Signaling with Private Monitoring*

Gonzalo Cisternas and Aaron Kolb

December 10, 2021

Abstract

We study dynamic signaling when the informed party does not observe the signals generated by her actions. A forward-looking sender signals her type continuously over time to a myopic receiver who privately monitors her behavior; in turn, the receiver transmits his private inferences back through an imperfect public signal of his actions. Preferences are linear-quadratic and the information structure is Gaussian. We construct linear Markov equilibria using belief states up to the sender's *second-order belief*. Because of the private monitoring, this state is an explicit function of the sender's past play, leading to a novel separation effect through the second-order belief channel. Applications to models of organizations and reputation are examined.

*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, Marina Halac, Stephen Morris, Alessandro Pavan, 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.

1 Introduction

This paper introduces a new class of dynamic signaling games featuring private signals. These games can be seen as continuous-time versions of repeated noisy signaling games in which the traditional receiver privately sees imperfect signals of the sender's actions—thus, the receiver endogenously develops a *private belief* about the sender's type. In settings where this form of private information is relevant for the sender's behavior, a complex problem in which players simultaneously form beliefs about each other's beliefs arises. We offer the first tractable treatment of this largely unexplored problem within signaling games.

Incorporating private signals of behavior is an important agenda. In the theory of *organizations*, the literature emanating from the statistical theory of teams of Marschak and Radner (1972) recognizes that decision makers possess private information, albeit exogenous, about the external environment; individuals then attempt to transmit their information, after which they take actions and the interaction typically ends.[1] Organizations, however, are dynamic, which means that now endogenous information linked to actions can be subject to the same information frictions—dispersion, difficult transmission—a topic completely ignored in the analyses of teams. Similarly, virtually all models of *reputation* assume that the party interested in building a reputation is certain about the relevant audience's perception; but this precludes the private possession of imperfect signals of behavior that is relevant in many markets,[2] or natural features such as the subjective interpretation of information.

The scarcity of results in this area is likely related to the technical difficulties encountered when examinining models with these characteristics. First, there is higher-order uncertainty due to the players attempting to "forecast the forecasts of others" (Townsend, 1983). Second, these games are inherently asymmetric: when facing a sender of a *fixed* type, the receiver develops *evolving* private information in the form of a belief—and the latter player can signal information back. Third, most analyses will be nonstationary due to ongoing learning effects.

In this paper, we examine a class of linear-quadratic-Gaussian games of incomplete information and private monitoring in continuous time. A forward-looking sender (she) and a myopic receiver (he), both with quadratic preferences, interact over a finite horizon. The sender has a normally distributed type. Our key innovation is that the receiver privately observes a noisy signal of the sender's action; in turn, the sender gets feedback about the receiver via a public signal of the latter's behavior. Using shocks that are additive and Brownian, we construct linear Markov equilibria (LMEs) with the players' beliefs as states.[3]

[1]See Dessein and Santos (2006) and Rantakari (2008), also relying on quadratic preferences.

[2]E.g., financial markets (Yang and Zhu, 2019) or for consumer data (Bonatti and Cisternas, 2020).

[3]We note that our receiver develops private information, and transmits it back. A forward-looking receiver would have no impact on our construction, methods and applications—see Remark 1 and Section 6.

Equilibrium construction and signaling. It is well known that the construction of nontrivial equilibria in games of private monitoring can be a daunting task. In fact, to estimate rivals' continuation behavior under any strategy, players usually have to make an inference about their opponents' private histories. Not knowing what their rivals have seen, the players will then rely on their past play, but this implies that the players' inferences will vary with their *own* private histories. Thus, (i) probability distributions over histories must be computed, and (ii) the continuation games at off- versus on-path histories may differ.

With incomplete information, one expects this statistical inference problem to become one of the estimation of belief states that summarize the payoff-relevant aspects of the players' private histories—our approach offers a parsimonious treatment of this issue. The quadratic preferences permit equilibria in which players' strategies are linear in their posterior beliefs' means (henceforth, beliefs). Conjecturing such strategies, learning is (conditionally) Gaussian: the receiver's belief is linear in the history of his private signals, and the sender's *second-order belief*—her belief about the receiver's private belief—is linear in the histories of the public signal and her *past play*. The estimation of histories described in (i) is thus simplified by the fact that these are aggregated linearly.

The sender's second-order belief is also private, as her actions depend on her type; the receiver must therefore forecast this state. The problem of the state space expanding is circumvented by a key *representation* of the second-order belief under linear Markov strategies in terms of the sender's type and the belief about it using the public signal exclusively (Lemma 1). Performing equilibrium analysis then requires a second-order belief that is only spanned by the other states along the path of play ((ii) above). With Markov states, the sender's best-response problem is one of stochastic control and we use dynamic programming.

A key property is that the sender controls her own second-order belief, generalizing the traditional control of a public belief under imperfect public monitoring. But since this state is now an explicit function of past play, private monitoring has novel implications for signaling. Specifically, since different types behave differently in equilibrium, their past behavior leads them to expect the receiver to hold different beliefs. This perception of different continuation games across types then opens an additional channel for separation, and is encoded in the value that the second-order belief takes in the representation.

We refer to the signaling implications of the previous separation effect as the *history-inference effect* on signaling. Importantly, this effect will be at play whenever an ex ante informed party does not see all the signals of her behavior — from this perspective, settings in which beliefs are public are clearly an exception. Our applications then intend to illustrate how traditional logic and forces derived in such public settings are, via this effect, altered by the presence of higher-order uncertainty.

Applications. In the class of games studied, different sender types may have different incentives depending on where the belief of the follower stands: *revelation* motives, when the sender wants to direct the follower's belief towards the type; and *concealment motives*, influencing the belief away from it. Our applications intentionally isolate each motive.

In Section 4.1, we analyze a novel coordination game that we use to explore the connection between learning and performance in organizations when information frictions are present. In our setting, a team's performance increases with the proximity of the leader's actions to both a state of the world (adaptation) and to a follower's action (coordination). The leader maximizes the team's performance while the follower tries to coordinate, resulting in incentives being aligned. Thus, the leader wishes to reveal the state of the world to the follower, but the follower's learning is imperfect—hence, gradual—and private. The players then engage in natural guessing of the other's understanding as play unfolds.

