = \sigma_Y^2 \ln(\frac{\gamma^o}{\gamma_T})$, which falls in $\gamma_T < \gamma^o$.

4.2 Application 2: Reputation for Neutrality

Recall the reputation game of Section 2: up to positive factors, the players' payoffs are:

$$\text{sender:} \; -\int_0^T e^{-rt}(a_t-\theta)^2 dt - e^{-rT}\psi \hat{a}_T^2,\; \psi>0; \qquad \text{receiver:} \; -(\hat{a}_t-\theta)^2.$$

The type θ captures a sender's form of bias (e.g., a politician's true stance on a topic). The mean μ is normalized to zero and is interpreted as an *unbiased type*. The receiver (e.g., a media outlet) wants to accurately predict the bias, i.e., $\hat{a}_t = \hat{M}_t$. Hence, the lump-sum terminal payoff in the sender's payoff captures a long-term concern for being perceived as unbiased: she wants $\hat{M}_T$ to match $\mu = 0$. Thus, all sender types have a *concealment* motive, but this conflicts with their short-run temptations ($-(a_t-\theta)^2$) term. Does better information, as measured by a more precise signal X, help the sender to manage her reputation?

The following result characterizes equilibrium behavior.

Proposition 4. *Suppose that $r \geq 0$ and $\sigma_X \in (0,\infty)$. In any LME, the sender's strategy satisfies $\beta_{0t}=0$ and $\beta_{1t}, \beta_{2t} \leq 0 < \beta_{3t} \leq 1$ for all $t\in[0,T]$, with all inequalities strict over $[0,T)$, while $\hat{a}_t = \hat{M}_t$ (i.e., $\delta_1 = 1$ and $\delta_0=\delta_2\equiv 0$). Moreover, $\alpha_{3t} := \beta_{3t}+\beta_{1t}\chi_t \in (0,1)$.*

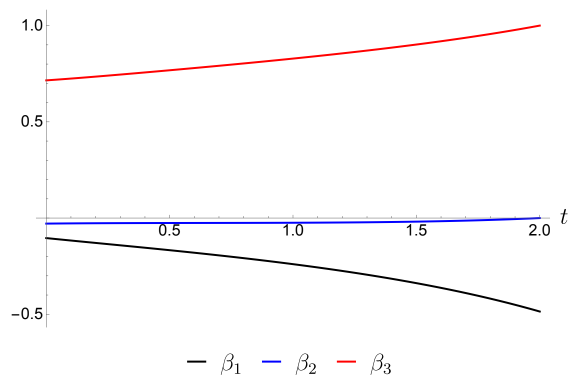


Figure 3: Sender's strategy coefficients in the reputation game: $(\psi, \gamma^o, r, \sigma_X, \sigma_Y) = (1, 2, .1, 1, 1)$.

Let us use Figure 3 to understand these coefficients. First, if the sender were myopic, she would attach a weight of 1 to θ at all times, precisely the terminal value of β_3. The sender then deviates from this value at earlier times in an effort to manage her reputation. Specifically, from a time-t perspective, her reputational concerns are captured by

$$-e^{-r(T-t)}\psi \mathbb{E}_t[\hat{M}_T^2] = -e^{-r(T-t)}\psi(\mathbb{E}_t[M_T^2]+\chi_T\gamma_T), \tag{18}$$

where $\chi_T\gamma_T$ is the variance of the sender's second-order belief (Lemma 2). Two conclusions immediately follow. First, since higher types take higher actions ($\alpha_3 > 0$) due to their higher biases, these types will anticipate greater upward drift in their reputation M all else equal. To preempt a large terminal loss, the sender moderates her actions, resulting in $\beta_{3t} < 1$ until time T; this deviation is stronger earlier in the game, as more time is left to reap the benefits of it. Second, senders with biased reputations M_t from their perspective expect to be perceived as biased at the end, so they will attempt corrective actions early on: the weight β_1 on M is negative so as to prevent this state from growing. And since M_t becomes a better predictor of M_T as time progresses, such corrections becomes stronger: β_{1t} is decreasing.

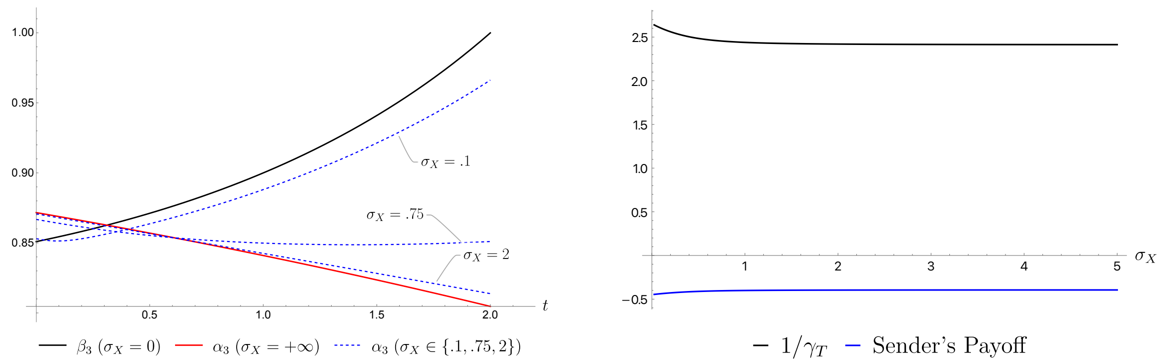
Finally, note that L is used in the strategies despite *not* appearing in the players' payoffs. This is because the receiver needs to anticipate the sender's action to update his belief after changes in Y, and this updating matters for the sender's intertemporal incentives. Indeed, since the receiver, via the representation, expects $\hat{\mathbb{E}}_t[a_t] = \alpha_{3t}\hat{M}_t + [\beta_{1t}(1-\chi_t) + \beta_{2t}]L_t$, the sender predicts $\hat{M}_t$ to update in the direction of $\mathbb{E}_t[dY_t - \hat{\mathbb{E}}_t[a_t]\,dt] = [a_t - (\alpha_{3t}M_t + [\beta_{1t}(1-\chi_t) + \beta_{2t}]L_t)]\,dt$, i.e., the drift of M in (13) (up to a constant). Because $\beta_1 < 0$, the sender's second-order belief has a tendency to drift up if $L > 0$, which induces the sender to correct his behavior today—$\beta_2 < 0$. The case $\theta = 0$, at an off-path history where $M_t = 0$, illustrates this issue: despite being unbiased and perceived as such, this type deviates from $a_t = 0$ because her reputation would otherwise deteriorate at rate $\beta_1(1-\chi_t)L_t$ given the receiver's identification problem. As this predictability ceases to matter at T, $\beta_{2T} = 0$ in Figure 3.

History-inference effect and concealment. As more extreme types take more extreme actions in equilibrium, such types will also develop more extreme beliefs about themselves; hence, those types will correct their reputations more aggressively than less extreme types. The history-inference effect now goes against information transmission: $\beta_1\chi$ and β_3 have opposite signs in α_3. But this creates scope for less separation, and hence a better chance to conceal the bias. A subtle trade-off emerges: with higher-order uncertainty, the sender loses her ability to take the best actions to manage her reputation, but she may transmit less information in the first place. Conversely, in a public environment, the sender can perfectly tailor her actions to her current reputation $\hat{M}$, but the history-inference effect is absent: while higher types do take higher actions in the analog LME, given any fixed public history all types agree on their reputation and hence use $\hat{M}$ to correct their actions in the same way.

To make our main point, we again examine the cases $\sigma_X = 0$ and $\sigma_X = +\infty$, which are particularly tractable for computing outcomes: the former being a public setting, and the latter maximizing the potential for the history-inference effect shutting down separation.

Proposition 5. *Suppose that $r = 0$ and $\psi < \sigma_Y^2/\gamma^o$. Then for all $T > 0$, there exists a unique LME when $\sigma_X = 0$ or $+\infty$. If $\sigma_X = +\infty$, the receiver's terminal belief is less precise, and the sender's ex ante payoff is higher, than when $\sigma_X = 0$.*

We conclude that a career-concerned agent can benefit from being uncertain about how she is perceived (e.g., because the information seen is imperfect). A sufficient condition for this to happen is $\psi < \sigma_Y^2/\gamma^o$: a very concave terminal payoff (ψ large) makes changes in the receiver's belief costly on average; and these changes are more drastic with large initial uncertainty (γ^o is large) or informative signals (σ_Y is low). Then, the "direct" cost of being unable to perfectly tailor actions to one's reputation dominates the "strategic" benefit of less information transmission. Figure 4 shows that the logic extends to $\sigma_X \in (0, +\infty)$ for fixed $r > 0$: as σ_X grows, α_3 rotates clockwise, falling for most of the game (left); this increases the sender's ex ante payoff and reduces the receiver's learning (right).



(a) Sender's signaling coefficient as σ_X varies.

(b) Learning and payoffs

Figure 4: Reputation game: $(\psi, \gamma^o, r, \sigma_Y) = (1, 1, .25, 1)$.

4.3 Application 3: Trading and Leakage

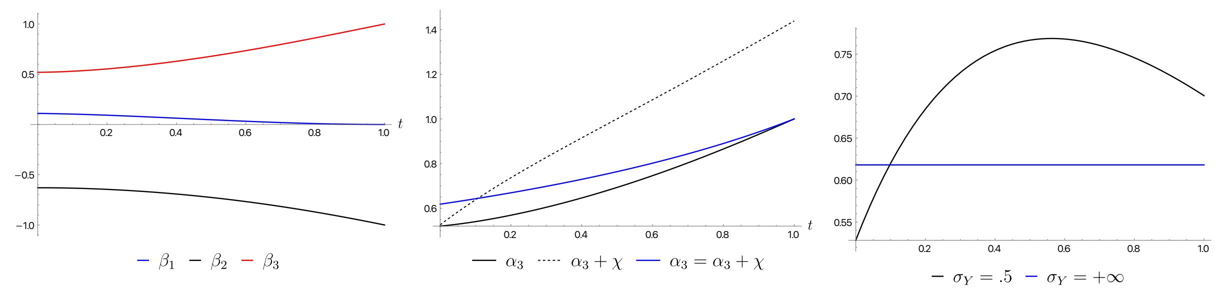
Recall the trading game from Section 2, where the players' payoffs are given by

$$\text{sender}: \int_0^T \left[a_t(\theta - \mathbb{E}[\theta|\mathcal{F}_t^X]) - \frac{a_t^2}{2} \right] dt; \quad \text{receiver}: \hat{a}_t(\theta - \mathbb{E}[\theta|\mathcal{F}_t^X]) - \frac{\hat{a}_t^2}{2}.$$

The sender is a trader who knows the value θ of an asset, while the receiver is an ex-ante uninformed trader who sees a noisy private signal ("leakage") of the sender's behavior. The asset's (public) price is $\mathbb{E}[\theta|\mathcal{F}_t^X]$, based on the the total order flow $dX_t = (a_t + \hat{a}_t)dt + \sigma_X dZ_t^X$, carrying the players' trades a_t and $\hat{a}_t$ as well as confounding "noise trading." In models of this kind (e.g., Kyle, 1985) a key question is how the sender dynamically exploits mispricing $\theta - \mathbb{E}[\theta|\mathcal{F}_t^X]$ accounting for how her trades affect future prices. Here, we will explore how these incentives change when those same trades generate private information for others.

Our analysis from Sections 2–3 can be extended to this case where both players' actions affect the public signal. In particular, the identity $L_t = \mathbb{E}[\theta|\mathcal{F}_t^X]$ still holds in the analogous representation result for this case, so we can explore environments where $\mathbb{E}[\theta|\mathcal{F}_t^X]$ is directly payoff-relevant. The following proposition characterizes equilibrium coefficients. (Section S.4 in the Supplementary Appendix provides an existence result for this game.)

Proposition 6. *Suppose $\sigma_X \in (0, \infty)$. In any LME, $\beta_{0t} = 0$ and $\beta_{1t} + \beta_{2t} + \beta_{3t} = 0$. Thus, in equilibrium, $a_t = \alpha_{3t}(\theta - L_t)$ with $\alpha_{3t} > 0$; $\hat{a}_t = \hat{M}_t - L_t$; and the price satisfies $dL_t = \Lambda_t \, dX_t$ where $\Lambda_t := \frac{\gamma_t^X(\alpha_{3t}+\chi_t)}{\sigma_X^2}$. Moreover, if T is not too large, the coefficients can be signed analytically: $\beta_{1t} > 0$, $\beta_{3t} \in (0,1)$, and $\beta_{2t} < 0$; while $\frac{d\alpha_{3t}}{dt} > 0$, $\frac{d\beta_{3t}}{dt} > 0$, and $\frac{d\beta_{1t}}{dt} < 0$; and $\beta_{1t}\chi_t$ is nonmonotone and is maximized at an interior time.*



(a) Sender's strategy coefficients ($\sigma_Y = .5$)

(b) α_3 and $\alpha_3 + \chi$ for $\sigma_Y = .5$ (black) and $\sigma_Y = +\infty$ (blue)

(c) Price impact as σ_Y varies.

Figure 5: Trading game: $(\gamma^o, r, \sigma_X) = (1, 0, 1)$.

As higher sender types expect a larger profit per unit traded, the equilibrium weight on the type β_3 is positive; and obviously, purchases must fall as the price increases, and so $\beta_2 < 0$. These signs are consistent with the proposition and are confirmed more generally in Figure 5a. Further, as is ubiquitous in this literature, the sender's equilibrium trades are based on the size of the current mispricing via $\alpha_3(\theta - L_t)$, with $\alpha_3 > 0$ also shaping the responsiveness of prices via Λ. In other words, endogenous price impact makes it costly for the sender to place large trades. As the endgame approaches and the concern about future prices fades, however, β_3 and β_2 move toward the myopic solution monotonically.

Our model departs from the literature in the form that price impact takes, due to the private monitoring at play. To illustrate, suppose that the sender deviated by trading more aggressively in the past, resulting in a higher M than that implied by the representation (7). In practice, this means that the sender thinks that the receiver is optimistic about the asset and, critically, that this optimism will eventually get incorporated into prices: through the lens of our baseline model, M—a persistent state—enters the law of motion (14) for the price L. In other words, from the sender's perspective, an extra layer of price impact emerges.

From an incentives viewpoint, this extra layer is a form of price predictability that can be exploited by the sender. Indeed, using that $\beta_{1t} + \beta_{2t} + \beta_{3t} = 0$, her extended strategy reads $a_t = \beta_{3t}(\theta - L_t) + \beta_{1t}(M_t - L_t)$, thus capturing that the sender exploits both forms of superior information relative to market makers (who believe $\mathbb{E}[\hat{M}_t|\mathcal{F}_t^X] = L_t$); the fact that $\beta_1 > 0$ reflects an incentive to buy more aggressively in anticipation of higher future prices, an effect that decays over time ($\beta_{1T} = 0$)—see Figure 5a. This phenomenon is similar to that in the reputation game, albeit in reversed form, where the sender used L to predict M.

The resulting extra persistence shapes equilibrium price impact: Λ in the proposition features β_3 augmented by $\beta_1\chi + \delta_1\chi = \beta_1\chi + \chi$, the sender's history-inference effect plus the receiver's own trades. The appearance of χ in the latter stems from θ and $\hat{M}$ becoming more correlated over time from the perspective of "market makers," which is a form of history-inference effect linked to the receiver's signaling: indeed, by not observing Y, such price setters must construct a (second-order) belief about $\hat{M}$ using the sender's conjectured equilibrium play.[25] The sender then scales back along the β_3 dimension for fear of high future prices: in Figure 5b, α_3 for $\sigma_Y < +\infty$ falls below the "no-leak" benchmark case $\sigma_Y = +\infty$.

The total signaling coefficient $\alpha_3 + \chi$ from the perspective of market makers is shown in dashed in Figure 5b: it is low early on due to the sender's reduced signaling and the total history-inference effect just getting started, but it increases as the latter effect builds. As a result, in panel 5c, price impact begins below the no-leak benchmark, but it eventually surpasses it (falling at the end due to the market maker's learning). For the most part, then, high price impact is consistent with a low degree of insider trading, as we formalize next.

Proposition 7. *Fix $\sigma_Y \in (0, +\infty)$, and suppose that a LME exists over $[0, T]$. For any such LME, there exists a nonzero measure of times t for which $\Lambda_t^{leak} > \Lambda_t^{no\ leak}$.*

5 Existence of Linear Markov Equilibria

In this section we show that the problem of finding LMEs is effectively one of solving a system of ODEs carrying a mix of initial and terminal conditions—a "boundary value problem" (BVP). We provide time horizons for which such a problem admits a solution.

Setting up a BVP We postulate a quadratic value function for the sender of the form

$$V(\theta, m, \ell, t) = v_{0t} + v_{1t}\theta + v_{2t}m + v_{3t}\ell + v_{4t}\theta^2 + v_{5t}m^2 + v_{6t}\ell^2 + v_{7t}\theta m + v_{8t}\theta\ell + v_{9t}m\ell,$$

where $v_{i\cdot}$, $i = 0, ..., 9$ are differentiable functions of time. We can then write the Hamilton-Jacobi-Bellman (HJB) equation for the sender's problem: for all $t < T$,

$$rV = \sup_{a'} \left\{ \tilde{u}_t(a', \mathbb{E}_t[\hat{a}_t], \theta) + V_t + \mu_M(a')V_m + \mu_L V_\ell + \frac{\sigma_M^2}{2}V_{mm} + \sigma_M\sigma_L V_{m\ell} + \frac{\sigma_L^2}{2}V_{\ell\ell} \right\}, \quad (19)$$

[25] Indeed, by the representation, $\mathrm{Cov}(\hat{M}_t, \theta)$ conditional on the public information takes the value $\chi_t\gamma_t^X$.

where $\tilde{u}_t(\cdot) := u(\cdot) + \frac{1}{2}u_{\hat{a}\hat{a}}\delta_{1t}^2\gamma_t\chi_t$, $\mu_M(a')$ and μ_L (respectively, σ_M and σ_L) denote the drifts (respectively, volatilities) in (13) and (14), and $\hat{a}_t$ is determined via the static best response (16). Recall that implicit in this problem is a tuple $(\beta_0, \beta_1, \beta_2, \beta_3)$ used by the receiver to construct his best response and form beliefs about the sender; this tuple thus affects the sender's flow utility u and the drift and volatility terms in both M and L.