In this context, when there is higher-order uncertainty, the history-inference effect can lead to more information being transmitted relative to the public case in which the follower's belief is commonly known. Yet, the team's performance is lower. Thus, organizations with worse information channels can exhibit a better understanding of their economic environments, as measured by the follower's terminal learning, and learning need not reflect performance. In fact, as we show, learning is a measure of coordination costs in an organization.

At the other extreme, Section 4.2 examines concealment motives in a new model of reputation where the informed party does not know her reputation with certainty. The sender is a politician or regulator who finds it costly to take actions away from her *bias*—the type—on a relevant issue. The concealment motive arises in that types would like to be perceived as neutral at a terminal time (e.g., a reappointment): the sender suffers a terminal quadratic loss in the distance between the belief of a news outlet (the receiver) and the type's prior (capturing the unbiased type). The politician receives feedback regarding her reputation from the outlet's reporting; naturally, the latter is public and imperfect.

Clearly, the direct effect of more precise public feedback is that it allows the politician to better tailor her actions to her reputation. With higher-order uncertainty, however, there is also a strategic effect. Specifically, since higher types take higher actions due to their higher biases, those types will perceive their reputation to be more upward biased. Because higher types will then attempt to offset higher beliefs, the history-inference effect reduces the informativeness of the politician's action, effectively enabling her to better conceal her true type—having access to *worse* feedback can thus increase her payoffs.

Finally, a common element of both applications is a non-monotonic signaling coefficient due to the history-inference at play, which we link to predictions of behavior in each case.

Existence of LME and technical contribution. The games studied are *asymmetric*, both in terms of the players' preferences and their private information (a fixed state versus a changing one). Thus, the players can signal at different rates. As we explain next, this issue is a major hurdle for showing the existence of a LME, which we address in Section 5.

Specifically, due to the Gaussian structure, the belief states that we employ are stochastic posterior means that are also coupled with two deterministic second moments: the receiver's posterior variance (shaping the sensitivity of the receiver's belief to his private signals), and the weight of the sender's type in the representation (linked to the sender's learning, shaping the history-inference effect). Using dynamic programming, one can transform the problem of the existence of an LME to finding a solution to a *boundary value problem* (BVP) including ODEs for the two aforementioned functions of time and for the coefficients in the sender's strategy. The two *learning* ODEs endow the BVP with exogenous initial conditions, while the rest carry terminal ones arising from the static game of two-sided incomplete information at the end of the interaction. The ODEs are obviously coupled: the learning coefficients depend on the signaling that takes place over time, but the latter depends on the path of the learning coefficients because these are taken as given in the best response problem.

With ODEs in both directions, proving the existence of a solution to such a BVP is a complex "shooting" problem: not only must solutions to all ODEs exist, but they must land at specific (potentially endogenous) values.[4] Symmetric environments—the paradigm for games of multi-sided Gaussian learning—are tractable because all players can signal at the same rate, leading to a single learning ODE. There, the problem is one-dimensional, and the traditional shooting method applies: one traces a candidate initial condition of the ODE to be shot over an interval such that the intermediate value theorem ensures the target is hit.[5]

This intuitive continuity method clearly does not extend to settings where the target is multidimensional. Our contribution then lies in framing the problem as a novel fixed-point one. Specifically, given functions that proxy for solutions to our learning ODEs, we obtain candidate equilibrium coefficients by solving their respective ODEs backwards. Equipped with the latter tuple, we obtain solutions to the learning ODEs by solving them forward. We can then construct an infinite-dimensional fixed-point problem over candidate learning functions, to which Schauder's theorem applies. Via this approach, Theorem 1 establishes the existence of LMEs for all horizon lengths up to a threshold that is inversely proportional to the environment's initial uncertainty, irrespective of the discount rate.

Our approach is a major step forward in the literature, both conceptually and method-

[4] For instance, one must find initial values for the coefficients in the strategy so that the resulting terminal values match the static Nash equilibrium at the end of the game, which can be endogenous due to learning.

[5] See Bonatti et al. (2017) studying Cournot competition. In insider trading models such as Back et al. (2000), linear payoffs further simplify this issue because terminal conditions are pinned down by no arbitrage.

ologically. Conceptually, the infinite-dimensional technique is in fact the best avenue for fully exploiting the economics of the problem: namely, that learning is determined iteratively forward in time given conjectured behavior, while behavior is determined backward—by backward induction—given a conjectured informativeness of the signals. Indeed, this is demonstrated by the fact that the thresholds on the time horizon for which we can guarantee existence have the same order as those found when the knife-edge one-dimensional shooting method applies. Methodologically, by being able to handle multiple different learning ODEs, our method offers a general approach for addressing linear-quadratic games with Gaussian learning and unobserved actions at a substantially broad level, an issue that we discuss in Section 6. Thus, the paper not only enables a tractable analysis of dynamic settings where higher-order beliefs matter, but it also uncovers a fundamental structure to the solution of these games while delivering new tools that are portable to new environments.

Related Literature. The literature on private monitoring has developed mostly in the context of repeated games with *complete* information, a framework in which the issue of inferences of private histories has been handled with methods that are very different to ours. Closest in spirit is Phelan and Skrzypacz (2012), where such inferences are coarsened into beliefs over a finite set of states. Other approaches include looking for mixed-strategy equilibria where the inference of histories becomes irrelevant (Ely and Välimäki, 2002), or examining the set of attainable payoffs in the patient limit case (Sugaya, 2021). Instead, in our approach with incomplete information, we construct belief-dependent equilibria where discount rates are fixed and the players must infer their rivals' entire private histories.

Regarding traditional signaling models, in *static* (i.e., sequential-move, one-shot) noisy signaling games (e.g., Matthews and Mirman, 1983; Carlsson and Dasgupta, 1997), the signal realization is trivially hidden from the sender at the moment of action, but the environment is public in that the receiver's prior belief is common knowledge at that time. In dynamic environments, the receiver's belief is also public in settings with observable actions and an exogenous, public stochastic process (e.g., Kremer and Skrzypacz, 2007; Daley and Green, 2012; Kolb, 2019; Gryglewicz and Kolb, 2021). On the other hand, beliefs can be private when there are *exogenous* private signals of the sender's type (Feltovich et al., 2002; Cetemen and Margaria, 2020; Kolb et al., 2021). By contrast, in our setting all players' beliefs are private and the associated signals are actively affected by behavior.

Combining unobserved actions and private information in dynamic settings is challenging because the players develop private beliefs, and hence the need to forecast those states can arise. In this line, linear-quadratic-Gaussian models have proven useful, 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 ones. 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, effectively constitute a perfect public signal channel through which the firms' private beliefs are revealed. Thus, in none of these papers are higher-order (private) beliefs needed as states; and because any imperfect learning is either symmetric across players, or simply one-sided, the need to solve a multidimensional BVP does not arise.

To conclude, this paper contributes to a growing literature using continuous-time methods to analyze dynamic incentives. Sannikov (2007) examines two-player games of imperfect public monitoring; Faingold and Sannikov (2011) reputation effects with behavioral types; Cisternas (2018) games of ex ante symmetric incomplete information; and Bergemann and Strack (2015) dynamic revenue maximization.

2 Model

We consider two-player *dynamic noisy signaling* games of a linear-quadratic-Gaussian (LQG) nature where the ex ante informed player does not directly observe the signals of her actions.

Model basics. A forward-looking sender (she) interacts with a myopic receiver (he). Time runs continuously over a finite interval $[0, T]$, $T < \infty$. The environment is parametrized by the realization of a random variable θ that is the sender's private information, or her *type*. We assume that θ is normally distributed with mean $\mu \in \mathbb{R}$ and variance $\gamma^o > 0$; the latter are exogenous parameters in the model.

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. Given realized action paths $(a_t)_{t\in[0,T]}$ and $(\hat{a}_t)_{t\in[0,T]}$, the sender's 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 a discount rate. For simplicity, we assume that the terminal payoff function $\psi : \mathbb{R} \to \mathbb{R}$ is a concave quadratic—its dependence on the receiver's action resembles the sequentiality of traditional sender-receiver games.

The receiver, on the other hand, is concerned only about maximizing his flow utility at all instants of time. Given realized actions a_t and $\hat{a}_t$, his 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.

Turning to the information structure, the sender knows θ before play begins. Instead, the receiver only knows its distribution $\theta \sim \mathcal{N}(\mu, \gamma^o)$, and this is common knowledge. There are also two noisy signals that are linear in the players' actions,

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

where Z^X and Z^Y are orthogonal Brownian motions and $\sigma_X, \sigma_Y \in \mathbb{R}_+$ are volatility parameters. Our key innovation is to make Y—which carries information about the sender's actions—*privately observed by the receiver*; instead, 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 is natural in the applications we study.

Due to the full-support structure of (3), 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)* $|u_{a\hat{a}}| + |u_{\hat{a}\hat{a}}| + |\psi_{\hat{a}\hat{a}}| \neq 0$*.*

Parts (i) and (ii) ensure that the players condition their actions on their "first-order" private information: the sender's action is sensitive to her type, while the receiver's will be sensitive to his private belief. Part (iii) in turn guarantees the use of a second-order belief: either a non-trivial strategic interaction term ($u_{a\hat{a}} \neq 0$), or a nonlinearity coming 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 behavior. Further technical conditions that we use to ensure the existence of the equilibria studied are presented in Section 5.[7]

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

[7]These conditions are minimal in that they pertain to the existence of a non-trivial linear Bayes' Nash equilibrium in the static game of two-sided incomplete information that arises at the terminal time T.

Applications. With quadratic preferences, types can only face one of two incentives depending on what they perceive the receiver's belief to be: *revelation* or *concealment* motives. The revelation motive manifests in the sender steering the receiver's belief in the direction of the true type. If this motive stops before reaching the type, or continues beyond it, we say that the concealment motive kicks in: manipulating the belief away from the type.

Our applications therefore purposely intend to isolate each of these cornerstone motives, recognizing that all environments will display incentives that are a mix of these two extremes. In the *coordination game* that we study, all types have a revelation motive, but one that is trumped by the presence of information frictions. In the *reputation game* that we study, all types generically want to conceal their identities, but this turns out to be costly. Our goal is to understand how the presence of higher-order uncertainty can affect behavior, learning, and payoffs relative to public analyses where beliefs are common knowledge. To this end, we vary the volatility σ_X of the public feedback X: as a measure of the quality of the feedback received by the sender, it shapes the extent of higher-order uncertainty in each setting.

Application 1: A coordination game. Our first example aims to shed light on a classic problem faced by organizations: how to best adapt to new economic conditions when coordinating activities is desired yet information is dispersed among decision-makers? We interpret θ as the realized value of a new state of the world and consider the payoffs

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

A "leader" (e.g., top management of an organization—the sender) and a "follower" (e.g., a key division—the receiver) form a team. The performance of the team depends on both the leader's adaptation to the environment and coordination. The leader wants to maximize the performance of this form of organization, while the follower simply wants to coordinate; we discuss the latter assumption shortly. Finally, the scalars attached simply deliver the desired normalization $u_{aa}=\hat{u}_{\hat{a}a}=-1$.

Our starting point is the wide recognition that coordination needs are a central element driving organizational performance (e.g., Milgrom and Roberts, 1992). Thus, while leaders may be able to visualize how to respond to change, they will need other decision-makers to understand the new paradigms to adapt efficiently. A leader can then begin taking actions to move the organization in her desired direction, but the challenge is that the organization's inferences are likely to be subjective. A natural, yet largely unexplored, two-way inference problem arises in those settings: both parties are trying to guess each other's understanding simultaneously as decisions are being made. Our approach to making progress in this area is via a two-player signaling game in which any attempt by the leader to signal the new state of the world will inevitably result in the receiver developing a private belief.