Let $a(\theta, m, \ell, t)$ denote the maximizer of the right-hand side in the HJB equation. It is easy to see that the first-order condition (FOC) reads

$$\frac{\partial u}{\partial a}(a(\theta, m, \ell, t), \delta_{0t} + \delta_{1t}m + \delta_{2t}\ell, \theta) + \underbrace{\frac{\gamma_t\alpha_{3t}}{\sigma_Y^2}}_{dM_t/da_t}\underbrace{[v_{2t} + 2v_{5t}m + v_{7t}\theta + v_{9t}\ell]}_{V_m(\theta,m,\ell,t)} = 0, \tag{20}$$

which is a linear equation in $a(\theta, m, \ell, t)$ and (θ, m, ℓ). One can then solve for $a(\theta, m, \ell, t)$ in (20) and impose the equilibrium condition $a(\theta, m, \ell, t) = \beta_{0t} + \beta_{1t}m + \beta_{2t}\ell + \beta_{3t}\theta$. Since the resulting equation must hold at all possible values of (θ, m, ℓ), the equilibrium condition boils down to equating the terms in each of the variables in (θ, m, ℓ) and the constant terms. This procedure links strategy coefficients $(\beta_0, \beta_1, \beta_2, \beta_3)$ to (v_2, v_5, v_7, v_9) in the value function.

At this stage, one can always obtain a system of ODEs for v_i, $i = 0, ..., 9$ after returning to the HJB equation. The structure of the problem, however, permits a reduction. Concretely:

1. We can solve for (v_2, v_5, v_7, v_9) directly in terms of $\vec{\beta}$ and (γ, χ) (see (C.1)-(C.4) in the Appendix); the associated mapping is well-defined provided that α_3 and γ never vanish, which will be the case in the equilibrium we construct (more in this shortly);

2. We can then insert the expressions from the previous step into the HJB equation, along with $a(\theta, m, \ell, t) = \beta_{0t} + \beta_{1t}m + \beta_{2t}\ell + \beta_{3t}\theta$, to obtain a system of ODEs for both the coefficients $(\beta_0, \beta_1, \beta_2, \beta_3)$ and the remaining coefficients in the value function. These ODEs are coupled with those of (γ, χ) because the learning coefficients affect the law of motion of (M, L). The resulting system of ODEs can be further reduced by eliminating $(v_0, v_1, v_3, v_4, \beta_0)$ which are "downstream" of the remaining variables.[26]

This procedure yields a system of ODEs for $(\beta_1, \beta_2, \beta_3, v_6, v_8, \gamma, \chi)$, which can be found in Appendix C.[27] We need to complement this system with *boundary conditions*. First, γ and χ satisfy exogenous initial conditions $\gamma_0 = \gamma^o > 0$ and $\chi_0 = 0$ reflecting the players' initial

[26]Note that (v_0, v_1, v_4) are the coefficients of the constant, θ- and θ^2-terms in the leader's value function, none of which the leader controls, so they have no impact on the rest of the system. Meanwhile, the equations for (β_0, v_3) are coupled as these encode the *deterministic* component of the leader's incentive to manipulate beliefs, which by definition is independent of the values that the beliefs take.

[27]We use a change of variables there to simplify the ODEs. The fact that v_6 and v_8 cannot be eliminated from the system is a consequence of the sender indirectly controlling L via changes in M (see (14)).

uncertainty. Second, there are *endogenous* terminal values for the remaining variables that are determined by the static (Bayes) Nash equilibrium played at time T. For expositional simplicity, we provide these for the case in which there are no terminal payoffs, i.e., $\psi \equiv 0$:

$$\beta_{0T} = \frac{u_0 + u_{a\hat{a}}\hat{u}_0}{1 - u_{a\hat{a}}\hat{u}_{\hat{a}a}}, \ \beta_{1T} = \frac{u_{a\hat{a}}[u_{a\theta}\hat{u}_{\hat{a}a} + \hat{u}_{\hat{a}\theta}]}{1 - u_{a\hat{a}}\hat{u}_{\hat{a}a}\chi_T}, \ \beta_{2T} = \frac{u_{a\hat{a}}^2\hat{u}_{\hat{a}a}[u_{a\theta}\hat{u}_{\hat{a}a} + \hat{u}_{\hat{a}\theta}](1-\chi_T)}{(1 - u_{a\hat{a}}\hat{u}_{\hat{a}a})(1 - u_{a\hat{a}}\hat{u}_{\hat{a}a}\chi_T)}, \qquad (21)$$
$$\beta_{3T} = u_{a\theta}, \qquad v_{6T} = v_{8T} = 0.$$

This fully specifies a BVP that the coefficients $(\beta_1, \beta_2, \beta_3, v_6, v_8)$, along with (γ, χ), must satisfy in a LME. (The general expressions for the terminal conditions in the presence of a terminal payoff are presented in Section S.3.2 in the Supplementary Appendix.)

Technical conditions on primitives and existence technique Since the players engage in a static game of two-sided incomplete information at time T, a minimal requirement is that this static game always admits an equilibrium—in the terminal conditions (21), this amounts to all the denominators being different from zero *after all possible histories* of the game, which are encoded in the value that χ_T takes at the endgame. For intuition, notice that absent any incomplete information, the sender's *best-response function* is linear in $\hat{a}$ with slope $u_{a\hat{a}}$, while the receiver's counterpart has slope $\hat{u}_{\hat{a}a}$ on a (due to $u_{aa} = \hat{u}_{\hat{a}\hat{a}} = 1$); thus, the players' best responses would (generically) never intersect if $u_{a\hat{a}}\hat{u}_{\hat{a}a} = 1$. Since in our setting we require that both $1 - u_{a\hat{a}}\hat{u}_{\hat{a}a}$ and $1 - u_{a\hat{a}}\hat{u}_{\hat{a}a}\chi_T$ never vanish, and χ_T takes values in $[0, 1)$ (Lemma 1), the requirement that the best response functions always intersect boils down to $1 - u_{a\hat{a}}\hat{u}_{\hat{a}a}\chi_T$ never changing sign, and so we require that $u_{a\hat{a}}\hat{u}_{\hat{a}a} < 1$.

Second, we want the sender to signal her type at all times. While this requirement can be perceived as minimal too, we note also that it guarantees that there is information transmission via the second-order belief channel: if $\alpha_3 := \beta_1\chi + \beta_3$ never vanishes, in equilibrium different types do take different actions after each history of the public signal, as the sender's action takes the form $\alpha_0 + \alpha_2 L + \alpha_3\theta$ along the path of play. As it turns out, to guarantee that the coefficient α_3 never changes sign it suffices to ensure that its terminal value α_{3T} never vanishes. To compute the latter, we use (21) again to obtain:

$$\alpha_{3T} = \beta_{1T}\chi_T + \beta_{3T} = \frac{u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_T}{1 - u_{a\hat{a}}\hat{u}_{\hat{a}a}\chi_T}.$$

Since $\chi_T \in [0, 1)$, the numerator is guaranteed to never vanish when $u_{a\theta}$ and $u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}$ have the same sign.[28] Our technical conditions then read as follows:

[28]If $\alpha_3 < 0$ in equilibrium, higher types take lower actions (e.g., more negative), yet naturally develop higher second-order beliefs. This is because the weights that M attaches to past actions are negative due to the receiver responding negatively to large realizations of Y. Also, note that we do allow the receiver to

Assumption 2. *Flow payoffs satisfy (i)* $u_{a\hat{a}}\hat{u}_{a\hat{a}} < 1$ *and (ii)* $u_{a\theta}(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}) > 0$.

Establishing the existence of a solution to the BVP is nontrivial not only because solutions to the ODEs must exist over the whole time horizon, but also because these solutions must land at potentially endogenous values. This problem is particularly challenging due to the presence of multiple ODEs in both directions: the "behavior" ODEs for $(\vec{\beta}, v_6, v_8)$ are traced backward from their terminal values by backward induction, while the "learning" ODEs for (γ, χ) are traced forward from their initial values. In BVPs where only one variable has an initial condition and the remaining variables have terminal conditions (or vice-versa), a traditional one-dimensional shooting argument applies: introduce a guess variable for the candidate terminal value of the solution to the ODE going forward, trace all variables backward in time using that guess variable as the initial condition, and argue via the intermediate value theorem that some guess hits the target (the exogenous initial condition). With multiple ODEs in both directions, this method does not apply.[29]

The problem of existence of LME, however, is fundamentally a *fixed-point problem of functions*: the evolution of the learning coefficients (γ, χ) depends on the signaling that takes place during the game, but the signaling coefficients depend on the path of the learning coefficients because these are taken as given by the sender in the best-response problem. Thus, we translate the BVP into a fixed-point equation in the space of functions (γ, χ). That is, our fixed-point argument is infinite-dimensional. It works as follows.

First, we choose an arbitrary pair $\lambda := (\gamma, \chi)$ in a closed-convex domain Λ that nests all functions (γ, χ) that can be obtained as solutions to their coupled ODEs (10)–(11) for continuous (β_1, β_3) satisfying a particular uniform bound. Taking λ as an input, we "shoot back": we pose an *initial value problem* in time-reversed form consisting of the ODEs for $(\vec{\beta}, v_6, v_8)$ taking λ as an input, and where initial conditions for the ODEs are given by the static time-T conditions of the game (which may depend on λ_T). We then derive a sufficient condition on the time horizon such that: (i) this initial value problem has a unique solution for all λ in the domain; (ii) the solution satisfies the uniform bound referred to above (we expand on this after the theorem); and (iii) the solution is continuous in λ. We then "shoot forward": we feed the resulting (β_1, β_3) pair into the learning ODEs for (γ, χ) to get a solution for this system that we denote $\boldsymbol{\lambda}$. As we prove, the mapping from input pairs λ to $\boldsymbol{\lambda}$ is continuous, and $\boldsymbol{\lambda}$ lies in Λ, making Schauder's infinite dimensional fixed-point theorem applicable. By construction, the fixed-point coefficients found induce an LME. Figure 6

suspend information transmission, which happens when $\delta_1 = \hat{u}_{\hat{a}\theta} + \hat{u}_{\hat{a}a}[\beta_{3t} + \beta_{1t}\chi_t]$ vanishes. Indeed, this is not an issue: the sender simply ignores the public signal, while the second-order belief continues updating through the use of the sender's past history of play. That said, if this occurs, it is only temporary due to (β_1, β_3) themselves changing over time—and it can never happen if exactly one of $(\hat{u}_{\hat{a}\theta}, \hat{u}_{\hat{a}a})$ is zero.

[29] Special cases for which the one-dimensional shooting is applicable are discussed in Section 6.

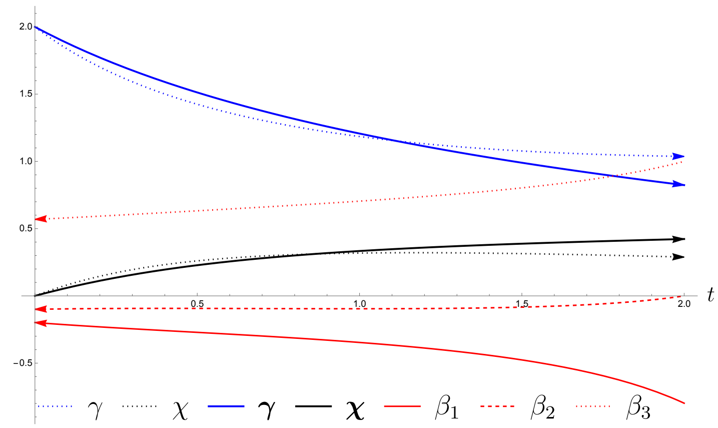


Figure 6: One iteration under our fixed-point method for the reputation game. Take as given candidate learning coefficients (dotted curves pointing to the right). Next, working backwards, generate candidate equilibrium coefficients (curves pointing to the left). Finally, use the latter functions to generate solutions to the learning ODEs (solid curves pointing to the right).

illustrates one iteration of this procedure.

We can now state our main theorem, which guarantees existence of LME for time horizons that are *robust to the discount rate, for the entire class of games.* Recall that ψ captures the sender's terminal payoff function depending on the receiver's terminal action.

Theorem 1. *Suppose Assumptions 1 and 2 hold. If either ψ is linear (including $\psi \equiv 0$), or ψ is not too concave, there exists a scalar $C > 0$ independent of (r, γ^o), such that, if $T < C/\gamma^o$, there exists an LME for all $r \geq 0$. In this equilibrium, α_3 never vanishes.*

To understand why the horizons found are proportional to $1/\gamma^o$, recall that beliefs are naturally less responsive to new information as the initial uncertainty, γ^o, falls, which means there is less scope for manipulating beliefs by the sender. Mathematically, the ODEs for the equilibrium coefficients are proportional to γ, and so smaller values of γ^o lower the pressure on the uniform bounds that we find for $(\vec{\beta}, v_6, v_8)$; less stringent bounds, in turn, enable us to find longer horizons over which solutions to the behavior ODEs can exist. On the other hand, the presence of a terminal payoff can make the static Nash equilibrium arising at T more complex due to "last minute" incentives; a purely technical lower bound on the second derivative of ψ then allows us to extract a sufficiently regular selection of static equilibria for all possible (χ_T, γ_T) over $[0, 1] \times [0, \gamma^o]$, which we need for continuity purposes in our argument. Importantly, this lower bound depends on parameters, and sometimes it never binds (i.e., ψ'' can take any value in $(-\infty, 0)$), as is the case for our reputation game.

This is obviously a very general result. Taking a step back, however, it is natural to ask why one needs this infinite-dimensional approach. The reason, interestingly, lies purely in the economics of the problem. First, it is only when the ODEs for the behavior coefficients are traced backward that greater discounting limits their growth; we exploit this to find times for existence that apply for all $r \geq 0$. Second, the learning ODEs always admit solutions if traced forward, but not necessarily backwards from generic values. Thus, the approach fully exploits the natural tractability of the system in each direction: behavior determined in a backward fashion via backward induction and Bayesian updating naturally evolving forward. But to leverage this structure, only a subset of the ODEs can be used in each "shooting" step, which means that candidate solutions for the remaining ODEs are needed as inputs. The need for input *functions* means that the approach must be infinite dimensional.

Finally, all the steps in this existence technique can be refined: we can include more general terminal payoffs, obtain better uniform bounds (we only use the degree of the polynomials involved), and potentially find horizons of existence that increase with the discount rate; the latter because behavior must be closer to myopic as r increases, and an LME for myopic players exists for all T by Assumption 2.[30] In the next section we discuss this method in light of the existing literature, as well as areas for future applicability.

6 Discussion

Forward-looking receiver A forward-looking receiver would actively control L via her actions affecting X, but no states beyond $(t, \hat{M}, L)$ would be necessary for this player. Also, the same strategies found in Applications 1 and 2 would remain an LME. To see why, consider the coordination game, and suppose that the receiver deviates from choosing the myopic best response $\hat{\mathbb{E}}_t[a_t]$ over $[t, t+dt)$, thus incurring in a loss over that instant. Importantly, because the deviation is hidden, the receiver continues thinking that the sender takes actions according to $\alpha_3\theta + (1-\alpha_3)L$. Since α_3 is deterministic, however, only L is affected by the deviation, but the latter state is always perfectly observed anyways. In other words, the receiver cannot affect the informativeness of Y (α_3 is unaffected), and hence cannot affect his speed of learning. This means that the deviation does not improve the receiver's ability to predict a_t at future times, so there is no future benefit associated with the deviation. The same logic applies to the reputation game, and more generally to *prediction* problems.

Proposition 8. *Suppose that $\hat{u}(a, \hat{a}, \theta) = -\frac{1}{2}(c_0 + c_1\theta + c_2 a - \hat{a})^2$, with $c_0, c_1, c_2 \in \mathbb{R}$, and that an LME in our baseline model exists. Then, for all $\hat{r} \geq 0$, the same LME arises when the receiver has the payoff $\int_0^T e^{-\hat{r}t}\hat{u}(a_t, \hat{a}_t, \theta)\, dt + e^{-\hat{r}T}\hat{u}(a_T, \hat{a}_T, \theta)^2$.*

[30] To find horizons that apply for all $r \geq 0$, we perform two modifications to the BVP before constructing a fixed point. For expositional ease, we defer a detailed explanation of those modifications and the underlying motivation to the proof in Appendix C (see '**Centering**' and '**Auxiliary Variable**' steps).

Beyond these settings, non-trivial dynamic incentives for the receiver can arise. First, non-strategic intertemporal "smoothing" motives: e.g., if the terminal payoff were different from the flow counterpart, a forward-looking receiver would depart more from the static best-response as the endgame approaches, while the myopic policy would exhibit a discontinuity at T. Second, more interestingly, there can be strategic incentives to manipulate the sender's belief. Since these incentives operate through affecting a public signal, however, they correspond to traditional signal-jamming motives (e.g., Holmström, 1999). Importantly, our methods can be adapted to find LME in these cases: no additional learning ODEs would arise, but the "backward part" of the BVP would have to be augmented to account for ODEs that characterize the (now) dynamic coefficients $(\delta_0, \delta_1, \delta_2)$ in the receiver's strategy. Our two-step shooting method would then have to be applied to this new system.[31]

Private-value environments and one-dimensional shooting The presence of two learning dynamics γ and χ is at the core of the complications behind our BVP. Economically, this results from the players potentially signaling at very different rates, so it is natural to examine environments with some symmetry. With *private values*, i.e., $\hat{u}_{\hat{a}\theta} = 0$, the receiver strategically cares about the sender's action only, and so the players signal at proportional rates ($\delta_1 = \hat{u}_{\hat{a}a}\alpha_3$). In those settings, a one-to-one mapping between γ and χ exists.

Proposition 9. *If $\hat{u}_{\hat{a}\theta} = 0$, $\chi_t = \frac{c_1 c_2(1-[\gamma_t/\gamma^o]^d)}{c_1+c_2[\gamma_t/\gamma^o]^d}$ for some positive scalars c_1, c_2 and d. Thus, $\chi_t \in [0, c_2)$ when $\gamma_t \in (0, \gamma^o]$, where c_2 is increasing in σ_X with $\lim_{\sigma_X \to 0} c_2 = 0$ and $\lim_{\sigma_X \to +\infty} c_2 = 1$.*

With private values, the shooting problem is one-dimensional and traditional continuity arguments apply—see Bonatti et al. (2017). Three observations are instructive. First, the upper bound c_2 for χ confirms our intuition that less weight is given to the type when the public signal improves. Second, while the multidimensional case is both conceptually and technically more challenging, the general horizons for which we can guarantee the existence of LME in Theorem 1 are of the same order as in the one-dimensional case. The reason is that the horizons found are pinned down, in both settings, by uniformly bounding the ODEs associated with the behavior coefficients exclusively (i.e., the dependence of the learning ODEs is only implicit); thus, our infinite-dimensional method establishes itself as the "right" extension of the one-dimensional shooting case. Third, Proposition 9 is a contribution in itself: analog results for first-order private beliefs had been derived in settings where types come from symmetric distributions (e.g., Foster and Viswanathan, 1996); instead, the sender's type is fixed and exogenous here, while the receiver's type is evolving and its dis-

[31]See `spm.nb` on our websites. There, (as discussed) the system is stated in terms of the value function coefficients (rather than the strategy counterparts) because that domain is more convenient for the receiver's problem. (This stems from the fact that $\chi = 0$ at $t = 0$, and thus the receiver cannot initially affect L.)

tribution determined in equilibrium. More generally, by extending to a second-order belief and involving a system of ODEs, our representation is a novel result in the literature.

Further applications of the existence technique Our fixed-point technique is useful in BVPs featuring multiple ODEs in both directions, and where a non-trivial feedback between the forward and backward components is at play. LQG games of incomplete information are a natural area of exploitation—forward ODEs encoding learning, with backward counterparts arising from quadratic value functions—and there are two sub-areas in which our methods immediately apply. First, games of one-sided noisy signaling involving multidimensional types: there, the receiver would in general need to keep track of a nontrivial *variance-covariance matrix* when updating from a linear strategy in the multidimensional type of an informed player. Second, in games with multi-sided private information and noisy public signals, where types have different prior variances: any player would again have to construct a variance-covariance matrix to form beliefs about rivals, as different precisions of prior beliefs are a natural trigger for different signaling rates.[32] In both cases, such a matrix evolves deterministically; but in the second, beliefs can be private.

That said, our methods can be applied to other economic settings, the constraint being that equilibrium variables must be encoded in a system of ODEs. One potential area of study are models of search and bargaining in continuous time, akin to Duffie et al. (2005), who study over-the-counter markets. There, different agents' (investors, dealers) willingness to pay for an asset satisfy ODEs stemming from Bellman equations. Such "behavior" ODEs depend on the masses of agents looking to buy/sell an asset, because the number of agents determine the contact rates—and hence, changes in utility—when matching is random; further, these masses can also obey deterministic dynamics. With stationary solutions as the typical object of study, the ODEs becomes algebraic equations and the distinction between forward and backward ODEs is absent. Out of steady state, however, this need not be the case: if a crisis hits and, say, some dealers are not willing to intermediate, the initial size of the dealer segment will matter for recovery. Further, if subsequent entry is allowed, this decision will naturally depend on future market profitability—utilities now enter the ODEs that govern the evolution of the aforementioned masses, and a full feedback is at play.[33] Finite horizon versions of settings like these can be used to approximate an infinite horizon market.[34]

[32]See Cetemen (2020), who uses a finite-dimensional fixed-point method from an earlier version of our paper, suited for undiscounted games. Multiple learning variables also arise in Foster and Viswanathan (1994), where types can be multi-dimensional and asymmetric; their fixed-point problem is confronted numerically.

[33]Our fixed point arises from behavior depending on past learning, which depends on past behavior, which depends on future learning/behavior via backward induction. This temporal circularity arises at the filtering stage in LQG macroeconomic models with forward-looking variables (Svensson and Woodford, 2003).

[34]Bonatti et al. (2017) use this "sequence" approach to show the existence of an LME in an infinite-horizon version of their model of dynamic oligopoly with incomplete information.

Private-public information structure Our mixed private-public information structure is crucial for closing the state space at the level of the sender's second-order belief. If instead the receiver's actions were privately monitored, the receiver would have to resort to his past history of play to forecast the sender's "M" (as opposed to using our representation). The problem is that the resulting linear aggregate of past actions can become a non-trivial state variable: because past actions carry the receiver's past beliefs, and beliefs change over time, the linear aggregate need not coincide with the receiver's contemporaneous first-order belief. This means that the sender has to form a second type of second-order belief—one about a historical average of past values of the receiver's first-order belief—which the receiver would have to forecast again relying on his past play, and so forth. So, not only would the players have to move up in the belief hierarchy (e.g., the receiver constructing a non-trivial third-order belief), but in their forecasting exercises the players also would move "horizontally," constructing different types of first- and second-order beliefs. Whether this information structure is manageable—and, equally important, how relevant for behavior and outcomes is this movement up and across the belief hierarchy—is an open question.

7 Concluding Remarks

We have developed a dynamic model of strategic behavior that is at the intersection of two long-standing areas in game theory: signaling and private monitoring games. With respect to signaling games, we have uncovered a new higher-order belief channel for separation. Importantly, this channel rests on an intuitive logic: when agents need to rely on their past actions to forecast what others have seen, different types necessarily develop different beliefs—despite this clearly being the generic case (rarely is all information public), this area has largely been unexplored. On the other hand, with respect to private monitoring games, our setup demonstrates that examining asymmetric environments—here, a combination of one-sided incomplete information with one-sided private monitoring—can be a fruitful endeavor, especially if the focus is on studying equilibria that are belief-based.

The tractability of a linear-quadratic-Gaussian structure has been key in this regard. While LQG models have been exploited in many static settings, it is far less obvious what to expect in dynamic environments featuring complex information structures like ours. This paper demonstrates that, despite the substantial gap in difficulty when transitioning to the latter world, it is still possible to obtain new answers and insights, while at the same time contributing methodological tools that can be implemented in other domains. It is our belief that the stylized nature of these games, rather than being a limitation, is an asset that helps uncover forces that are robust to other, more nonlinear settings.

Appendix A: Proofs for Section 3

Preliminary results. We state standard results on ODEs (Teschl, 2012) which we use in the proofs that follow. Let $f(t,x)$ be continuous from $[0,T]\times\mathbb{R}^n$ to $\mathbb{R}^n$, where $T>0$.

- *Peano's* Theorem (Theorem 2.19, p. 56): There exists $T'\in(0,T)$, such that there is at least one solution to the IVP $\dot{x}=f(t,x),\ x(0)=x_0$ over $t\in[0,T')$.

If, moreover, f is locally Lipschitz continuous in x, uniformly in t, then:

- *Picard-Lindelöf* Theorem (Theorem 2.2, p. 38): For $(t_0,x_0)\in[0,T)\times\mathbb{R}^n$, there is an open interval I over which the IVP $\dot{x}=f(t,x),\ x(t_0)=x_0$ admits a unique solution.

- *Comparison* theorem (Theorem 1.3, p. 27): If $x(\cdot),y(\cdot)$ are differentiable, $x(t_0)\le y(t_0)$ for some $t_0\in[0,T)$, and $\dot{x}_t-f(t,x(t))\le\dot{y}_t-f(t,y(t))$ $\forall t\in[t_0,T)$, then $x(t)\le y(t)$ $\forall t\in[t_0,T)$. If, moreover, $x(t)<y(t)$ for some $t\in[t_0,T)$, then $x(s)<y(s)$ $\forall s\in[t,T)$.

Proof of Lemma 1. Let L in (7) denote a process that is measurable with respect to X. Inserting (7) into (5) yields $a_t=\alpha_{0t}+\alpha_{2t}L_t+\alpha_{3t}\theta$ which the receiver thinks drives Y, where $\alpha_{0t}=\beta_{0t}$, $\alpha_{2t}=\beta_{2t}+\beta_{1t}(1-\chi_t)$, and $\alpha_{3t}=\beta_{3t}+\beta_{1t}\chi_t$.

The receiver's filtering problem is then conditionally Gaussian. Specifically, define

$$d\hat{Y}_t \;:=\; dY_t-[\alpha_{0t}+\alpha_{2t}L_t]dt=\alpha_{3t}\theta dt+\sigma_Y dZ_t^Y,$$

which are in the receiver's information set, and where the last equalities hold from his perspective. By Theorems 12.6 and 12.7 in Liptser and Shiryaev (1977), his posterior belief is Gaussian with mean $\hat{M}_t$ and variance γ_{1t} (simply γ_t in the main body) that evolve as

$$d\hat{M}_t=\frac{\alpha_{3t}\gamma_{1t}}{\sigma_Y^2}[d\hat{Y}_t-\alpha_{3t}\hat{M}_t dt] \;\text{ and }\; \dot{\gamma_{1t}}=-\frac{\gamma_{1t}^2\alpha_{3t}^2}{\sigma_Y^2}. \tag{A.1}$$

(These expressions still hold after deviations, which go undetected.)

The sender can affect $\hat{M}_t$ via her choice of actions. Indeed, using that $d\hat{Y}_t=(a_t-\alpha_{0t}-\alpha_{2t}L_t)dt+\sigma_Y dZ_t^Y$ from her standpoint,

$$d\hat{M}_t \;=\; (\kappa_{0t}+\kappa_{1t}a_t+\kappa_{2t}\hat{M}_t)dt+B_t^Y dZ_t^Y,\;\text{ where} \tag{A.2}$$

$$\kappa_{1t}=\alpha_{3t}\gamma_{1t}/\sigma_Y^2,\quad \kappa_{0t}\;=-\kappa_{1t}[\alpha_{0t}+\alpha_{2t}L_t],\ \kappa_{2t}=-\alpha_{3t}\kappa_{1t},\ B_t^Y=\alpha_{3t}\gamma_{1t}/\sigma_Y. \tag{A.3}$$

On the other hand, since the sender always thinks that the receiver is on path, the public signal evolves, from her perspective, as $dX_t=(\delta_{0t}+\delta_{1t}\hat{M}_t dt+\delta_{2t}L_t)dt+\sigma_X dZ_t^X$. Because the dynamics of $\hat{M}$ and X have drifts that are affine in $\hat{M}$—with intercepts and slopes that are in the sender's information set—and deterministic volatilities, the pair $(\hat{M},X)$ is conditionally

Gaussian. Thus, by the filtering equations in Theorem 12.7 in Liptser and Shiryaev (1977), $M_t := \mathbb{E}_t[\hat{M}_t]$ and $\gamma_{2t} := \mathbb{E}_t[(M_t - \hat{M}_t)^2]$ satisfy

$$dM_t = \underbrace{(\kappa_{0t} + \kappa_{1t}a_t + \kappa_{2t}M_t)dt}_{=\mathbb{E}_t[(\kappa_{0t}+\kappa_{1t}a_t+\kappa_{2t}\hat{M}_t)dt]} + \frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2}[dX_t - (\delta_{0t} + \delta_{1t}M_t + \delta_{2t}L_t)dt] \tag{A.4}$$

$$\dot{\gamma}_{2t} = 2\kappa_{2t}\gamma_{2t} + (B_t^Y)^2 - (\gamma_{2t}\delta_{1t}/\sigma_X)^2, \tag{A.5}$$

where $dZ_t := [dX_t - (\delta_{0t} + \delta_{1t}M_t + \delta_{2t}L_t)dt]/\sigma_X$ is a Brownian motion from the sender's standpoint.[35] Observe that since (A.4) is linear, one can solve for M_t as an *explicit* function of past actions $(a_s)_{s<t}$ and past realizations of the public history $(X_s)_{s<t}$.

Inserting $a_t = \beta_{0t} + \beta_{1t}M_t + \beta_{2t}L_t + \beta_{3t}\theta$ in (A.4) and collecting terms yields $dM_t = [\hat{\kappa}_{0t} + \hat{\kappa}_{1t}M_t + \hat{\kappa}_{2t}L_t + \hat{\kappa}_{3t}\theta]dt + \hat{B}_t dX_t$, where,

$$\begin{aligned}
\hat{\kappa}_{0t} &= \left(\frac{\alpha_{3t}\gamma_{1t}}{\sigma_Y^2}\right)(\beta_{0t} - \alpha_{0t}) - \delta_{0t}\frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2} \\
\hat{\kappa}_{1t} &= \left(\frac{\alpha_{3t}\gamma_{1t}}{\sigma_Y^2}\right)(\beta_{1t} - \alpha_{3t}) - \delta_{1t}\frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2} \\
\hat{\kappa}_{2t} &= \left(\frac{\alpha_{3t}\gamma_{1t}}{\sigma_Y^2}\right)(\beta_{2t} - \alpha_{2t}) - \delta_{2t}\frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2} \\
\hat{\kappa}_{3t} &= \left(\frac{\alpha_{3t}\gamma_{1t}}{\sigma_Y^2}\right)\beta_{3t}, \qquad \hat{B}_t = \frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2}.
\end{aligned}$$

Now let $R(t,s) = \exp(\int_s^t \hat{\kappa}_{1u}du)$. Since $M_0 = \mu$, we have

$$M_t = R(t,0)\mu + \theta\int_0^t R(t,s)\hat{\kappa}_{3s}ds + \int_0^t R(t,s)[\hat{\kappa}_{0s} + \hat{\kappa}_{2s}L_s]ds + \int_0^t R(t,s)\hat{B}_s dX_s.$$

Imposing (7) yields the equations

$$\begin{aligned}
\chi_t &= \int_0^t R(t,s)\hat{\kappa}_{3s}ds \\
L_t &= [R(t,0)\mu + \int_0^t R(t,s)[\hat{\kappa}_{0s} + \hat{\kappa}_{2s}L_s]ds + \int_0^t R(t,s)\hat{B}_s dX_s]/[1-\chi_t].
\end{aligned}$$

The validity of the construction boils down to finding a solution to the previously stated

[35] Theorem 12.7 in Liptser and Shiryaev (1977) is stated for actions that depend on (θ, X) exclusively, but it also applies to those that condition on past play (i.e., on M). Indeed, from (A.2), $\hat{M}_t = \hat{M}_t^\dagger + A_t$ where $\hat{M}_t^\dagger = \hat{M}_t^\dagger[Z_t^Y; s<t]$ and $A_t = \int_0^t e^{\int_0^s \kappa_{2u}du}\kappa_{1s}a_s ds$. Applying the theorem to $(\hat{M}_t^\dagger, X_t)_{t\in[0,T]}$, yields a posterior mean $M_t^\dagger$ and variance $\gamma_{2t}^\dagger$ for $\hat{M}^\dagger$ such that $M^\dagger + A_t = M_t$ as in (A.4) and $\gamma_{2t} = \gamma_{2t}^\dagger$.

equation for χ that takes values in $[0,1)$. Indeed, when this is the case, it is easy to see that

$$dL_t = \frac{L_t[\hat{\kappa}_{1t} + \hat{\kappa}_{2t} + \hat{\kappa}_{3t}]dt + \hat{\kappa}_{0t}dt + \hat{B}_t dX_t}{1-\chi_t}, \tag{A.6}$$

from which it is easy to conclude that L is a (linear) function of X as conjectured.

We will find a solution to the χ-equation that is C^1 with values in $[0,1)$. Differentiating $\chi_t = \int_0^t R(t,s)\hat{\kappa}_{3s}ds$ then yields an ODE for χ as below that is coupled with γ_1 and γ_2:

$$\dot{\gamma}_{1t} = -\gamma_{1t}^2(\beta_{3t} + \beta_{1t}\chi_t)^2/\sigma_Y^2 \tag{A.7}$$

$$\dot{\gamma}_{2t} = -2\gamma_{2t}\gamma_{1t}(\beta_{3t} + \beta_{1t}\chi_t)^2/\sigma_Y^2 + \gamma_{1t}^2(\beta_{3t} + \beta_{1t}\chi_t)^2/\sigma_Y^2 - (\gamma_{2t}\delta_{1t})^2/\sigma_X^2 \tag{A.8}$$

$$\dot{\chi}_t = \gamma_{1t}(\beta_{3t} + \beta_{1t}\chi_t)^2(1-\chi_t)/\sigma_Y^2 - (\delta_{1t}\chi_t)(\gamma_{2t}\delta_{1t})/\sigma_X^2. \tag{A.9}$$

In the proof of Lemma A.1, we take the system above as a primitive and establish that $\chi = \gamma_2/\gamma_1$. Equipped with this, we set $\gamma_2 = \chi\gamma_1$ in the third ODE, and after writing γ for γ_1, the first and third ODEs become (10)–(11). The same Lemma A.1 further establishes the bounds $0 < \gamma_t \leq \gamma^o$ and $0 \leq \chi_t < 1$, with strict inequalities for all $t > 0$ if $\beta_{3,0} \neq 0$.