In this context, how does the leader, via her signaling, manage the transition and what are the corresponding implications on performance and learning? Because our ultimate goal is to uncover the effects of higher-order uncertainty on outcomes, we have aligned the players' incentives by making the receiver simply interested in following at all times—this choice is analogous to the common-interest assumption in the *statistical theory of teams* (e.g., Marschak and Radner, 1972) intended to isolate the effects that information frictions can have on organizations' performance. Thus, all leader types will happen to have a revelation motive, which will vary with the quality of the feedback received, σ_X. We note that our methods, however, can accommodate biases—and hence, concealment motives—too.[8]

Application 2: A reputation game. For notational simplicity, let us normalize the prior mean μ to zero. We consider the following payoffs for our players:

$$\frac{1}{2}\left[-\int_0^T e^{-rt}(a_t-\theta)^2dt - e^{-rT}\psi\hat{a}_T^2\right], \text{ with } \psi>0, \text{ and } \hat{u}(a_t,\hat{a}_t,\theta) = -\frac{1}{2}(\hat{a}_t-\theta)^2.$$

In this specification, the sender finds it costly to take actions away from her type ($-(a_t-\theta)^2$ term) and she benefits from the receiver's terminal action $\hat{a}_T$ being close to the prior mean $\mu=0$. In turn, the receiver aims to solve a classic prediction problem at all times: he will choose the best predictor of the type given his information.

We interpret this model as one of *reputation for neutrality*. Consider a politician or expert—the sender—with $\theta \sim N(0,\gamma^o)$ representing the intensity of her (horizontal) bias on a relevant issue (e.g., vaccine mandates vis-à-vis individual liberties during a pandemic, or environmental regulations vis-à-vis economic growth in light of climate change); the type $\theta=0$ captures an *unbiased type.*[9] The receiver is a news outlet that gets private signals Y of the politician's past behavior[10] and that reports its perception of the bias. The resulting

[8]With aligned preferences, noiseless talk renders the problem trivial. Two remarks are then in order. First, imperfect communication is standard in analyses of organizations, with much of the common-interest team theory literature involving imperfect message transmission; ours can then be seen as an action-based counterpart, a direction suggested by Marschak (1955) (p. 157). Second, conveying information verbally can be difficult in real organizational problems, as much of the relevant knowledge there is hard to codify and transfer (or *tacit*; Garicano, 2000). This dimensionality issue—that productive know-how is more complex than the communication channels available—is what shutting down talk delivers in our model: the channel's dimension shrinks to 0 while the state of the world has dimension 1. Interestingly, work in the strategic management literature bridges both remarks: due to its complexity, tacit knowledge transmission is linked to observing its application, and hence rooted in actions (Nonaka, 1991; Grant, 1996).

[9]Also in politics, but with a different motivation, Bouvard and Lévy (2019) study a horizontal model of reputation with quadratic preferences in which uncertainty is symmetric.

[10]Actions such as voting, contributions, favors, statements to groups of influence, etc. often have a private

reporting process X is naturally public; imperfect (e.g., journalist turnover, idiosyncratic opinions, etc.); and fair on average (the shocks have zero mean on average). We are interested in the case in which the politician wants the outlet's time-T perception to be as close as possible to $\mu = 0$—that is, she wants to appear neutral at a final future date (e.g., a time at which she is to be appointed to a high-profile position).

Unlike in the previous example, all types now have a *concealment* motive. Specifically, since the prior is that the politician unbiased, all types would like to "pool" on a particular action, or even refrain from taking actions at all, so that no information is conveyed and their reputation remains at the prior. However, all types face short-run temptations that prevent them from doing so ($-(a_t - \theta)^2$ term). Clearly, more precise reporting as measured by low values of σ_X helps the politician in that she can better tailor her actions to her reputation. But how does this better ability to tailor interact with her "commitment" problem?

Remark 1. *For both applications, the equilibria we characterize remain equilibria when the receiver becomes forward looking, due to this player solving a* prediction problem *in each application. Refer to Section 6 for more details.*

We close the section with the definition of strategies and equilibrium.

Strategies and Equilibrium Concept. The full-support monitoring implies that the only off-path histories for each player are those in which that same player has deviated. Thus, we use the Nash equilibrium concept for defining the equilibrium of the game and leave off-path behavior unspecified at this point, as imposing sequential rationality does not further refine the set of equilibrium outcomes in games with unobserved actions.[11]

A (pure, action-free) admissible strategy for the sender is 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.[12] 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)**.** *An admissible pair $(a_t, \hat{a}_t)_{t\geq 0}$ 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.*

nature, and hence are likely to be leaked with error, justifying the noise in Y.

[11]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.

[12]Square integrability refers to $\int_0^T a_t^2 dt$ and $\int_0^T \hat{a}_t^2 dt$ being finite in expectation. This ensures that a strong solution to (3) exists (Ch. 1.3 and 3.2 in Pham, 2009) and thus the outcome of the game is well defined.

The LQG structure suggests looking for Nash equilibria where the strategies are *linear* functions of the signals observed by each player. While this is a simple task in static settings, it is a far more challenging enterprise in dynamic environments. Specifically, the core of the issue is that evaluating the candidacy of an equilibrium profile requires assessing the value of deviations; but with incomplete information and imperfect monitoring, the sender will find it optimal to condition on more information than (θ, X) in the continuation game after she deviated. The next section formalizes this idea by means of a *belief-based recursive method* for finding linear Nash equilibria that has two key properties: first, it demonstrates that finding such equilibria is inherently linked to imposing full sequential rationality under a richer set of strategies (of course, not restricted to the linear or Markov classes); second, it shows that the ensuing linear aggregation of signals is the outcome of players naturally relying on their beliefs to guide their behavior.

3 Equilibrium Analysis: Linear Markov Equilibria

In this section we lay out our method for finding a Nash equilibrium in linear strategies. The starting point is that any recursive approach will naturally demand the players to form beliefs about each other's private information in order to be able to assess the continuation game as the interaction unfolds. The next subsection offers an overview of the belief states employed in our construction, and it builds intuition for why a second-order belief is needed.

3.1 Belief States: An Overview

We will characterize equilibria in which, on and off the path of play, the sender and receiver behave according to the linear *Markov* strategies

$$\begin{aligned} a_t &= \beta_{0t} + \beta_{1t} M_t + \beta_{2t} L_t + \beta_{3t}\theta & (4)\\ \hat{a}_t &= \delta_{0t} + \delta_{1t}\hat{M}_t + \delta_{2t} L_t, & (5)\end{aligned}$$

with the coefficients β_{it}, $i = 0, 1, 2, 3$, and δ_{jt}, $j = 0, 1, 2$, differentiable functions of time, and

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

That is, $\hat{M}_t$ is the receiver's first-order belief, M_t the sender's second-order counterpart, and L_t is the belief about θ using the public information exclusively, $t \in [0, T]$. (For brevity, we refer to the means of posterior beliefs as simply "beliefs.") The deterministic nature of the coefficients will encode learning and time-horizon effects.

Consider our coordination game: $u(\theta, a, \hat{a}) \propto -(a-\theta)^2-(a-\hat{a})^2$ and $\hat{u}(\theta, a, \hat{a}) \propto -(a-\hat{a})^2$. Since the sender has an adaptation motive, θ is a relevant state; but because the receiver wants to match the sender's action, $\hat{M}$ is relevant for this player's behavior. The sender's coordination motive then forces her to forecast $\hat{M}$, and the second-order belief M appears. More generally, (iii) in Assumption 1 ensures that M is needed as a proxy for $\hat{M}$ for the sender to compute her continuation payoff.

To guide intuition, we preview key properties of this second-order belief state that we establish in this section. These properties illustrate: the "beliefs about beliefs" problem that players face; how this problem interacts with the determination of optimal behavior; and why the state space does not grow indefinitely as a result of an infinite regress problem. As a byproduct, the appearance of the public state L is explained. Specifically:

(i) 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 $(X_s)_{s<t}$, $t \in [0, T]$;

(ii) $(M_t)_{t\in[0,T]}$ is the only state variable directly controlled by the sender;

(iii) M_t is a convex combination of θ and L_t, $t \in [0, T]$, if (4)–(5) are followed.

That M depends on X is intuitive: $\hat{M}$, via the receiver's action, enters the public signal. Due to the private monitoring, however, the sender also relies on her *past play* to forecast $\hat{M}$: higher past actions are statistically informative of higher values of Y observed by the receiver, so M should be higher for any fixed public history of X. The contrast to public monitoring is clear: if Y were public, past behavior would be irrelevant in this forecasting exercise, as the receiver's belief would be uniquely determined by the realizations of Y.

The explicit dependence on past actions implies that M is controlled by the sender. Further, this is the only state directly controlled because L, as a public state, is ultimately a function of the history of X only, and this latter signal is not directly affected by the sender (see (3)). Thus, finding (linear) Nash equilibria *requires determining optimality of the sender's behavior with respect to* M. This observation is important because, by (iii), if the sender is on path, her action does not depend explicitly on M due to the latter state becoming exclusively a function of θ and L, in turn suggesting that one could dispense with M in the analysis. This, however, is not possible because of the need to evaluate deviations when assessing the optimality of any (θ, L)-dependent strategy. In doing so, the sender must rely on M to calculate her continuation value, but after deviations M is not spanned by θ and L due to the dependence on past play. Further, as the unique controlled state, changes in the continuation value necessarily occur via changes in M—the use of this state is critical.

A second implication of the explicit dependence on past actions is that M is *private information to the sender in equilibrium*: by property (iii), the second-order belief is a function of the sender's type along the path of play, as her equilibrium actions depend on θ. The receiver must then forecast this state itself, but the state space does not grow further because M is linear in θ and L under the linear Markov strategy (4). Since L is used in the receiver's forecasting exercise, it becomes a relevant state for both players.

The next two subsections formally demonstrate properties (i)–(iii) in order to pose a well-defined best-response problem for the sender. We begin this task with (iii): namely, with a *representation of the second-order belief* under linear Markov strategies.

3.2 Belief Representation and History-Inference Effect

We begin by establishing a representation for the second-order belief M under the linear-Markov strategies (4)–(5). (To avoid repetition, we defer an expression under deviations to the next section.) This step is key in our analysis, both conceptually and technically.

Indeed, as different types proceed with taking different actions, their reliance on past play to assess the continuation game will lead them to hold different beliefs—in the sender's strategy (4), this leads to M becoming a source of heterogeneity in addition to the direct contribution that the type has on behavior. In equilibrium, the receiver must account for this channel to form a correct belief about the type, but this amounts to anticipating the form of M in a candidate linear Markov equilibrium. Conceptually, then, the representation encodes the *signaling* that occurs via the second-order belief channel.

Given the LQG structure, it is natural to expect a representation of the form

$$M_t = \chi_t \theta + (1 - \chi_t) L_t, \tag{6}$$

where $L_t := \mathbb{E}[\theta | \mathcal{F}_t^X]$ and $(\chi_t)_{t \in [0,T]}$ is deterministic. Intuitively, behind (6) is how the sender balances the information conveyed by her private histories—$\chi_t \theta$ term—with that conveyed by the public signal—$(1 - \chi_t) L_t$ term—in her forecasting exercise.[13]

A functional form like (6) is operational only if we know how (χ, L) depends on the candidate equilibrium linear Markov strategies (4)–(5). Indeed, because the receiver must anticipate how $(\chi_t)_{t \in [0,T]}$ will evolve in equilibrium to formulate both his belief and best-response, the sender is forced to anticipate the same weight to evaluate her possible courses of action. Technically, therefore, having up-front knowledge of a representation is a requirement for being able to set up the sender's best response problem.[14]

[13] With pure strategies, the outcome of the game should be a function of the signals available to the players, so M must be a function of θ and X; the time-varying weights and the additive separability are consequences of the induced Gaussian learning under linear strategies. Also, from (6) and the fact that deviations go undetected, the receiver does not need to rely on additional states beyond L and $\hat{M}$.