Using (i)–(v) that define $(\vec{\hat{\kappa}}, \hat{B})$, (A.6) becomes $dL_t = (\ell_{0t} + \ell_{1t}L_t)dt + B_t dX_t$, where

$$l_{0t} = -\frac{\gamma_t\chi_t\delta_{0t}\delta_{1t}}{\sigma_X^2(1-\chi_t)},\ l_{1t} = -\frac{\gamma_t\chi_t\delta_{1t}(\delta_{1t}+\delta_{2t})}{\sigma_X^2(1-\chi_t)},\ B_t = \frac{\gamma_t\chi_t\delta_{1t}}{\sigma_X^2(1-\chi_t)}. \tag{A.10}$$

That L_t coincides with $\mathbb{E}[\theta|\mathcal{F}_t^X]$ is proved in the Supplementary Appendix. □

Proof of Lemma 2. Using (A.3), (A.4) becomes

$$dM_t = \frac{\gamma_t\alpha_{3t}}{\sigma_Y^2}(a_t - [\alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}M_t])dt + \frac{\chi_t\gamma_t\delta_{1t}}{\sigma_X}dZ_t,$$

where $dZ_t := [dX_t - (\delta_{0t} + \delta_{1t}M_t + \delta_{2t}L_t)dt]/\sigma_X$ a Brownian motion from the sender's standpoint. As for the law of motion of L, this follows from (12) using (A.10) and that $dX_t = (\delta_{0t} + \delta_{2t}L_t + \delta_{1t}M_t)dt + \sigma_X dZ_t$ from the sender's perspective.

We conclude with three observations. First, from (A.2) and (A.4), $\hat{M}_t - M_t$ is independent of the strategy followed, and hence so is Z_t due to $\sigma_X dZ_t = \delta_{1t}(\hat{M}_t - M_t)dt + \sigma_X dZ_t^X$ under the true data-generating process. This strategic independence enables us to fix an exogenous Brownian motion Z and then solve the best-response problem with Z in the laws of motion of M and L—i.e., the so-called *separation principle* for control problems with unobserved states applies (see, for instance, Liptser and Shiryaev, 1977, Chapter 16).

Second, it is clear from (15), (A.4)–(A.5), and the proof of Lemma A.1 that no additional state variables are needed due to $\gamma_{2t} := \mathbb{E}_t[(M_t - \hat{M}_t)^2] = \chi_t\gamma_t$ holding irrespective of the strategy chosen. Third, the set of admissible strategies for the best-response problem then consists of all square-integrable processes that are progressively measurable with respect to (θ, M, L). This set is clearly the appropriate set, and richer than that in Definition 1. $\square$

Lemma A.1 (Learning ODEs)**.** *Suppose that* $(\beta_1, \beta_3, \delta_1)$ *are differentiable. Then, there is a unique solution to* (10)–(11)*, and this solution satisfies* $0 < \gamma_t \leq \gamma^o$ *and* $0 \leq \chi_t < 1$ *for all* $t \in [0, T]$*, with strict inequalities over* $(0, T]$ *if* $\beta_{3,0} \neq 0$*. The same conclusions hold if* $\delta_{1t} = \hat{u}_{\hat{a}\theta} + \hat{u}_{a\hat{a}}\alpha_{3t}$*.*

Proof. Consider the system in $(\gamma_1, \gamma_2, \chi)$ from the proof of Lemma 1. By Peano's Theorem, a solution exists in some interval $[0, T')$ where $T' > 0$. And since the system is locally Lipschitz continuous in $(\gamma_1, \gamma_2, \chi)$ uniformly in $t \in [0, T]$, the solution is unique over any interval of existence by the Picard-Lindelöf Theorem. By applying the comparison theorem to γ_1 and the zero function, we obtain $\gamma_1 > 0$; and clearly, $\dot{\gamma}_1 \leq 0$ so $\gamma_1 \leq \gamma^o$. Hence, γ_2/γ_1 is well-defined, and it is easy to verify that it satisfies the χ-ODE. Since the solution is unique whenever it exists, we conclude that $\chi = \gamma_2/\gamma_1$, as promised in Lemma 1; in other words, $\chi_t = \mathbb{E}_t[(M - \hat{M})^2]/\hat{\mathbb{E}}[(\theta - \hat{M})^2]$. We can therefore substitute $\gamma_2 = \chi\gamma_1$ into (A.7) and (A.9) and abbreviate γ_1 to γ to obtain (10)-(11). Next, we apply the comparison theorem to (11): first, with the zero function, we obtain $0 \leq \chi$, and second, with the constant function 1, we obtain $\chi < 1$.

Using these bounds, we argue that the solution to (10)-(11) exists over $[0, T]$. Suppose by way of contradiction that the maximum interval of existence is $[0, \tilde{T})$. Then since (γ, χ) and their derivatives are bounded, the solution can be extended to $\tilde{T}$. If $\tilde{T} = T$, we are done, and if $\tilde{T} < T$, by Peano's Theorem the solution can be further extended to $\tilde{T} + \epsilon$ for some $\epsilon > 0$, contradicting that $[0, \tilde{T})$ is the maximum interval of existence. We conclude that the solution exists over the whole horizon $[0, T]$.

If, moreover, $\beta_{3,0} \neq 0$, then $\dot{\gamma}_{1,0} < 0$ and $\dot{\chi}_0 > 0$. Hence, by continuity of $\dot{\gamma}_1$ and $\dot{\chi}$, there exists $\epsilon > 0$ such that $\gamma_{1t} < \gamma^o$ and $\chi_t > 0$ for all $t \in (0, \epsilon)$, and by the comparison theorem, these strict inequalities hold up to time T.

Lastly, suppose that $\delta_{1t} = \hat{u}_{\hat{a}\theta} + \hat{u}_{a\hat{a}}\alpha_{3t} = \hat{u}_{\hat{a}\theta} + \hat{u}_{a\hat{a}}(\beta_{1t}\chi_t + \beta_{3t})$, where (β_1, β_3) are differentiable. Then the system (10)-(11) changes in that the functional form of the operator is altered, but importantly, it still satisfies the conditions for the Peano and Picard-Lindelöf theorems, and the arguments above go through. $\square$

Appendix B: Proofs for Section 4

In this section, we prove Proposition 1, and we highlight the main steps for proving Proposition 2; the corner cases of the reputation game follow similar steps. All other results and assertions made in Section 4 are proved in the Supplementary Appendix.

Proof of Proposition 1. As in the proof of Theorem 1, by the Picard-Lindelöf theorem applied to the time-reversed ODEs, the strategy coefficients are pinned down by their terminal values. It is straightforward to check that $(\beta_0, v_1, v_3) = (0, 0, 0)$, $v_6 = \sigma_Y^2[-1 + 2\beta_1(1-\chi) + \alpha_3]/(4\alpha_3\gamma) - v_8/2$ and $\beta_2 = 1 - \beta_1 - \beta_3$ satisfy their respective ODEs and terminal conditions in any LME, so by uniqueness, we have $\beta_0 = 0$ and $\beta_1 + \beta_2 + \beta_3 = 1$. As for α_3, note that its terminal value is $\beta_{1T}\chi_T + \beta_{3T} = \frac{1}{2-\chi_T} > 0$, and its ODE is

$$\dot{\alpha}_{3t} = r\alpha_{3t}[\alpha_{3t}(2-\chi_t) - 1] - \frac{2\alpha_{3t}^3\gamma_t\chi_t}{\sigma_X^2\sigma_Y^2(1-\chi_t)}\left\{\sigma_Y^2\chi_t[1 - \alpha_{3t} - \beta_{1t}(1-\chi_t)] + \alpha_{3t}\gamma_t v_{8t}\right\}.$$

Applying the comparison theorem to α_3 (going backward in time) establishes $\alpha_3 > 0$. Now on the equilibrium path, $a_t = \alpha_{3t}\theta + \alpha_{2t}L_t = \alpha_{3t}\theta + (1-\alpha_{3t})L_t$, where $\alpha_2 \equiv 1 - \alpha_3$ follows from $\beta_1 + \beta_2 + \beta_3 = 1$. The receiver thus plays $\hat{a}_t = \hat{\mathbb{E}}_t[a_t] = \alpha_{3t}\hat{M}_t + \alpha_{2t}L_t$.

For the non-monotonicity result, assume $r > 0$. Using the strict inequalities in Lemma A.1, we have $\alpha_{3T} > 1/2$ and $\dot{\alpha}_{3T} = -\frac{2\alpha_{3T}^3\gamma_T\chi_T}{\sigma_X^2\sigma_Y^2(1-\chi_T)}\left\{\sigma_Y^2\chi_T\frac{1-\chi_T}{2(2-\chi_T)}\right\} < 0$ (i.e. α_3 is eventually decreasing). Now at $t = 0$, we have $\chi_0 = 0$ and thus $\dot{\alpha}_{3,0} = r\alpha_{3,0}(2\alpha_{3,0} - 1)$; it follows that $\dot{\alpha}_{3,0} > 0$ iff $\alpha_{3,0} > \frac{1}{2}$. Consider two cases: (i) $\alpha_{3,0} > \frac{1}{2}$ and (ii) $\alpha_{3,0} \leq \frac{1}{2}$. In case (i), we have $\dot{\alpha}_{3,0} > 0$. In case (ii), we have $\alpha_{3T} > \frac{1}{2} \geq \alpha_{3,0}$, so by the mean value theorem, $\dot{\alpha}_{3t} > 0$ for some $t \in (0, T)$. In either case, since $\dot{\alpha}_{3T} < 0$, α_3 is non-monotonic.

For the last statement in the proposition, we make use of the proof of Theorem 1 and (within it) the proof of Theorem C.1. Fixing $\rho, K > 0$, for sufficiently small T, it guarantees the existence of a solution to the BVP in $(\gamma, \chi, \beta_1, \tilde{\beta}_2, \beta_3, v_6, v_8)$ (where $\tilde{\beta}_2 := \beta_2/(1-\chi)$) with the following properties: the sender's strategy coefficients $(\beta_1, \tilde{\beta}_2, \beta_3)$ differ from their myopic counterparts $(\beta_{1t}^m, \tilde{\beta}_{2t}^m, \beta_{3t}^m) = \left(\frac{1}{2(2-\chi_t)}, \frac{1}{2(2-\chi_t)}, \frac{1}{2}\right)$ by at most K; $(\beta_1^m, \tilde{\beta}_2^m, \beta_3^m)$ are bounded in magnitude by ρ; and (γ, χ) are Lipschitz continuous with uniform Lipschitz constants (i.e., constants that depend on ρ and K but not T). Hence, given any constant K, for sufficiently small horizons, we can also ensure that $|\gamma_t - \gamma^o| \leq K$ and $|\chi_t| \leq K$. Now as $\chi_T \to 0$, we have $\beta_{1T} \to 1/4$, $\tilde{\beta}_{2T} \to 1/4$, and $\alpha_{3T} \to 1/2$, while $\beta_{3T} = 1/2$, so choosing K and T sufficiently small, $(\beta_1^m, \tilde{\beta}_2^m, \beta_3^m)$ can be made arbitrarily close to those same values. In turn, since $(\beta_1, \tilde{\beta}_2, \beta_3)$ and $(\beta_1^m, \tilde{\beta}_2^m, \beta_3^m)$ differ by at most K, choosing K and T sufficiently small ensures that β_{3t} and α_{3t} lie in $(0,1)$ and β_{1t} and $\tilde{\beta}_{2t}$ lie in $(0, 1/2)$; the latter implies $\beta_2 = (1-\chi_t)\tilde{\beta}_{2t} \in (0, 1/2)$.

To bound β_3 below by $1/2$, write the ODE for β_3 as $\dot{\beta}_{3t} = f^{\beta_3}(\gamma, \chi, \beta_1, \tilde{\beta}_2, \beta_3, v_6, v_8)$, where f^{β_3} is of class C^1. It is easy to check that $f^{\beta_3}(\mathbf{z}_0) < 0$, where $\mathbf{z}_0 := (\gamma^o, 0, 1/4, 1/4, 1/2, 0, 0)$. Hence, for sufficiently small K and associated T as above, β_3 is strictly decreasing. Given $\beta_{3T} = 1/2$, this implies $\beta_{3t} \geq 1/2$ for all t. □

Next, we highlight the main steps for characterizing LME of the coordination game in the cases $\sigma_X = 0$ and $\sigma_X = +\infty$, where the resulting boundary value problem is simpler. Omitted details can be found in the Supplementary Appendix.

Main steps for $\sigma_X = 0$ case We aim to characterize an LME in which the leader backs out the follower's belief from his action at all times, with strategies of the form $a_t = \beta_{0t} + \beta_{1t}\hat{M}_t + \beta_{3t}\theta$ and $\hat{a}_t = \hat{\mathbb{E}}_t[a_t] = \beta_{0t} + (\beta_{1t} + \beta_{3t})\hat{M}_t$, where $\beta_{1t} + \beta_{3t} \neq 0$, $t \in [0, T]$. The laws of motion for the follower's belief are

$$d\hat{M}_t = \frac{\beta_{3t}\gamma_t}{\sigma_Y^2}\{dY_t - [\underbrace{(\beta_{0t} + (\beta_{1t} + \beta_{3t})\hat{M}_t)}_{=\hat{\mathbb{E}}_t[a_t]} dt]\} \text{ and } \dot{\gamma}_t = -\left(\frac{\beta_{3t}\gamma_t}{\sigma_Y}\right)^2, \tag{B.1}$$

with initial values $\hat{M}_0 = \mu$ and $\gamma_0 = \gamma^o$. Clearly, $(\theta, \hat{M}_t, t)$ are the relevant states for the leader's problem, as $\hat{M}_t$ is public. Let $V : \mathbb{R}^2 \times [0, T] \to \mathbb{R}$ denote the leader's value function. Given the law of motion for $\hat{M}_t$, the HJB equation is

$$rV = \sup_{a \in \mathbb{R}} \left\{ \frac{1}{4}[-(a-\theta)^2 - (a - \hat{a}_t)^2] + \frac{\beta_{3t}\gamma_t}{\sigma_Y^2}[a - \beta_{0t} - (\beta_{1t} + \beta_{3t})m]V_m + \frac{\beta_{3t}^2\gamma_t^2}{2\sigma_Y^2}V_{mm} + V_t \right\}.$$

We guess a quadratic solution $V(\theta, m, t) = v_{0t} + v_{1t}\theta + v_{2t}m + v_{3t}\theta^2 + v_{4t}m^2 + v_{5t}\theta m$ and derive a system of ODEs for $(\beta_0, \beta_1, \beta_3)$ subject to terminal conditions $(\beta_{0T}, \beta_{1T}, \beta_{3T}) = (0, 1/2, 1/2)$; these ODEs depend on γ, which evolves according to $\dot{\gamma}_t = -\gamma_t^2\beta_{3t}^2/\sigma_Y^2$ with initial condition $\gamma_0 = \gamma^o$. The key step for establishing existence is establishing a solution to this BVP; after that, it is easy to recover the value function coefficients.

Since this BVP only involes one ODE going forward, it can be solved using a traditional shooting method. Specifically, we transform it into a backward IVP by reversing time and using a parametrized initial value for γ. We then show that by the intermediate value theorem, there is $\gamma^F > 0$ such that $\gamma_T = \gamma^o$ in the backward system while all the other ODEs admit solutions. As in Bonatti et al. (2017), it suffices to show that the solutions are uniformly bounded when $\gamma_t \in [0, \gamma^o]$ for $t \in [0, T]$. Using the comparison theorem, we show that $\beta_0, \beta_1, \beta_3 \in [0, 1]$ as long as γ does not explode, so there exists a solution to the BVP, and hence an LME. The remaining arguments are carried out in the Supplementary Appendix.

Main steps for $\sigma_X = +\infty$ case We look for an equilibrium in which the leader plays $a_t = \beta_{1t}M_t + \beta_{2t}\mu + \beta_{3t}\theta$. We first derive a representation for the leader's second-order belief.

Lemma B.1 (Belief Representation)**.** *Assume $\sigma_X = +\infty$. Suppose the follower expects $a_t = \alpha_{2t}\mu + \alpha_{3t}\theta$, where $\alpha_2 = \beta_2 + \beta_1(1-\chi)$, $\alpha_3 = \beta_3 + \beta_1\chi$, $\chi = 1 - \gamma/\gamma^o$, and $\gamma_t := \hat{\mathbb{E}}_t[(\theta - \hat{M}_t)^2]$. Then $\dot{\gamma}_t = -\left(\frac{\gamma_t\alpha_{3t}}{\sigma_Y^2}\right)^2$. Moreover, if the leader follows $a_t = \alpha_{2t}\mu + \alpha_{3t}\theta$, $M_t = \chi_t\theta + (1-\chi_t)\mu$ holds at all times.*

The states (θ, M_t, t) are sufficient on and off path for the leader, since the follower's strategy will be linear in $(\mu, \hat{M}_t)$, and then in the leader's expected flow payoff, $\mathbb{E}_t[\hat{M}_t] = M_t$ and $\mathbb{E}_t[\hat{M}_t^2] = M_t^2 + \gamma_t\chi_t$ after all private histories. (See Supplementary Appendix for details.)