[14] As for $L_t = \mathbb{E}[\theta | \mathcal{F}_t^X]$, the representation is still key in that it delivers a law of motion for this state.

The main result of this section (Lemma 1) establishes the representation (6) where χ and L are characterized via *laws of motion*. To build towards the result, it is instructive to elaborate on how it is obtained. First, our approach is constructive: it starts by assuming (6) with $(L_t)_{t\in[0,T]}$ a general process depending only on the public information. The receiver then always assumes that the representation holds, as deviations are never detected due to the full support of Y. Inserting (6) into (4), the receiver expects, at all times,

$$a_t = \alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}\theta \tag{7}$$

$$\text{where} \qquad \alpha_{0t} := \beta_{0t}, \qquad \alpha_{2t} := \beta_{2t} + \beta_{1t}(1-\chi_t), \qquad \text{and} \qquad \alpha_{3t} := \beta_{3t} + \beta_{1t}\chi_t. \tag{8}$$

Note that information transmission is ultimately guided by the total weight on the type, so it is suitable to refer to α_{3t} as the *signaling coefficient*. We note, however, that this coefficient, via β_1 and β_3, is a mix of both non-strategic (i.e., static) and strategic (i.e., dynamic) motives—our choice over alternatives (such as α_3 net of a myopic counterpart) is purely for ease of exposition.

The receiver then filters θ assuming that Y is driven by (7). This problem is (conditionally) Gaussian (Liptser and Shiryaev, 1977, Theorems 12.6 and 12.7), so this player'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, and where the evolution of $\hat{M}$ depends on γ. Importantly, the linearity of the signal structure renders the pair $(\hat{M}, X)$ under (5) (conditionally) Gaussian too. The sender's problem of filtering $\hat{M}$ using X then yields another mean-variance pair, but with the corresponding mean M_t now depending explicitly on her past actions: for any given history of the public signal, changes in the sender's history of play will shift the mean of her belief. One can then solve for M_t under the linear strategy (4) to obtain differential equations for (χ, L).

Lemma 1. *Suppose that (X,Y) is driven by (4)–(5) and the receiver believes that (6) holds, with $(L_t)_{t\in[0,T]}$ a process that depends only on the public information.*[15] *Then (6) holds at all times if and only if $L_t = \mathbb{E}[\theta|\mathcal{F}_t^X]$ and $\chi_t = \mathbb{E}_t[(M_t - \hat{M}_t)^2]/\gamma_t$, where*

$$\dot{\gamma}_t \;=\; -\frac{\gamma_t^2(\beta_{3t} + \beta_{1t}\chi_t)^2}{\sigma_Y^2}, \qquad\qquad \gamma_0 = \gamma^o, \tag{9}$$

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

$$\dot{\chi}_t = \frac{\gamma_t(\beta_{3t}+\beta_{1t}\chi_t)^2(1-\chi_t)}{\sigma_Y^2} - \frac{\gamma_t\chi_t^2\delta_{1t}^2}{\sigma_X^2}, \qquad \chi_0 = 0, \tag{10}$$

$$dL_t = (l_{0t}+l_{1t}L_t)dt + B_t dX_t, \qquad L_0 = \mu, \tag{11}$$

with (l_{0t}, l_{1t}, B_t) deterministic and given in (A.7).

The lemma confirms that the public state satisfies $L_t = \mathbb{E}[\theta|\mathcal{F}_t^X]$, and characterizes the weight χ_t in (6) as a ratio of the players' posterior variances. It further offers dynamics for their evolution that are coupled with that in (9) for γ. The latter ODE is standard and states that the variance of the receiver's belief decays at a deterministic rate that is increasing in the signaling coefficient. The ODE (10) for χ on the other hand captures how types progressively separate over time due to their gradually differing beliefs. Finally, the linearity of L in the histories of X—by virtue of the Gaussian learning—is clear from (11).

Using these findings, we can uncover the economics behind the representation $M_t = \chi_t\theta + (1-\chi_t)L_t$. Consider first the beginning of the game. Since there is no second-order uncertainty at the outset, $M_0 = \mu$, an expression that is also delivered by the representation due to $\chi_0 = 0$ and $L_0 = \mu$ from the lemma. As soon as the sender conditions her actions on her type, however, the signal Y becomes informative and the sender loses track of the receiver's belief (i.e., $\mathbb{E}_t[(M_t - \hat{M}_t)^2] > 0$). Past actions are then useful in forecasting $\hat{M}$, a phenomenon that is readily apparent when there is non-trivial signaling at time 0: in (10), $\dot{\chi}_0 = \gamma^o\beta_{30}^2/\sigma_Y^2 > 0$ if $\beta_{30} \neq 0$, and the connection between M and θ present in (6) emerges.

To understand why the ODEs (9) and (10) are coupled, assume that the last term in (10), $\gamma_t\chi_t^2\delta_{1t}^2/\sigma_X^2$, is absent. Then, it is easy to verify that any solution to (9)–(10) satisfies

$$\chi = 1 - \frac{\gamma_t}{\gamma^o}.$$

Intuitively, if the sender has signaled more aggressively, she will expect the receiver to be more certain about her type, so lower values of γ are inherently linked to higher values of χ. Clearly, the rates of change of each depend on the signaling coefficient $\beta_{3t} + \beta_{1t}\chi_t$.

The presence of a public feedback channel nevertheless alters the previous relationship. This effect is captured by the last term in (10), where the key term is the signal-to-noise ratio of X, δ_{1t}^2/σ_X^2. Either an increase in the magnitude of δ_1, which we call the receiver's signaling coefficient, or a reduction in the volatility σ_X, improves the quality of X, thereby inducing the sender to rely more on the public information to forecast $\hat{M}$: in the χ-ODE, more downward pressure is put on the growth of χ as δ_1^2/σ_X^2 grows. In the limit as $\delta_1^2/\sigma_X^2 \nearrow \infty$, (10) cannot "take off," and so $\chi \equiv 0$; also, someone who has access only to the public signal learns the receiver's belief in real time. Thus, $M_t = \chi_t\theta + (1-\chi_t)L_t$ reduces $M_t = \hat{M}_t$, i.e., the environment becomes *public*.

To anticipate the main phenomenon under study, notice that when the environment is public—because Y public or because $\sigma_X = 0$—an LME naturally entails the sender and receiver acting linearly in $(\theta, \hat{M})$ and $\hat{M}$, respectively.[16] The common knowledge of $\hat{M}$ implies that the extent of signaling is then solely determined by the weight attached to the type in the sender's strategy: for a given history of Y, all types agree on the value of $\hat{M}$. This notion breaks with private monitoring, as reflected in the correction

$$\beta_{1t}\chi_t$$

in the signaling coefficient (8). We refer to it as the *history-inference effect* on signaling.

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

The representation implies that, on the equilibrium path, the sender's actions depend only on his type θ and the public belief L as in (7). To perform equilibrium analysis, however, the sender must evaluate deviations from (7); but at off-path histories, the representation does not hold due to its reliance on (4) being followed. The sender must then account for M and L separately: the former as an estimate of the receiver's private belief, and the latter because the receiver uses it in forecasting the sender's second-order belief.[17]

Laws of motion for M and L for arbitrary strategies of the sender (up to technical conditions specified shortly) are presented next.

Lemma 2 (Controlled dynamics)**.** *Suppose that the receiver follows* (5) *and believes that* (4) *and* (6) *hold. Then, if the sender follows* $(a'_t)_{t\in[0,T]}$,

$$dM_t = \frac{\gamma_t\alpha_{3t}}{\sigma_Y^2}(a'_t - [\alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}M_t])dt + \frac{\chi_t\gamma_t\delta_{1t}}{\sigma_X}dZ_t \quad (12)$$

$$dL_t = \frac{\chi_t\gamma_t\delta_{1t}}{\sigma_X^2(1-\chi_t)}[\delta_{1t}(M_t - L_t)dt + \sigma_X dZ_t], \quad (13)$$

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 from the sender's perspective. Also,* $\mathbb{E}_t[(M_t - \hat{M}_t)^2] = \gamma_t\chi_t$ *for any such* $(a'_t)_{t\in[0,T]}$.

[16] From the viewpoint of the optimality of the receiver's behavior, this notion of LME is "perfect" when Y is public, but only *Nash* when Y is private but $\sigma_X = 0$.

[17] That M is no longer spanned by θ and L after deviations reflects the divergence in the game's structure at on- versus off-path histories. This also happens in models of incomplete information with public signals such as Foster and Viswanathan (1996), Bonatti et al. (2017) and Cisternas (2018), where instead *exogenous* private beliefs must be accounted for after deviations. See also the survey article of (Kandori, 2002) describing the lack of recursive structure in traditional games of private monitoring.

The dynamics (12)–(13) are involved, so let us focus on a few economic properties. First, revisiting (i) from Section 3.1, M is indeed an explicit function of the past actions of the sender and past realizations of X, an issue at the heart of our construction. To see this, we can insert the definition of Z_t into the law of motion (12) 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 (11) shows that L_s is an explicit function of $(X_\tau)_{\tau<s}$.

Second, from the drifts of M and L we confirm that M is the only state directly controlled by the sender. In particular, not only does this non-trivial state "appear" after deviations, but it is through establishing optimality of this state that equilibrium behavior must be pinned down—this justifies our expositional choice to start with an extended strategy (4) that treats M and L differently as opposed to the outcome (7). From (12), moreover, the sender expects the receiver to be more responsive to her actions when the receiver is more uncertain (high γ) or when there is more signaling (high α_3).

Finally, the dynamic of L—which, again, is always an exclusive function of past values of X—reflects the predictable (drift) and unpredictable (Brownian) components from the sender's perspective, a distinction that matters for optimization. It is simply obtained from (11) by using the evolution of the public signal as forecasted by the sender (who assumes the receiver is always on path). In the predictable part, we note that M feeds into the drift of L. This implies that the sender expects to influence the public belief despite not being able to directly affect the public signal X: higher actions suggest higher values of the receiver's belief, which ultimately influences X. Changes in L then matter for the sender's incentives because the receiver uses this state to forecast M.[18]

As a technical, but important, observation, we note that the dynamics (12)–(13) depend on solutions (γ, χ) to the ODEs (9)–(10). This dependence originates from the receiver's learning process: since deviations are hidden, this player always assumes that the representation (6) holds when constructing his belief. The next result shows that the drifts and volatilities are all well defined when the coefficients in the linear Markov strategies are continuous. To this end, we note that it can be easily verified from the receiver's first-order condition that his best reply attaches weight $\hat{u}_{\hat{a}\theta} + \hat{u}_{\hat{a}a}[\beta_{3t} + \beta_{1t}\chi_t]$ to $\hat{M}$.

Lemma 3 (Learning ODEs). *Suppose (β_1, β_3) is continuous and $\delta_{1t} = \hat{u}_{\hat{a}\theta} + \hat{u}_{\hat{a}a}[\beta_{3t} + \beta_{1t}\chi_t]$. Then (9)–(10) governing (γ, χ) has a unique solution. In this solution, $0 < \gamma_t \leq \gamma^o$ and*

[18] To gain more intuition on the dynamics, note from the drift of (13) that L moves towards M on average, i.e., someone who only observes X must gradually learn the type over time. On the other hand, 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 own behavior. The volatility terms capture how both states will respond to the future information conveyed by the public signal; if the latter is uninformative (e.g., $\delta_1/\sigma_X \equiv 0$), M evolves deterministically and $L \equiv \mu$.

$0 \leq \chi_t < 1$ *for all* $t \in [0,T]$, *with strict inequalities over* $(0,T]$ *if* $\beta_{30} \neq 0$.