We can now set up the HJB equation. The leader controls M, which evolves as

$$dM_t = \frac{\alpha_{3t}\gamma_t}{\sigma_Y^2}\left(a - \alpha_{2t}\mu - \alpha_{3t}M_t\right)dt$$

with $M_0 = \mu$. The HJB equation is thus

$$\begin{aligned} rV = \sup_{a\in\mathbb{R}} \Bigg\{ &\frac{1}{4}[-(a-\theta)^2 - \left(a^2 - 2a[\alpha_{2t}\mu + \alpha_{3t}m] + \alpha_{2t}^2\mu^2 + 2\alpha_{2t}\alpha_{3t}\mu m + \alpha_{3t}^2[m^2 + \gamma_t\chi_t]\right)] \\ &+ V_t + \frac{\alpha_{3t}\gamma_t}{\sigma_Y^2}\left(a - \alpha_{2t}\mu - \alpha_{3t}m\right)V_m \Bigg\}. \end{aligned}$$

We then guess $V(\theta, m, \mu, t) = v_{0t} + v_{1t}\theta + v_{2t}m + v_{3t}\mu + v_{4t}\theta^2 + v_{5t}m^2 + v_{6t}\mu^2 + v_{7t}\theta m + v_{8t}\theta\mu + v_{9t}m\mu$ and take analogous steps to those in the proof for the $\sigma_X = 0$ case. In particular, we obtain a boundary value problem in $(\beta_1, \beta_2, \beta_3, \gamma)$, transform it in to an initial value problem, and solve it using the same one-dimensional shooting method as for $\sigma_X = 0$ case.

Appendix C: Proofs for Section 5

Overview of approach Our overall proof strategy consists of reducing the HJB equation (19) subject to the equilibrium condition (20) to a suitable boundary value problem that we then solve using a fixed-point argument. The BVP will contain ODEs linked to behavior—hence, involving terminal conditions—and also the learning ODEs for (γ, χ) that have initial conditions. The fixed point will be over pairs of functions (γ, χ): a pair (γ^*, χ^*) that generates mutual best responses that in turn induce learning ODEs whose solution is (γ^*, χ^*).

This overarching goal requires several intermediate steps, which we label *core subsystem*, *centering*, *auxiliary variable*, *fixed point* and *verification*; we provide brief explanations of these when they arise. Throughout the proof, we refer to the *myopic equilibrium coefficients*

$$(\beta_{0t}^m, \beta_{1t}^m, \beta_{2t}^m, \beta_{3t}^m) = \left(\frac{u_0 + u_{a\hat{a}}\hat{u}_0}{1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}}, \frac{u_{a\hat{a}}(u_{a\theta}\hat{u}_{a\hat{a}} + \hat{u}_{\hat{a}\theta})}{1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t}, \frac{u_{a\hat{a}}^2\hat{u}_{a\hat{a}}(u_{a\theta}\hat{u}_{a\hat{a}} + \hat{u}_{\hat{a}\theta})(1-\chi_t)}{(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t)}, u_{a\theta} \right),$$

which correspond to the sender's strategy coefficients in the unique linear Bayes Nash equilibrium involving states $(\theta, M, \hat{M}, L)$ of the static game with flow utilities $(u, \hat{u})$ if the receiver believes $M_t = \chi_t\theta_t + (1-\chi_t)L_t$. By Assumption 2, $(\beta_{0t}^m, \beta_{1t}^m, \beta_{2t}^m, \beta_{3t}^m)$ is well defined and $\alpha_t^m := \beta_{1t}^m\chi_t + \beta_{3t}^m \neq 0$ for all $\chi_t \in [0,1]$. Henceforth, given χ_t, we write β_{it}^m and α_t^m to refer to these functions of χ_t, suppressing the dependence on χ_t, and we abbreviate α_3 to α.

Core subsystem: *We show that the problem of existence of LME reduces to a core subsystem in* $(\gamma, \chi, \vec{\beta}, v_6, v_8)$*, where* $\vec{\beta} := (\beta_1, \beta_2, \beta_3)$*, and perform a change of variables for* (β_2, v_6, v_8)*; we denote the new system by* $(\gamma, \chi, \beta_1, \tilde{\beta}_2, \beta_3, \tilde{v}_6, \tilde{v}_8)$.

The first thing to note is that $\alpha_t := \beta_{1t}\chi_t + \beta_{3t} \neq 0$ for all $t \in [0,T]$ in any LME. Indeed, if $\alpha_t = 0$, it is then easy to verify from the HJB equation that $\beta_{it} = \beta_{it}^m$ for $i \in \{0,1,2,3\}$: since the sender's actions transmit no information, both players must be using myopic best responses. But this implies that $\alpha_t = \alpha_t^m \neq 0$ in such an LME, a contradiction. Second, since the coefficients $(\beta_0, \beta_1, \beta_2, \beta_3)$ and χ will be continuous, it follows that $\gamma_t > 0$ at all times by Lemma A.1. From the HJB equation, it is easy to see that

$$v_{2t} = -\sigma_Y^2[u_{ac} + u_{a\hat{a}}\hat{u}_{\hat{a}c} - (1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})\beta_{0t}]/(\alpha_t\gamma_t) \tag{C.1}$$

$$v_{5t} = -\sigma_Y^2[u_{a\hat{a}}\hat{u}_{\hat{a}\theta} + u_{a\hat{a}}\hat{u}_{a\hat{a}}\alpha_t - \beta_{1t}]/(2\alpha_t\gamma_t) \tag{C.2}$$

$$v_{7t} = -\sigma_Y^2[u_{a\theta} - \beta_{3t}]/(\alpha_t\gamma_t) \tag{C.3}$$

$$v_{9t} = -\sigma_Y^2[u_{a\hat{a}}\hat{u}_{a\hat{a}}\beta_{1t}(1-\chi_t) - \beta_{2t}(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})]/(\alpha_t\gamma_t). \tag{C.4}$$

Expressions (C.1)-(C.4) allow us to eliminate v_i and $\dot{v}_i$, $i \in \{2,5,7,9\}$, in the HJB equation to get a system of ODEs for $(\gamma, \chi, \beta_0, \vec{\beta}, v_0, v_1, v_3, v_4, v_6, v_8)$—as a last step we verify that our (α, γ) satisfy $|\alpha_t||\gamma_t| > 0$ all $t \in [0,T]$, recovering the value function through (C.1)-(C.4).

The expressions in this system can be found in the Mathematica file `spm.nb` on our websites—we omit them in favor of stating the core subsystem with which we will be working below. The omitted system has three properties easily verified by inspection in the same file:

(i) the ODEs for $(\vec{\beta}, v_6, v_8)$ do not contain $(v_0, v_1, v_3, v_4, \beta_0)$;

(ii) given $(\vec{\beta}, v_6, v_8)$, $(v_0, v_1, v_3, v_4, \beta_0)$ form a non-homogeneous linear ODE system; and

(iii) $(\vec{\beta}, v_6, v_8)$ carries $(1-\chi)$ in the denominator.

Parts (i) and (ii) imply that we can focus on the sub-system $(\vec{\beta}, v_6, v_8)$, as any linear system with continuous coefficients admits a unique solution for all times (Teschl, 2012, Corollary 2.6).[36] Part (iii) reflects that the dynamic for L carries a denominator of that form; by Lemma A.1, however, we know that $\chi \in [0,1)$ if the coefficients are continuous.

It is then convenient to use the change of variables $(\tilde{\beta}_2, \tilde{v}_6, \tilde{v}_8) = (\beta_2/(1-\chi), v_6\gamma/(1-\chi)^2, v_8\gamma/(1-\chi))$ that eliminates this denominator in the resulting system for the functions $(\gamma, \chi, \beta_1, \tilde{\beta}_2, \beta_3, \tilde{v}_6, \tilde{v}_8)$—because (χ, γ) only depend on (β_1, β_3) directly, it follows that $\chi \in [0,1)$ and $\gamma > 0$ in any solution to this system, and we trivially recover (β_2, v_6, v_8).[37]

We can now state the core subsystem of ODEs for $(\gamma, \chi, \beta_1, \tilde{\beta}_2, \beta_3, \tilde{v}_6, \tilde{v}_8)$ with which we will be working. Recall that $\delta_{1t} = \hat{u}_{\hat{a}\theta} + \hat{u}_{a\hat{a}}(\beta_{1t}\chi_t + \beta_{3t})$.

$$
\begin{aligned}
\dot{\tilde{v}}_{6t} &= \tilde{v}_{6t}[r + \alpha_t^2\gamma_t/\sigma_Y^2 + 2\delta_{1t}^2\gamma_t\chi_t/\sigma_X^2] - (\gamma_t/2)\Big\{\beta_{1t}^2\hat{u}_{a\hat{a}}[2u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}] \\
&\quad + \tilde{\beta}_{2t}(2\beta_{1t} + \tilde{\beta}_{2t})[-1 + 2u_{a\hat{a}}\hat{u}_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}^2]\Big\} \\
\dot{\tilde{v}}_{8t} &= \tilde{v}_{8t}[r + \delta_{1t}^2\gamma_t\chi_t/\sigma_X^2] - \gamma_t\Big\{(\tilde{\beta}_2 + \beta_{1t})[u_{a\theta} + u_{\hat{a}\theta}\hat{u}_{a\hat{a}}] - \beta_{1t}\beta_{3t}\Big\} \\
\dot{\beta}_{1t} &= r\frac{\alpha_t}{\alpha_t^m}[\beta_{1t} - \beta_{1t}^m] - \gamma_t[\sigma_X^2\sigma_Y^2(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)]^{-1}\times \\
&\quad \Big\{\tilde{\beta}_{2t}2(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})\sigma_Y^2\delta_{1t}^2\chi_t(u_{a\theta} + \beta_{1t}\chi_t) + \beta_{1t}^2[\sigma_X^2\alpha_t(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t) + (1 - 2u_{a\hat{a}}\hat{u}_{a\hat{a}})\sigma_Y^2\delta_{1t}^2\chi_t^2] \\
&\quad + \beta_{1t}\sigma_X^2\alpha_t[\hat{u}_{a\hat{a}}(u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}})\alpha_t^2\chi_t + \hat{u}_{\hat{a}\theta}(u_{\hat{a}\theta} + u_{\hat{a}\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t) + \alpha_t(-u_{a\theta} + u_{\hat{a}\theta}\hat{u}_{a\hat{a}} + 2u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)] \\
&\quad - \beta_{1t}\sigma_Y^2 u_{a\hat{a}}\delta_{1t}^2\chi_t(2u_{a\theta}\hat{u}_{a\hat{a}} + \hat{u}_{\hat{a}\theta}\chi_t) + \delta_{1t}^2\tilde{v}_{8t}\alpha_t\chi_t(\beta_{1t} - u_{a\hat{a}}\hat{u}_{\hat{a}\theta}) \\
&\quad - \sigma_X^2\delta_{1t}\alpha_t[u_{a\hat{a}}(u_{\hat{a}\theta}\hat{u}_{\hat{a}\theta} - u_{a\theta}\alpha_t) - u_{\hat{a}\hat{a}}u_{a\theta}\delta_{1t}]\Big\}. \\
\dot{\tilde{\beta}}_{2t} &= r\frac{\alpha_t}{\alpha_t^m}[\tilde{\beta}_{2t} - \tilde{\beta}_{2t}^m] - \gamma_t[\sigma_X^2\sigma_Y^2(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})]^{-1}\times \\
&\quad \Big\{\delta_{1t}^2\alpha_t\chi_t[2\tilde{v}_{6t}(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t) - u_{a\hat{a}}^2\hat{u}_{a\hat{a}}\hat{u}_{\hat{a}\theta}\tilde{v}_{8t}] \\
&\quad + \tilde{\beta}_{2t}\sigma_X^2\alpha_t[\hat{u}_{a\hat{a}}(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})(u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}})\alpha_t^2\chi_t + \hat{u}_{\hat{a}\theta}(u_{\hat{a}\theta} + u_{a\hat{a}}u_{a\theta} - u_{a\hat{a}}u_{\hat{a}\theta}\hat{u}_{a\hat{a}} + u_{\hat{a}\hat{a}}u_{a\theta}\hat{u}_{a\hat{a}} \\
&\quad + [u_{a\hat{a}}^2 + u_{\hat{a}\hat{a}}]\hat{u}_{\hat{a}\theta}\chi_t) + \alpha_t(u_{\hat{a}\theta}\hat{u}_{a\hat{a}}[1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}] + u_{a\theta}[-1 + 2u_{a\hat{a}}\hat{u}_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}^2] \\
&\quad + \hat{u}_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t[u_{a\hat{a}}^2 + 2u_{\hat{a}\hat{a}} - u_{a\hat{a}}u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}])] + \delta_{1t}[\sigma_X^2 u_{a\hat{a}}\hat{u}_{a\hat{a}}\alpha_t(u_{a\theta}u_{\hat{a}\hat{a}}\delta_{1t} + u_{a\hat{a}}[u_{a\theta}\alpha_t - u_{\hat{a}\theta}\hat{u}_{\hat{a}\theta}])] \\
&\quad - \tilde{\beta}_{2t}\sigma_Y^2(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})\delta_{1t}^2\chi_t[u_{a\theta}(1 - 2u_{a\hat{a}}\hat{u}_{a\hat{a}}) + \chi_t(u_{a\hat{a}}\hat{u}_{\hat{a}\theta} - \beta_{1t}[1 - 2u_{a\hat{a}}\hat{u}_{a\hat{a}}])] \\
&\quad + \alpha_t(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})[\tilde{\beta}_{2t}\tilde{v}_{8t}\delta_{1t}^2\chi_t - \sigma_X^2\beta_{1t}^2(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)] + 2\sigma_Y^2\delta_{1t}^2\tilde{\beta}_{2t}^2\chi_t^2(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})^2 \\
&\quad + \beta_{1t}\delta_{1t}[\sigma_X^2\alpha_t(u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}})(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t) + \sigma_Y^2\delta_{1t}\chi_t u_{a\hat{a}}u_{a\theta}\hat{u}_{a\hat{a}}(1 - 2u_{a\hat{a}}\hat{u}_{a\hat{a}})]\Big\}
\end{aligned}
$$

[36] Intuitively, (v_0, v_1, v_4) are the coefficients of the constant, θ- and θ^2-terms in the sender's value function, none of which the sender controls, so they do not affect the rest of the system. The equations for (β_0, v_3) are coupled and encode the *deterministic* component of the sender's incentive to manipulate beliefs; they do not enter the sub-system for $(\vec{\beta}, v_6, v_8)$ but depend on the latter through the signal-to-noise ratio in Y.

[37] Our method for finding intervals of existence of LME relies on bounding solutions to ODEs uniformly, and this denominator would unnecessarily complicate that task since there is no upper bound on $1/(1-\chi)$ that applies to all environments. This change of variables is akin to working with $\tilde{L} = (1-\chi)L$ instead of L.

$$\dot{\beta}_{3t} = r\frac{\alpha_t}{\alpha_t^m}[\beta_{3t} - \beta_{3t}^m] - \gamma_t[\sigma_X^2\sigma_Y^2(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)]^{-1}\times$$
$$\Big\{\tilde{\beta}_{2t}2(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})\sigma_Y^2\delta_{1t}^2\chi_t^2(\beta_{3t} - u_{a\theta}) - \beta_{1t}^2\chi_t[\sigma_X^2\alpha_t(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t) + \sigma_Y^2\delta_{1t}^2\chi_t^2(1 - 2u_{a\hat{a}}\hat{u}_{a\hat{a}})]$$
$$-\beta_{1t}\alpha_t\sigma_X^2[\hat{u}_{a\hat{a}}(u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}})\alpha_t^2\chi_t^2 + \hat{u}_{\hat{a}\theta}\chi_t(u_{\hat{a}\theta} + u_{\hat{a}\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)$$
$$+\alpha_t([u_{\hat{a}\theta}\hat{u}_{a\hat{a}} - u_{a\hat{a}}\hat{u}_{\hat{a}\theta}]\chi_t - u_{a\theta} + [u_{a\hat{a}} + 2u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}]\hat{u}_{\hat{a}\theta}\chi_t^2)] - \beta_{1t}\delta_{1t}^2\chi_t^2\sigma_Y^2(1 - 2u_{a\hat{a}}\hat{u}_{a\hat{a}})(u_{a\theta} - \alpha_t)$$
$$+\delta_{1t}^2\alpha_t\chi_t\tilde{v}_{8t}(\beta_{3t} + \chi_t u_{a\hat{a}}\hat{u}_{\hat{a}\theta}) + \sigma_X^2\delta_{1t}\alpha_t[(u_{a\hat{a}}u_{\hat{a}\theta} - u_{\hat{a}\hat{a}}u_{a\theta})\hat{u}_{\hat{a}\theta}\chi_t + (u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}})\alpha_t^2\chi_t$$
$$+\alpha_t(u_{\hat{a}\theta} - \chi_t[u_{a\theta}(u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}) - u_{\hat{a}\hat{a}}\hat{u}_{\hat{a}\theta}])]\Big\}$$
$$\dot{\gamma}_t = -(\beta_{1t}\chi_t + \beta_{3t})^2\gamma_t^2/\sigma_Y^2, \quad \dot{\chi}_t = \gamma_t\left[(\beta_{1t}\chi_t + \beta_{3t})^2(1 - \chi_t)/\sigma_Y^2 - \delta_{1t}^2\chi_t^2/\sigma_X^2\right].$$

This system has two initial conditions $(\gamma_0, \chi_0) = (\gamma^o, 0)$. It also has terminal conditions for $(\beta_{1T}, \tilde{\beta}_{2T}, \beta_{3T}, \tilde{v}_{6T}, \tilde{v}_{8T})$ that depend on whether there are terminal payoffs. In what follows, we focus on the case without terminal payoffs—i.e., where the terminal conditions are $(\beta_{1T}^m, \tilde{\beta}_{2T}^m, \beta_{3T}^m, 0, 0)$—postponing the discussion of terminal payoffs to the end of the analysis. We note that the remaining denominators never vanish thanks to Assumption 2, and that all the ODEs carry r-independent terms that scale linearly in γ; this latter property will allow us to find horizons for existence that are inversely proportional to γ^o.

Centering: *To exploit discounting, we focus on the* centered *system* $(\gamma, \chi, \beta_1^c, \tilde{\beta}_2^c, \beta_3^c, \tilde{v}_6, \tilde{v}_8)$, *where* $(\beta_1^c, \tilde{\beta}_2^c, \beta_3^c)$ *denotes* $(\beta_1, \tilde{\beta}_2, \beta_3)$ *net of the myopic counterpart.* The tuple $(\beta_1, \tilde{\beta}_2, \beta_3)$ is constructed going backward in time from its terminal value as with backward induction in discrete time. One would expect higher discount rates to pull these coefficients towards the myopic values more strongly, thereby facilitating the existence of LME. Indeed, the term $-r\frac{\alpha}{\alpha^m}(\beta_i - \beta_i^m)$ in the time-reversed version of the β_i-ODE reflects this fact as long as $\alpha := \beta_1\chi + \beta_3$ does not change sign. To exploit the effect of discounting when finding intervals of existence, it is then useful to introduce the *centered* coefficients, i.e., $x_{it}^c := x_{it} - x_{it}^m$ for $x \in \{\beta_1, \tilde{\beta}_2, \beta_3\}$, and work with the ODEs of $(\beta_1^c, \tilde{\beta}_2^c, \beta_3^c, \tilde{v}_6, \tilde{v}_8)$ in backward form.[38]

The next lemma states the key properties of this backward centered system, noting that (i) the RHS of the ODEs for $(\beta_1, \tilde{\beta}_2, \beta_3)$ above are polynomials in $(\beta_1, \tilde{\beta}_2, \beta_3) = (\beta_1^c + \beta_1^m, \tilde{\beta}_2^c + \tilde{\beta}_2^m, \beta_3^c + \beta_3^m)$, (ii) $(\beta_1^m, \tilde{\beta}_2^m, \beta_3^m)$ are functions of χ and are independent of r, (iii) $(\dot{\beta}_1^m, \dot{\tilde{\beta}}_2^m, \dot{\beta}_3^m)$ carry a factor of γ through $\dot{\chi}$, and (iv) $\alpha_t^m = \frac{u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t}{1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t}$. (The proof is straightforward and hence omitted.) Without fear of confusion, in the lemma and in what follows we denote the solution to the backward system by $(\beta_1^c, \tilde{\beta}_2^c, \beta_3^c, \tilde{v}_6, \tilde{v}_8)$ (and unless otherwise stated, we always refer to the backward system when invoking this tuple). Also, let $\vec{\beta}^c := (\beta_1^c, \tilde{\beta}_2^c, \beta_3^c)$.

[38] This centering step can be sometimes skipped when intervals of existence can be readily obtained without resorting to the "worst" $r = 0$ case. See the proofs of Propositions ?? and ??. We also note that a backward first-order ODE of a function f is obtained by differentiating $\tilde{f} = f(T - t)$, and hence only differs with the original one in the sign. We maintain the labels to avoid further notational burden.

Lemma C.1. *For $x \in \{\beta_1, \tilde{\beta}_2, \beta_3\}$ and $y \in \{\tilde{v}_6, \tilde{v}_8\}$, the (backward) ODEs that x^c and y satisfy have the form*

$$\dot{x}_t^c = -rx_t^c \frac{\alpha_t}{\alpha_t^m} + \frac{\gamma_t h_x(\vec{\beta}^c, \tilde{v}_{6t}, \tilde{v}_{8t}, \chi_t)}{\sigma_X^2 \sigma_Y^2 (u_{a\theta} + u_{a\hat{a}} \hat{u}_{\hat{a}\theta} \chi_t)^{n_{1,x}} (1 - u_{a\hat{a}} \hat{u}_{a\hat{a}} \chi_t)^{n_{2,x}} (1 - u_{a\hat{a}} \hat{u}_{a\hat{a}})^{n_{3,x}}}$$

$$\dot{y}_t = -y_t[r + \gamma_t R_y(\vec{\beta}^c, \tilde{v}_{6t}, \tilde{v}_{8t}, \chi_t)] + \frac{\gamma_t h_y(\vec{\beta}^c, \chi_t)}{\sigma_X^2 \sigma_Y^2 (u_{a\theta} + u_{a\hat{a}} \hat{u}_{\hat{a}\theta} \chi_t)^{n_{1,y}} (1 - u_{a\hat{a}} \hat{u}_{a\hat{a}} \chi_t)^{n_{2,y}} (1 - u_{a\hat{a}} \hat{u}_{a\hat{a}})^{n_{3,y}}},$$

where $n_{i,x}, n_{i,y} \in \mathbb{N}$, $i = 1, 2, 3$, and h_x, h_y, and $R_y \geq 0$ are polynomials.[39] *The initial conditions are $(\vec{\beta}_0^c, \tilde{v}_{60}, \tilde{v}_{80}) = (0, 0, 0, 0, 0)$.*

In particular, notice that (i) the terms not containing r continue scaling with γ, (ii) the denominators are bounded away from zero, and (iii) the discount rate term pushes any solution towards zero when α does not change sign. We turn to this issue in the next step.

Auxiliary variable: *To exploit discounting, we introduce an auxiliary variable $\tilde{\alpha} \neq 0$ and work with an ODE-system for $(\gamma, \chi, \beta_1^c, \tilde{\beta}_2^c, \beta_3^c, \tilde{v}_6, \tilde{v}_8, \tilde{\alpha})$.* Observe that α will indeed never vanish in any solution to the centered system. In fact, a tedious but straightforward exercise shows that in *backward* form, $\alpha = \beta_1 \chi + \beta_3$ satisfies

$$\begin{aligned}\dot{\alpha}_t = \alpha_t \Bigg\{ &-r\left(\frac{\alpha_t}{\alpha_t^m} - 1\right) + \gamma_t[\sigma_X^2 \sigma_Y^2 (u_{a\theta} + u_{a\hat{a}} \hat{u}_{\hat{a}\theta} \chi_t)]^{-1} \times \\ &\Big\{ \delta_{1t}[\beta_{1t}\chi_t + \beta_{3t}]\sigma_X^2 u_{\hat{a}\theta} + \delta_{1t}\chi_t[\delta_{1t}\chi_t \sigma_Y^2 (2\tilde{\beta}_{2t}[1 - u_{a\hat{a}} \hat{u}_{a\hat{a}}] + \beta_{1t}[1 - 2u_{a\hat{a}} \hat{u}_{a\hat{a}}]) \\ &+ (\beta_{1t}\chi_t + \beta_{3t})(\delta_{1t}\tilde{v}_{8t} + \sigma_X^2[u_{\hat{a}\hat{a}}\delta_{1t} + u_{a\hat{a}}(\beta_{1t}\chi_t + \beta_{3t})])]\Big\}\Bigg\}, \end{aligned} \tag{C.5}$$

with initial condition $\alpha_0 = \alpha_0^m = \frac{u_{a\theta} + u_{a\hat{a}} \hat{u}_{\hat{a}\theta} \chi_0}{1 - u_{a\hat{a}} \hat{u}_{a\hat{a}} \chi_0}$ (here, for consistency, χ_0 is the terminal value of χ going forward in time). By Assumption 2, α_0^m always has the same sign as $u_{a\theta}$ because $\chi_0 \in [0, 1]$. Also, the right-hand side of (C.5) is proportional to α, so it vanishes at $\alpha \equiv 0$. By the comparison theorem, α is always nonzero, as the ODE is locally Lipschitz continuous in α uniformly in time. Moreover, since α^m never changes sign, $\alpha / \alpha^m > 0$.

However, our fixed point argument will input *general* (γ, χ) pairs into the backward ODEs of Lemma C.1, pairs that need not solve the learning ODEs (or even be differentiable). Thus, we will not be able to use a comparison argument like that above to show that each induced $\alpha := \beta_1 \chi + \beta_3$ never changes sign for any (γ, χ), allowing us to exploit the discount rate.

To circumvent this difficulty, we augment the BVP with an auxiliary variable $\tilde{\alpha}$ to serve as a proxy for α in the r term in the centered system; by construction, it will share the sign of α^m and, in any solution to the BVP, will coincide with α. Specifically, observe that using the decomposition $x = x^c + x^m$ for $x \in \{\beta_1, \tilde{\beta}_2, \beta_3\}$ yields that the r-independent term inside

[39] More precisely, we have $n_{1,x} = 1$, $n_{1,y} = 0$, and $n_{3,\beta_1} = n_{3,\beta_3} = 0$.

the outer brace of (C.5) is of the form $\frac{\gamma_t h_\alpha(\vec{\beta}^c,\tilde{v}_6,\tilde{v}_8,\chi_t)}{\sigma_X^2\sigma_Y^2(u_{a\theta}+u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)^{n_{1,\alpha}}(1-u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t)^{n_{2,\alpha}}(1-u_{a\hat{a}}\hat{u}_{a\hat{a}})^{n_{3,\alpha}}}$, where h_α is a polynomial and $n_{j,\alpha} \in \mathbb{N}$, $j = 1, 2, 3$. We introduce the (backward) linear ODE

$$\dot{\tilde{\alpha}}_t = \tilde{\alpha}_t \left\{ -r\left(\frac{\alpha_t}{\alpha_t^m} - 1\right) + \frac{\sigma_X^{-2}\sigma_Y^{-2}\gamma_t h_\alpha(\vec{\beta}^c, \tilde{v}_6, \tilde{v}_8, \chi_t)}{(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)^{n_{1,\alpha}}(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t)^{n_{2,\alpha}}(1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})^{n_{3,\alpha}}} \right\} \tag{C.6}$$

with initial condition $\tilde{\alpha}_0 = \alpha_0^m$. That is, the right-hand side of (C.6) is exactly as the one in (C.5) except for $\tilde{\alpha}$ now multiplying the bracket. The exact same application of the comparison argument between α and 0 shows that $\tilde{\alpha}$ never vanishes over its interval of existence *for any pair* (γ, χ) *Lipschitz taking values in* $[0, \gamma^o] \times [0, 1]$, and $\tilde{\alpha}/\alpha^m > 0$.

Our augmented BVP then consists of the ODEs of $x^c = \beta_1^c, \tilde{\beta}_2^c, \beta_3^c$ in Lemma C.1 with a modified r-term of the form $-rx_t^c \frac{\tilde{\alpha}_t}{\alpha_t^m}$, i.e., with $\tilde{\alpha}$ replacing α in the numerator of the fraction accompanying r. It also includes: the ODEs of $y = \tilde{v}_6, \tilde{v}_8$; the learning ODEs (10)-(11); and the ODE (C.6) of $\tilde{\alpha}$.[40] The resulting system of ODEs—denote it $\dot{\mathbf{z}}_t = F(\mathbf{z}_t)$, where $\mathbf{z} := (\gamma, \chi, \vec{\beta}^c, \tilde{v}_6, \tilde{v}_8, \tilde{\alpha})$—is such that each component of $F(\mathbf{z})$ is a polynomial divided by a product of powers of $1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}$, $1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t$, and $u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t$. Since the latter are bounded away from zero, F is of class C^1. We verify at the end of the proof that any solution to this augmented BVP satisfies that $\alpha := \beta_1\chi + \beta_3$ coincides with $\tilde{\alpha}$ by construction.[41]

Fixed point: *Use a fixed-point argument to show that there are horizon lengths of order* $1/\gamma^o$ *such that the augmented BVP admits a solution.* We will prove the following result:

Theorem C.1. *Under Assumptions 1 and 2, there is a strictly positive function* $T(\gamma^o) \in \Omega(1/\gamma^o)$ *such that if* $T < T(\gamma^o)$, *there exists a solution to the BVP in* $\mathbf{z} = (\gamma, \chi, \vec{\beta}^c, \tilde{v}_6, \tilde{v}_8, \tilde{\alpha})$.

Proof. The proof consists of converting the BVP into a fixed point problem over pairs $\lambda := (\gamma, \chi)$ in a suitable set. Specifically, for a given λ we can first solve the backward initial value problem (IVP) in the variables $(\vec{\beta}^c, \tilde{v}_6, \tilde{v}_8, \tilde{\alpha})$ that takes λ as an input. Second, we can solve the forward IVP for the two learning coefficients that takes as an input the solution from the previous step. This procedure generates a continuous mapping from candidate λ paths in a suitable set to itself, to which we apply Schauder's fixed point theorem.

Step 1: *Define the domain for our fixed point equation.* Let $\mathcal{C}$ denote the Banach space of continuous functions from $[0, T]$ to $\mathbb{R}$, equipped with the sup norm $||\cdot||_\infty$ defined by $||x||_\infty := \sup\{|x_t| : t \in [0, T]\}$. (To economize on notation, we use $||\cdot||_\infty$ to denote the supremum norm for objects of all other dimensions too.) By the Arzela-Ascoli theorem (Ok, 2007, p. 198), the space of uniformly bounded functions with a common Lipschitz constant

[40] For consistency, the α_t in the r-term in (C.6) and in (10)-(11) must be written as $(\beta_{1t}^c + \beta_{1t}^m)\chi_t + \beta_{3t}^c + \beta_{3t}^m$.

[41] In a slight abuse of notation, $\dot{\mathbf{z}}_t = F(\mathbf{z}_t)$ assumes that the ODEs have been stated in only one direction.

is a compact subspace of $\mathcal{C}$. In particular, for all $\rho, K > 0$, define $\Gamma(\rho + K) \subset \mathcal{C}$ as the space of uniformly Lipschitz continuous functions $\gamma : [0, T] \to [0, \gamma^o]$ with uniform Lipschitz constant $(\gamma^o)^2(2[\rho + K])^2/\sigma_Y^2$ that satisfy $\gamma_0 = \gamma^o$. Likewise, let $X(\rho + K) \subset \mathcal{C}$ denote the space of Lipschitz continuous functions $\chi : [0, T] \to [0, 1]$ with uniform Lipschitz constant $\gamma^o\left[(2[\rho + K])^2/\sigma_Y^2 + (|\hat{u}_{\hat{a}\theta}| + |\hat{u}_{a\hat{a}}|2[\rho + K])^2/\sigma_X^2\right]$ that satisfy $\chi_0 = 0$. Thus, the product $\Lambda(\rho + K) := \Gamma(\rho + K) \times X(\rho + K)$ is a compact subspace of $\mathcal{C}^2$.

We note that the these Lipschitz constant are motivated by a bounding exercise of the γ and χ ODEs that uses $|\beta_i^c| < K$ and $|\beta_i^m| < \rho$, implying that $|\beta_i| < \rho + K$, $i = 1, 3$. Below, we shall construct horizons over which any solution satisfies this property.

Step 2: *Given $(\gamma, \chi) \in \Lambda(\rho + K)$, define a backward initial value problem (IVP) for $(\vec{\beta}^c, \tilde{v}_6, \tilde{v}_8, \tilde{\alpha})$, and establish sufficient conditions for this IVP to have a unique solution.* For any function x, let us use $\hat{x}_{(\cdot)} := x_{T-(\cdot)}$ to emphasize the time-reversed version of x whenever convenient (not to be confused with the hat notation used in the main body). Given any $\lambda \in \Lambda(\rho + K)$, where $(\rho, K) \in \mathbb{R}^2_{++}$, we can define a (backward) IVP consisting of the ODEs for $(\vec{\beta}^c, \tilde{v}_6, \tilde{v}_8, \tilde{\alpha})$ previously stated, but where $\hat{\lambda}$ is used in place of the solutions of the learning ODEs. We write this problem as

$$\dot{\mathbf{b}}_t = \mathrm{f}^{\hat{\lambda}}(\mathbf{b}_t, t) \ \text{ s.t. } \ \mathbf{b}_0 = (0, 0, 0, 0, 0, \alpha^m(\hat{\lambda}_0)), \tag{IVP$^{\text{bwd}}(\hat{\lambda})$}$$

where the use of boldface distinguishes solutions to this IVP from those of our original BVP. We write $\mathbf{b}(\cdot; \lambda)$ for the solution as a functional of the input λ. The extra dependence on time in the right hand side of (IVP$^{\text{bwd}}(\hat{\lambda})$) is due to the role of λ in the system.

For all $\lambda_t \in [0, \gamma^o] \times [0, 1]$, let $\mathrm{B}(\lambda_t) := (\beta_1^m(\lambda_t), \tilde{\beta}_2^m(\lambda_t), \beta_3^m(\lambda_t), 0, 0, \alpha^m(\lambda_t))$. From here, we define $\rho := \sup_{\lambda_t \in [0,\gamma^o]\times[0,1]} ||\mathrm{B}_{-6}(\lambda_t)||_\infty > 0$, with B_{-i} denoting as usual the vector B excluding B_i.[42] For arbitrary $K > 0$, we now establish sufficient conditions for (IVP$^{\text{bwd}}(\hat{\lambda})$) to have a unique solution for each $\lambda \in \Lambda(\rho + K)$.

Lemma C.2. *Fix $\gamma^o, K > 0$. There exists a threshold $T(\gamma^o; K) > 0$ such that if $T < T(\gamma^o; K)$, then for all $\lambda \in \Lambda(\rho + K)$, a unique solution $\mathbf{b}(\cdot; \lambda)$ to (IVP$^{\text{bwd}}(\hat{\lambda})$) exists over $[0, T]$ and satisfies $||\mathbf{b}_i(\cdot; \lambda)||_\infty < K$ for all $i \in \{1, \ldots, 5\}$. Moreover, $T(\gamma^o; K) \in \Omega(1/\gamma^o)$.*

Proof. Fix any $\lambda \in \Lambda(\rho + K)$. Since λ is continuous in t and $\mathrm{f}^{\hat{\lambda}}$ is of class C^1 with respect to $\mathbf{b}_t$, $\mathrm{f}^{\hat{\lambda}}$ is locally Lipschitz continuous in $\mathbf{b}_t$, uniformly in t. By Peano's theorem, a local solution exists; and by the Picard-Lindelöf theorem, solutions are unique given existence. Given $K > 0$, we now construct $T(\gamma^o; K)$ such that a solution exists over $[0, T]$ and satisfies

[42] We exclude $\tilde{\alpha}$ from the definition of ρ because it does not enter the ODEs for the learning coefficients explicitly, and hence it does not affect the definition of $\Lambda(\rho + K)$.

$||\mathbf{b}_i(\cdot;\lambda)||_\infty < K$ for $i \in \{1,\dots,5\}$.

We state two facts that hold over any interval of existence. First, using the ODEs adapted from Lemma C.1 (using $\tilde{\alpha}$ instead of α in the r terms), we have for $i \in \{1,2,3\}$ and $j \in \{4,5\}$

$$\mathbf{b}_{it} = \int_0^t e^{-r\int_s^t \frac{\tilde{\alpha}_u}{\alpha_u^m}du}\hat{\gamma}_s h_i(\mathbf{b}_s,\hat{\chi}_s)ds \ \text{ and } \ \mathbf{b}_{jt} = \int_0^t e^{-\int_s^t (r+\hat{\gamma}_u R_j(\mathbf{b}_u,\hat{\chi}_u))du}\hat{\gamma}_s h_j(\mathbf{b}_s,\hat{\chi}_s)ds.$$

Here, h_i and h_j include the denominators that were factored out of h_x and h_y in Lemma C.1, and do not contain $\tilde{\alpha}$; R_j is only a relabeling of R_y from the same lemma. Second, as long as the conjectured bounds $|\mathbf{b}_{it}| < K$ for $i \in \{1,2,\dots,5\}$ hold, a direct bounding exercise on h_i that uses $\chi_t \in [0,1]$ yields the existence of a scalar $h_i(K)$ such that $|\hat{\gamma}_s h_i(\mathbf{b}_s,\hat{\chi}_s)| \le \gamma^o h_i(K)$, $i \in \{1,2,\dots,5\}$, where we have used that $\gamma_t \in [0,\gamma^o]$ at all times.

Equipped with the equations above for $\mathbf{b}_i$ and with $h_i(K)$, $i \in \{1,\dots,5\}$, notice that the bound $|\mathbf{b}_{it}| < K$ clearly holds for small t. And as long as it holds, $\tilde{\alpha}$ is finite because $\mathbf{b}_{6t}$ has the form $\alpha_0^m e^{\int_0^t G_s ds}$ with $|G_s| < +\infty$ as the latter depends only on $(\mathbf{b}_{-6},\hat{\chi})$ at time $s \in [0,t]$. Moreover, $\tilde{\alpha}/\alpha_t^m > 0$ (see '**Auxiliary Variable**'). Thus, for $i \in \{1,2,3\}$ and $j \in \{4,5\}$,

$$\begin{aligned}
|\mathbf{b}_{it}| &\le \int_0^t e^{-r\int_s^t \frac{\tilde{\alpha}_u}{\alpha_u^m}du}\gamma^o h_i(K)ds \le \int_0^t \gamma^o h_i(K)ds = t\gamma^o h_i(K)\\
|\mathbf{b}_{jt}| &\le \int_0^t e^{-\int_s^t (r+\hat{\gamma}_u R_j(\mathbf{b}_u,\hat{\chi}_u))du}\gamma^o h_j(K)ds \le \int_0^t \gamma^o h_j(K)ds = t\gamma^o h_j(K),
\end{aligned}$$

where we have used that the exponential term is less than 1. Imposing that the right-hand sides above are themselves smaller than K leads us to $T(\gamma^o;K) := \min_{i\in\{1,\dots,5\}} \frac{K}{\gamma^o h_i(K)} > 0$ such that (IVP$^{\text{bwd}}(\hat{\lambda})$) with $T < T(\gamma^o;K)$ by construction admits a unique solution satisfying $|\mathbf{b}_{-6}| < K$ for all $\lambda \in \Lambda(\rho+K)$. Moreover, since $T(\gamma^o;K)$ is independent of r, the statement holds for all $r \ge 0$; also $T(\gamma^o;K) \in \Omega(1/\gamma^o)$.[43] $\square$

In what follows, assume $T < T(\gamma^o;K)$. Lemma C.2 implies that $\lambda \in \Lambda(\rho+K) \mapsto \mathbf{b}(\cdot;\lambda)$ is a well-defined function linking λ paths to corresponding solutions to the backward IVP. We can then define the functional

$$q(\lambda) := (\hat{\mathbf{b}}_1(\cdot;\lambda),\hat{\mathbf{b}}_3(\cdot;\lambda)) + (\mathrm{B}_1(\lambda_{(\cdot)}),\mathrm{B}_3(\lambda_{(\cdot)}))$$

that for each λ delivers the induced "total" 'β_1' and 'β_3' forward-looking coefficients—the centered components delivered by the previous IVP plus the myopic counterparts—that we will use as an input in the learning ODEs below. (Clearly, each $q(\lambda)$ function is a continuous

[43] It is clear from the argument that $\tilde{\alpha}$ is also uniformly bounded for all $\lambda \in \Lambda(\rho+K)$. Also, the linearity of the $\tilde{\alpha}$-ODE (C.6) implies that the interval of existence is constrained only by the ODEs for $\mathbf{b}_i$, $i \in \{1,\dots,5\}$.

function of time.) The continuity of this functional is key for our fixed-point argument.

Step 3: *The operator $\lambda \mapsto q(\lambda)$ is continuous and $||q(\lambda)||_\infty < \rho + K$ for all $\lambda \in \Lambda(\rho + K)$.* Let us show, more generally, that $\lambda \mapsto \hat{\mathbf{b}}(\cdot;\lambda)$ is continuous; since $\lambda \mapsto \mathrm{B}_i(\lambda_{(\cdot)})$ is clearly continuous due to $\beta_i^m = \beta_i^m(\chi_{(\cdot)})$ being of class C^1, $i \in \{1, 3\}$, the result will follow. To this end, we make use of the following lemma, proved in the Supplementary Appendix.

Lemma C.3. *Let $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$ and $U \subseteq \mathbb{R}^n$ be compact sets. Consider $F : X \times Y \to U$ of class C^1 and $\omega : Y \to X$. Suppose $\mathcal{Y} \subset C([0, T]; Y)$ is a collection of functions such that for all $y \in \mathcal{Y}$, the initial value problem IVP(y) defined by $\dot{x}_t = F(x_t, y_t)$ and $x_0 = \omega(y_0)$ admits a solution defined over $[0, T]$. Then there exist constants k_1 and k_2 (depending on T) such that for all $y^1, y^2 \in \mathcal{Y}$, the corresponding solutions x^i to IVP(y^i) satisfy*

$$||x_t^1 - x_t^2||_\infty \leq k_1||\omega(y_0^1) - \omega(y_0^2)||_\infty + k_2 \sup_{s\in[0,T]} ||y_s^1 - y_s^2||_\infty, \qquad \textit{for all } t \in [0, T].$$

Now consider any $\lambda^1, \lambda^2 \in \Lambda(\rho + K)$. We apply Lemma C.3 to: $x = \mathbf{b}$; $y^i = \hat{\lambda}^i$, $i = 1, 2$; $\omega(\cdot) = (0, 0, 0, 0, 0, \alpha^m(\cdot))$; $F(x_t, y_t) := \mathrm{f}^{\hat{\lambda}}(\mathbf{b}_t, t)$; and X and Y the hypercubes defined by the uniform bounds on $\mathbf{b}$ and λ, respectively. Using that $||x||_\infty = ||\hat{x}||_\infty$, we obtain

$$||\hat{\mathbf{b}}(\cdot;\lambda^1) - \hat{\mathbf{b}}(\cdot;\lambda^2)||_\infty = \sup_{t\in[0,T]} ||\mathbf{b}_t(\lambda^1) - \mathbf{b}_t(\lambda^2)||_\infty \leq k_1|\alpha^m(\lambda_T^1) - \alpha^m(\lambda_T^2)| + k_2||\lambda^1 - \lambda^2||_\infty,$$

for some constants k_1 and k_2. Since $\lambda_T \mapsto \alpha^m(\lambda_T)$ is continuous, it follows that $||\hat{\mathbf{b}}(\cdot;\lambda^1) - \hat{\mathbf{b}}(\cdot;\lambda^2)||_\infty \to 0$ as $||\lambda^1 - \lambda^2||_\infty \to 0$, yielding the desired result.

Finally, $||q(\lambda)||_\infty < \rho + K$ follows from $||\hat{\mathbf{b}}_i(\cdot;\lambda)||_\infty < K$ and $||\mathrm{B}_i(\lambda_T)||_\infty < \rho$, $i = 1, 3$.

Step 4: *Construct a continuous self-map on $\Lambda(\rho+K)$ using the IVP for the learning ODEs.* Take $\lambda \in \Lambda(\rho + K)$ and define the IVP for $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$

$$\dot{\boldsymbol{\lambda}}_t = \mathrm{f}^{q(\lambda)}(\boldsymbol{\lambda}_t, t) \;\text{ s.t. }\; \boldsymbol{\lambda}_0 = (\gamma^o, 0), \qquad (\text{IVP}^{\text{fwd}}(q(\lambda)))$$

consisting of the two (forward) learning ODEs (10)-(11) that use as input $q(\lambda) = (q_1(\lambda), q_2(\lambda))$ playing the role of (β_1, β_3)—here, the first (second) entry of the system corresponds to the γ-ODE (χ-ODE), while the boldface convention aims at distinguishing between inputs λ via q and induced solutions $\boldsymbol{\lambda}$ to this IVP. Importantly, because for all $\lambda \in \Lambda(\rho + K)$ the function $q(\lambda)$ is continuous in time, Lemma A.1 gives existence and uniqueness of a solution to (IVP$^{\text{fwd}}$($q(\lambda)$)) defined over $[0, T]$ that satisfies $\boldsymbol{\lambda}_t \in (0, \gamma^o] \times [0, 1)$ for all such times.

Next, we argue that $\boldsymbol{\lambda} \in \Lambda(\rho + K)$. By construction, $\boldsymbol{\lambda}_0 := (\boldsymbol{\lambda}_{1,0}, \boldsymbol{\lambda}_{2,0}) = (\gamma^o, 0)$, and as noted above, $\boldsymbol{\lambda}_t \in (0, \gamma^o] \times [0, 1)$ for all $t \in [0, T]$. Moreover, from the γ-ODE and χ-ODE, we have that

$$|\dot{\boldsymbol{\lambda}}_{1t}| = |-\frac{\boldsymbol{\lambda}_{1t}^2([q_2(\lambda)]_t + [q_1(\lambda)]_t\boldsymbol{\lambda}_{2t})^2}{\sigma_Y^2}| \leq (\gamma^o)^2(2[\rho+K])^2/\sigma_Y^2 \qquad \text{and similarly}$$
$$|\dot{\boldsymbol{\lambda}}_{2t}| \leq \gamma^o\left[(2[\rho+K])^2/\sigma_Y^2 + (|\hat{u}_{\hat{a}\theta}| + |\hat{u}_{a\hat{a}}|(2[\rho+K]))^2/\sigma_X^2\right]$$

for all $t \in [0,T]$. Since the Lipschitz bounds in the definition of $\Lambda(\rho+K)$ are satisfied, $\boldsymbol{\lambda} \in \Lambda(\rho+K)$.

Finally, by Lemma C.3 applied to (IVP$^{\text{fwd}}$($q(\lambda)$)) by setting $x = \boldsymbol{\lambda}$, $y = q(\lambda)$, $\omega(y_0) = (\gamma^o, 0)$, $F(x_t, y_t) = \mathrm{f}^{q(\lambda)}(\boldsymbol{\lambda}_t, t)$, $X = [0,\gamma^o]\times[0,1]$ and $Y = [-\rho-K, \rho+K]^2$, we conclude that $q \mapsto \boldsymbol{\lambda}(q)$ is continuous. Since $\lambda \mapsto q(\lambda)$ is continuous (Step 3), it follows that $g(\lambda) := \boldsymbol{\lambda}(q(\lambda))$ is a continuous map from $\Lambda(\rho+K)$ to itself.

Step 5: *Show that g has a fixed point.* By Step 1, $\Lambda(\rho+K)$ is a nonempty, compact, convex Banach space, and by Step 4, g is a continuous map from $\Lambda(\rho+K)$ to itself. By Schauder's Theorem (Zeidler, 1986, Corollary 2.13), there exists $\lambda^* \in \Lambda(\rho+K)$ such that $\lambda^* = g(\lambda^*)$. It is clear, by construction, that $(\lambda^*, \hat{\mathbf{b}}(\cdot;\lambda^*))$, with $\mathbf{b}(\cdot;\lambda^*)$ the solution to (IVP$^{\text{bwd}}$($\hat{\lambda}$)) under $\lambda = \lambda^*$, is a solution to the centered-augmented BVP under study. Finally, maximizing $T(\gamma^o;K)$ over $K>0$ yields a $T(\gamma^o)>0$ that has the form C/γ^o. □

Verification: *Recover first a solution to the original BVP, and then to the full HJB equation.* We verify that the solution to the centered-augmented BVP induces a solution to the original BVP stated in the '**Core subsystem**' section. To do this, we first note that any solution to the former BVP must satisfy the identity $\tilde{\alpha} \equiv \alpha$, where $\alpha_t := \beta_{1t}\chi_t + \beta_{3t}$, $\beta_{1t} := \beta_{1t}^c + \beta_{1t}^m$ and $\beta_{3t} := \beta_{1t}^c + \beta_{1t}^m$—consequently, $(\gamma, \chi, \vec{\beta}^c, \tilde{v}_6, \tilde{v}_8)$ solves the centered system defined in the '**Centering**' step. Indeed, using the definition of the myopic coefficients as well as the ODEs for χ, β_{1t}^c, and β_{3t}^c yields that α in backward form satisfies

$$\dot{\alpha}_t = -r\tilde{\alpha}_t(\alpha_t/\alpha_t^m - 1) + \alpha_t\frac{\gamma_t h_\alpha(\vec{\beta}^c, \tilde{v}_6, \tilde{v}_8, \chi_t)}{\sigma_X^2\sigma_Y^2(u_{a\theta}+u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)^{n_{1,\alpha}}(1-u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t)^{n_{2,\alpha}}(1-u_{a\hat{a}}\hat{u}_{a\hat{a}})^{n_{3,\alpha}}}.$$

Relative to (C.6), therefore, the r-term as well as the last fraction multiplying α coincide. Call this last term H_t—a continuous function of time—and observe that $p := \alpha - \tilde{\alpha}$ satisfies the ODE $\dot{p}_t = p_t H_t$ with initial condition $p_0 = 0$ due to $\alpha_0 = \tilde{\alpha}_0 = \alpha_0^m$ (recall that time is being reversed). By uniqueness, $p_t \equiv 0$ for all $t \in [0,T]$, confirming that $\alpha \equiv \tilde{\alpha}$.

Given this equivalence, it follows that $(\gamma, \chi, \beta_1, \tilde{\beta}_2, \beta_3, \tilde{v}_6, \tilde{v}_8) = (\lambda^*, \hat{\mathbf{b}}_{-6}(\cdot;\lambda^*) + \mathrm{B}_{-6}(\lambda^*))$ solves by construction the BVP stated in the '**Core subsystem**' section. Moreover, as argued in Step 4 in the proof of Theorem C.1, $\gamma > 0$ and $\chi < 1$, so we can invert the change of variables $(\tilde{\beta}_2, \tilde{v}_6, \tilde{v}_8) = (\beta_2/(1-\chi), v_6\gamma/(1-\chi)^2, v_8\gamma/(1-\chi))$ to obtain (β_2, v_6, v_8). And since $\alpha = \tilde{\alpha}$ never vanishes (see '**Auxiliary variable**' section) and $\gamma > 0$, we can recover the rest of the coefficients in the value function as explained in the same section.

We extend our existence result to the case of terminal payoffs in the following corollary, proved in the Supplementary Appendix. The bound on curvature ensures that we can select an equilibrium of the static terminal game with sufficient regularity for our method.

Corollary C.2. *There exist $C_\psi \in \{-\infty\} \cup (-\infty, 0)$ and $C_T > 0$, both independent of (r, γ^o), such that if $\psi_{\hat{a}\hat{a}} \in (C_\psi/\gamma^o, 0]$ and $T < C/\gamma^o$, a linear Markov equilibrium exists for all $r \geq 0$. Moreover, α_3 never vanishes.*

Appendix D: Proofs for Section 6

Before proving Proposition 8, we present the laws of motion for $(\hat{M}, L)$ in the receiver's best response problem. These are obtained from (A.1) and (12), using $\hat{a}'_t$ in dX_t.

Lemma D.1 (Controlled dynamics: receiver)**.** *From the receiver's perspective, if he follows $(\hat{a}'_t)_{t\in[0,T]}$,*

$$d\hat{M}_t = \frac{\alpha_{3t}\gamma_{1t}}{\sigma_Y} dZ_t \tag{D.1}$$

$$dL_t = \frac{\gamma_t^X \chi_t \delta_{1t}}{\sigma_X^2}[\hat{a}'_t - (\delta_{0t} + [\delta_{1t} + \delta_{2t}]L_t) + \sigma_X dZ_t^X], \tag{D.2}$$

where $Z_t := \frac{1}{\sigma_Y}[Y_t - \int_0^t (\alpha_{0s} + \alpha_{2s}L_s + \alpha_{3s}\hat{M}_s)ds]$ is a Brownian motion.

Proof of Proposition 8. It suffices to show that the myopic policy is optimal for the receiver's best response problem when the receiver is forward looking. After all, in an LME of our baseline model with myopic receiver, the sender's strategy is already (by definition) a best response to the myopic strategy of the receiver, and the learning variables (γ, χ) are already consistent with these strategies.

To show this, consider the receiver's HJB equation, which given the laws of motion for $(\hat{M}, L)$ is

$$\hat{r}\hat{V} = \sup_{\hat{a}'} \left\{ -\frac{1}{2}(c_0 + c_1\hat{m} + c_2[\alpha_{0t} + \alpha_{3t}\hat{m} + \alpha_{2t}\ell] - \hat{a}')^2 - \frac{1}{2}(c_1 + c_2\alpha_{3t})^2\gamma_t + \hat{V}_t \right. \tag{D.3}$$

$$\left. + \mu_{\hat{M}}\hat{V}_{\hat{m}} + \mu_L(\hat{a}')\hat{V}_\ell + \frac{\sigma_{\hat{M}}^2}{2}\hat{V}_{\hat{m}\hat{m}} + \frac{\sigma_L^2}{2}\hat{V}_{\ell\ell} \right\}, \tag{D.4}$$

where $\mu_{\hat{M}}$ $(= 0)$, $\sigma_{\hat{M}}$, $\mu_L(\hat{a}')$, and σ_L are the drift and noise in (D.1) and drift and noise in (D.2), respectively. There is no $\hat{V}_{\hat{m}\ell}$ term since the innovations in $\hat{M}$ and L are uncorrelated (i.e., they have zero quadratic covariation). The value function must also satisfy

$$\hat{V}(\hat{m},\ell,T)=\sup_{\hat{a}'}\frac{1}{2}\left\{-(c_0+c_1\hat{m}+c_2[\alpha_{0T}+\alpha_{3T}\hat{m}+\alpha_{2T}\ell]-\hat{a}')^2-(c_1+c_2\alpha_{3T})^2\gamma_T\right\}. \tag{D.5}$$

Now let $(\delta^m_{0t},\delta^m_{1t},\delta^m_{2t})=(c_0+c_2\alpha_{0t},c_1+c_2\alpha_{3t},c_2\alpha_{2t})$ denote the myopic strategy coefficients and $\hat{a}^m_t=\delta^m_{0t}+\delta^m_{1t}\hat{m}+\delta^m_{2t}\ell$ the myopic policy. It is easy to see that $\hat{a}^m_T$ attains the supremum in (D.5) and the first quadratic term vanishes, so (D.5) yields the terminal condition

$$\hat{V}(\hat{m},\ell,T)=-\frac{1}{2}(c_1+c_2\alpha_{3T})^2\gamma_T. \tag{D.6}$$

Note that this terminal payoff is independent of $(\hat{m},\ell)$. In the same spirit, we conjecture a solution to the HJB where the value function depends only on time; for such a solution, the HJB equation (D.3) reduces to

$$\hat{r}\hat{V}=\sup_{\hat{a}'}\left\{-\frac{1}{2}(c_0+c_1\hat{m}+c_2[\alpha_{0t}+\alpha_{3t}\hat{m}+\alpha_{2t}\ell]-\hat{a}')^2-\frac{1}{2}(c_1+c_2\alpha_{3t})^2\gamma_t+\hat{V}_t\right\}.$$

It is easy to see that for all $t<T$, the right hand side is maximized at the myopic policy $\hat{a}_t$, at which point the first quadratic loss term vanishes, so the HJB equation further reduces to

$$\hat{r}\hat{V}=-\frac{1}{2}(c_1+c_2\alpha_{3t})^2\gamma_t+\hat{V}_t. \tag{D.7}$$

Simple integration using (D.6) and (D.7) yields the solution

$$\hat{V}(t)=-\frac{1}{2}\int_t^T e^{-\hat{r}(s-t)}(c_1+c_2\alpha_{3s})^2\gamma_s\,dt-\frac{1}{2}e^{-\hat{r}(T-t)}(c_1+c_2\alpha_{3T})^2\gamma_T, \tag{D.8}$$

which is indeed a function only of time. We conclude that the myopic policy is optimal. □

Proof of Proposition 9. We first derive a candidate mapping. Suppose $\delta_1=\hat{u}_{a\hat{a}}\alpha_3$. The χ-ODE boils down to

$$\dot{\chi}_t=\gamma_t\alpha_{3t}^2\left(\frac{1-\chi_t}{\sigma_Y^2}-\frac{(\hat{u}_{a\hat{a}}\chi_t)^2}{\sigma_X^2}\right)=:-\gamma_t\alpha_{3t}^2Q(\chi_t).$$

If $f:[0,\bar{\chi})\to[0,\gamma^o]$, some $\bar{\chi}\in(0,1]$, is differentiable and $f(\chi_t)=\gamma_t$ for all $t\geq 0$, then $f'(\chi_t)\dot{\chi}_t=\dot{\gamma}_t$. When $\alpha_{3t}\neq 0$, $\frac{f'(\chi_t)}{f(\chi_t)}=\frac{\Sigma}{Q(\chi_t)}$. Hence, we solve the ODE $\frac{f'(\chi)}{f(\chi)}=\frac{\Sigma}{Q(\chi)}$ for $\chi\in(0,\bar{\chi})$ where $f(0)=\gamma^o$.

To this end, let $c_2:=\frac{\sqrt{1/\sigma_Y^4+4(\hat{u}_{a\hat{a}})^2/[\sigma_X\sigma_Y]^2}-1/\sigma_Y^2}{2(\hat{u}_{a\hat{a}}/\sigma_X)^2}$ and $-c_1:=\frac{-\sqrt{1/\sigma_Y^4+4(\hat{u}_{a\hat{a}})^2/[\sigma_X\sigma_Y]^2}-1/\sigma_Y^2}{2(\hat{u}_{a\hat{a}}/\sigma_X)^2}$

be the roots of the convex quadratic Q above. Note that these are well-defined since $\hat{u}_{\hat{a}\theta}$ and Assumption 1 part (ii) imply that $\hat{u}_{a\hat{a}} \neq 0$.

Clearly, $-c_1 < 0 < c_2$. Also, $c_2 \le 1$ as $Q(1) \ge 0$. Thus, $\frac{\Sigma}{Q(\chi)} = -\frac{\sigma_X^2 \Sigma}{(\hat{u}_{a\hat{a}})^2(c_1+c_2)}\left[\frac{1}{\chi+c_1} - \frac{1}{\chi-c_2}\right]$ is well defined (and negative) over $[0, c_2)$ with $1/(\chi + c_1) > 0$ and $-1/(\chi - c_2) > 0$ over the same domain. We can then set $\bar{\chi} = c_2$ and solve $\int_0^\chi \frac{f'(s)}{f(s)} ds = -\frac{\sigma_X^2 \Sigma}{(\hat{u}_{a\hat{a}})^2(c_1+c_2)} \log\left(\frac{\chi+c_1}{c_2-\chi}\frac{c_2}{c_1}\right)$, which yields the decreasing function $f(\chi) = f(0)\left(\frac{c_1}{c_2}\right)^{1/d}\left(\frac{c_2-\chi}{\chi+c_1}\right)^{1/d}$, where $1/d = \sigma_X^2\Sigma/[(\hat{u}_{a\hat{a}})^2(c_1+c_2)] > 0$. Imposing $f(0) = \gamma^o$ and inverting yields $\chi(\gamma) = f^{-1}(\gamma)$ as given in the lemma. Note that $\chi(\gamma^o) = 0$ and $\chi(0) = c_2$.

We now verify that $\chi(\gamma)$ satisfies the χ-ODE (even when $\alpha_3 = 0$). We have

$$\frac{d(\chi(\gamma_t))}{dt} = \frac{\alpha_{3t}^2 \gamma_t}{\sigma_Y^2[c_1 + c_2(\gamma/\gamma^o)^d]^2} c_1 c_2 d[c_1 + c_2]\left(\frac{\gamma_t}{\gamma^o}\right)^d.$$

By construction, moreover, $c_1 c_2 = c_1 - c_2 = \frac{\sigma_X^2}{\sigma_Y^2(\hat{u}_{a\hat{a}})^2}$, which follows from equating the first- and zero-order coefficients in $Q(\chi) = \hat{u}_{a\hat{a}}^2\chi^2/\sigma_X^2 + \chi/\sigma_Y^2 - 1/\sigma_Y^2 = \hat{u}_{a\hat{a}}^2(\chi - c_2)(\chi + c_1)/\sigma_X^2$. Thus, $dc_1c_2 = c_1 + c_2$. On the other hand,

$$\frac{[\hat{u}_{a\hat{a}}\chi(\gamma)]^2}{\sigma_X^2} = \frac{\hat{u}_{a\hat{a}}^2}{\sigma_X^2}\left[c_1c_2\frac{1-(\gamma/\gamma^o)^d}{c_1 + c_2(\gamma/\gamma^o)^d}\right]^2 = \frac{c_1^2(1-c_2)}{\sigma_Y^2}\left[\frac{1-(\gamma/\gamma^o)^d}{c_1 + c_2(\gamma/\gamma^o)^d}\right]^2$$

where we used that $c_1^2c_2^2/\sigma_X^2 = c_1^2(1-c_2)/\sigma_Y^2$ follows from $\hat{u}_{a\hat{a}}^2 c_2^2/\sigma_X^2 = (1-c_2)/\sigma_Y^2$ by definition of c_2. Thus, the right-hand side of the χ-ODE evaluated at our candidate $\chi(\gamma)$ satisfies

$$\gamma_1\alpha_3^2\left(\frac{1-\chi}{\sigma_Y^2} - \frac{(\hat{u}_{a\hat{a}}\chi)^2}{\sigma_X^2}\right)\Bigg|_{\chi=\chi(\gamma)} = \frac{\alpha_3^2\gamma_1}{\sigma_Y^2}\left(1 - \chi - c_1^2(1-c_2)\left[\frac{1-(\gamma/\gamma^o)^d}{c_1 + c_2(\gamma/\gamma^o)^d}\right]^2\right).$$

Thus, using that $c_1c_2d = c_1 + c_2$ in our expression for $d(\chi(\gamma_t))/dt$, it suffices to show that

$$[c_1 + c_2]^2\left(\frac{\gamma_t}{\gamma^o}\right)^d = (1-\chi)[c_1 + c_2(\gamma/\gamma^o)^d]^2 - c_1^2(1-c_2)[1-(\gamma/\gamma^o)^d]^2.$$

Using that $\chi[c_1 + c_2(\gamma/\gamma^o)^d] = 1 - (\gamma/\gamma^o)$, it is easy to conclude that this equality reduces to three equations $0 = c_1^2 - c_1^2c_2 - c_1^2 + c_1^2c_2$, $(c_1 + c_2)^2 = 2c_1c_2 - c_1c_2(c_2 - c_1) + 2c_1^2(1-c_2)$ and $0 = c_2^2 + c_1c_2^2 - c_1^2(1-c_2)$, capturing the conditions on the constant, $(\gamma/\gamma^o)^d$ and $(\gamma/\gamma^o)^{2d}$, respectively. The first condition is trivially satisfied, and the third is easy to verify; by canceling common terms, the second condition is also a rearrangement of this identity. Thus, $\chi(\gamma)$ as postulated satisfies the χ-ODE; by uniqueness, $\chi = \chi(\gamma)$.

We now prove the final statement of the lemma. When $\gamma_t \in (0, \gamma^o]$, we have $\chi_t =$

$\frac{c_1 c_2(1-[\gamma_t/\gamma^o]^d)}{c_1+c_2[\gamma_t/\gamma^o]^d} < \frac{c_1 c_2}{c_1} = c_2$. Now c_2 simplifies to $\frac{\sqrt{\sigma_X^4+4\sigma_Y^2\sigma_X^2\hat{u}_{a\hat{a}}^2}-\sigma_X^2}{2\hat{u}_{a\hat{a}}^2\sigma_Y^2} = \frac{4\sigma_Y^2\hat{u}_{a\hat{a}}^2}{2\hat{u}_{a\hat{a}}^2\sigma_Y^2\left(\sqrt{1+4\sigma_Y^2\hat{u}_{a\hat{a}}^2/\sigma_X^2}+1\right)}$, which by inspection is increasing in σ_X and has limits $\lim_{\sigma_X\to 0} c_2 = 0$ and $\lim_{\sigma_X\to+\infty} c_2 = 1$ $\square$

References

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, 1041–1063.

Back, K., C. H. Cao, and G. A. Willard (2000): "Imperfect competition among informed traders," *The Journal of Finance*, 55, 2117–2155.

Bergemann, D. and P. Strack (2015): "Dynamic revenue maximization: A continuous time approach," *Journal of Economic Theory*, 159, 819–853.

Bhaskar, V. and I. Obara (2002): "Belief-based equilibria in the repeated prisoners' dilemma with private monitoring," *Journal of Economic Theory*, 102, 40–69.

Bonatti, A. and G. Cisternas (2020): "Consumer scores and price discrimination," *The Review of Economic Studies*, 87, 750–791.

Bonatti, A., G. Cisternas, and J. Toikka (2017): "Dynamic oligopoly with incomplete information," *The Review of Economic Studies*, 84, 503–546.

Bouvard, M. and R. Lévy (2019): "Horizontal Reputation and Strategic Audience Management," *Journal of the European Economic Association*.

Carlsson, H. and S. Dasgupta (1997): "Noise-proof equilibria in two-action signaling games," *Journal of Economic Theory*, 77, 432–460.

Cetemen, D. (2020): "Efficiency in Repeated Partnerships," *Available at SSRN 3724923*.

Cetemen, D., G. Cisternas, A. Kolb, and S. Viswanathan (2023): "Activist Trading Dynamics," *FRB of New York Staff Report*.

Cetemen, D. and C. Margaria (2020): "Signaling What You Don't Want," *Available at SSRN 3732674*.

Cisternas, G. (2018): "Two-sided learning and the ratchet principle," *The Review of Economic Studies*, 85, 307–351.

Daley, B. and B. Green (2012): "Waiting for News in the Market for Lemons," *Econometrica*, 80, 1433–1504.

Dessein, W. and T. Santos (2006): "Adaptive Organizations," *Journal of Political Economy*, 114, 956–995.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005): "Over-the-counter markets," *Econometrica*, 73, 1815–1847.

Ely, J. C. and J. Välimäki (2002): "A robust folk theorem for the prisoner's dilemma," *Journal of Economic Theory*, 102, 84–105.

Faingold, E. and Y. Sannikov (2011): "Reputation in continuous-time games," *Econometrica*, 79, 773–876.

Feltovich, N., R. Harbaugh, and T. To (2002): "Too Cool for School? Signalling and Countersignalling," *The RAND Journal of Economics*, 33, 630–649.

Foster, F. D. and S. Viswanathan (1994): "Strategic trading with asymmetrically informed traders and long-lived information," *Journal of Financial and Quantitative Analysis*, 29, 499–518.

——— (1996): "Strategic trading when agents forecast the forecasts of others," *The Journal of Finance*, 51, 1437–1478.

Garicano, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874–904.

Gryglewicz, S. and A. Kolb (2021): "Dynamic Signaling with Stochastic Stakes," *Theoretical Economics*, forthcoming.

Holmström, B. (1999): "Managerial incentive problems: A dynamic perspective," *The review of Economic studies*, 66, 169–182.

Kandori, M. (2002): "Introduction to repeated games with private monitoring," *Journal of economic theory*, 102, 1–15.

Karatzas, I. and S. E. Shreve (1991): *Brownian motion and stochastic calculus*, vol. 113, Springer Science & Business Media.

Kolb, A., M. Pease, D. W. Sacks, and J. Quick (2021): "Blind Disclosure," *Available at SSRN 3765529*.

Kolb, A. M. (2019): "Strategic real options," *Journal of Economic Theory*, 183, 344–383.

Kremer, I. and A. Skrzypacz (2007): "Dynamic signaling and market breakdown," *Journal of Economic Theory*, 133, 58–82.

Kyle, A. S. (1985): "Continuous auctions and insider trading," *Econometrica*, 1315–1335.

Liptser, R. S. and A. Shiryaev (1977): *Statistics of Random Processes 1, 2*, Springer-Verlag, New York.

MacLeod, W. B. (2003): "Optimal contracting with subjective evaluation," *American Economic Review*, 93, 216–240.

Mailath, G. J. and S. Morris (2002): "Repeated games with almost-public monitoring," *Journal of Economic Theory*, 102, 189–228.

Mailath, G. J. and L. Samuelson (2006): *Repeated Games and Reputations*, Oxford University Press.

Marschak, J. and R. Radner (1972): *Economic Theory of Teams*, New Haven, CT: Yale University Press.

Matthews, S. A. and L. J. Mirman (1983): "Equilibrium limit pricing: The effects of private information and stochastic demand," *Econometrica*, 981–996.

Ok, E. A. (2007): *Real analysis with economic applications*, vol. 10, Princeton University Press.

Pham, H. (2009): *Continuous-time stochastic control and optimization with financial applications*, vol. 61, Springer Science & Business Media.

Phelan, C. and A. Skrzypacz (2012): "Beliefs and private monitoring," *Review of Economic Studies*, 79, 1637–1660.

Rantakari, H. (2008): "Governing Adaptation," *Review of Economic Studies*, 75, 1257–1285.

Sannikov, Y. (2007): "Games with imperfectly observable actions in continuous time," *Econometrica*, 75, 1285–1329.

Subrahmanyam, A. (1998): "Transaction taxes and financial market equilibrium," *The Journal of Business*, 71, 81–118.

Svensson, L. E. and M. Woodford (2003): "Indicator variables for optimal policy," *Journal of monetary economics*, 50, 691–720.

Teschl, G. (2012): *Ordinary differential equations and dynamical systems*, vol. 140, American Mathematical Society.

Townsend, R. M. (1983): "Forecasting the forecasts of others," *Journal of Political Economy*, 91, 546–588.

Yang, L. and H. Zhu (2020): "Back-running: Seeking and hiding fundamental information in order flows," *The Review of Financial Studies*, 33, 1484–1533.

Zeidler, E. (1986): *Nonlinear functional analysis vol. 1: Fixed-point theorems*, Springer-Verlag.