We are now in a position to state the sender's best-response problem. By the last part of Lemma 2, the sender's posterior variance, $\mathbb{E}_t[(M_t - \hat{M}_t)^2]$, is independent of the strategy followed, with a value of $\gamma_t\chi_t$; intuitively, the additive structure in the signals and the Gaussian noise imply that changes in the sender's actions simply shift the receiver's belief for any realization of the shocks (Z^X, Z^Y).[19] Using this and the definition of M, we have $\mathbb{E}_t[u(a_t, \delta_{0,t} + \delta_{1,t}\hat{M}_t + \delta_{2t}L_t, \theta)] = \mathbb{E}_t[u(a_t, \delta_{0t} + \delta_{1,t}M_t + \delta_{2t}L_t, \theta)] + \frac{1}{2}u_{\hat{a}\hat{a}}\delta_{1t}^2\gamma_t\chi_t$, and likewise for the terminal payoff ψ in place of u at $t = T$; in other words, no moments of higher order are needed as additional state variables. We conclude that, up to an additive constant in the total payoff, the sender's best-response problem consists of maximizing

$$\mathbb{E}_0\left[\int_0^T e^{-rt}u(a_t, \delta_{0t} + \delta_{1t}M_t + \delta_{2t}L_t, \theta)dt + e^{-rT}\psi(\delta_{0T} + \delta_{1T}M_T + \delta_{2T}L_T)\right] \tag{14}$$

subject to the dynamics (12)–(13) of (M, L) and the ODEs (9)–(10) for (γ, χ).[20]

The set 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. It is important to stress two aspects of this set. First, it is richer than that used in the Nash equilibrium concept due to the explicit conditioning on past behavior via M. Second, this set is extremely general, well beyond the linear class: in particular, it allows for strategies that are non-linear in the states, that condition on entire histories, that are discontinuous, etc.—that is, we are not restricting the set of deviations to linear Markov strategies with differentiable coeffients in our search for an equilibrium with those properties.

A tuple $(\beta_0, \beta_1, \beta_2, \beta_3)$ of deterministic functions 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 $(\delta_0, \delta_1, \delta_2)$ in the receiver's strategy are optimal and correct, i.e., when

$$\hat{a}_t := \delta_{0t} + \delta_{1t}\hat{M}_t + \delta_{2t}L_t = \arg\max_{\hat{a}'\in\mathbb{R}} \hat{\mathbb{E}}_t[\hat{u}(\alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}\theta, \hat{a}', \theta)]. \tag{15}$$

This notion of equilibrium is clearly *perfect* in that it specifies optimal behavior by the sender after deviations. Finally, along the path of play of such a policy, the representation (6) holds by construction, and so the sender's behavior is given by $a_t = \alpha_{0t} + \alpha_{2t}L_t + \alpha_{3t}\theta$, where $(L_t)_{t\in[0,T]}$ follows (11) in Lemma 1—a Nash equilibrium in linear strategies ensues.[21]

[19] If the public signal is perfectly informative, i.e., $\delta_1/\sigma_X = +\infty$, we have that $\chi \equiv 0$ and so $\mathbb{E}_t[(M_t - \hat{M}_t)^2] \equiv 0$, confirming that the environment becomes public.

[20] 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.

[21] Regarding the receiver, while deviations by this player do affect L, it is clear that no additional states

In the next section we specialize the model to the two benchmark applications from Section 2. This allows us to illustrate the new economic insights that private monitoring—via the history-inference effect created—brings to the analysis of noisy signaling games. The reader interested in the question of the existence of LME can immediately skip to Section 5, where we address the best-response problem in its general form.

4 Applications

We begin with our coordination game within organizations, which is an ideal pedagogical example for fleshing out the new insights on behavior and outcomes that we uncover. Having developed this application, we offer a more streamlined exposition of our reputation model.

4.1 Coordination: Learning and Performance in Organizations

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

$$\text{“leader/team”:}\ \int_0^T e^{-rt}\{-\underbrace{(a_t-\theta)^2}_{\text{adaptation}}-\underbrace{(a_t-\hat{a}_t)^2}_{\text{coordination}}\}dt;\ \text{“follower”:}\ -(a_t-\hat{a}_t)^2.$$

That is, the leader aims to take actions adapted to the economic environment (i.e., θ) while accounting for the need to coordinate activities; the follower simply wants to coordinate. How does the leader guide actions towards θ, and what are the implications on learning—captured by the follower's terminal posterior variance γ_T—and on performance—captured by the leader's total payoff? Note that, due to the alignment of preferences, if the leader were able to transmit her knowledge about θ, the organization would incur no losses going forward; a more informed organization is a priori suggestive of better performance.

The public benchmark Suppose that the signal Y is public or the public signal noiseless, i.e., $\sigma_X = 0$: in either case, the follower's belief $\hat{M}$ is known to the leader or to someone who only observes the public signal, so the environment is effectively public.[22] Mathematically, $\hat{M} = M = L$ at all times, so there is a single state to track, $\hat{M}$. How does play unfold?

At time T, the players face a static game given any value of the follower's terminal belief, $\hat{M}_T$—the (Bayes) Nash equilibrium of this one-shot interaction is $a_T = \frac{1}{2}\theta + \frac{1}{2}\hat{M}_T$,

than $(t, L, \hat{M})$ are needed after deviations (see (A.1) in the Appendix for the law of motion of $\hat{M}$). Also, all the payoff-relevant histories are reachable on path, so the sequential rationality requirement is trivial for this player in an LME. All this is true if this player is forward looking.

[22] When $\sigma_X = 0$, this happens if the belief can be inverted from the action observed, which holds here.

$\hat{a}_T = \hat{\mathbb{E}}_T[a_T] = \hat{M}_T$. For $t < T$, a linear Markov equilibrium generalizes these strategies to

$$a_t = \beta_{0t} + \beta_{1t}\hat{M} + \beta_{3t}\theta \;\text{ and }\; \hat{a}_t = \hat{\mathbb{E}}_t[a_t] = \beta_{0t} + (\beta_{1t} + \beta_{3t})\hat{M}_t,$$

where the weights will differ from their static counterparts due to the leader's dynamic incentives. Equipped with these strategies, the follower's belief evolves as

$$d\hat{M}_t = \frac{\beta_{3t}\gamma_t}{\sigma_Y^2}\{dY_t - [\underbrace{(\beta_{0t} + (\beta_{1t} + \beta_{3t})\hat{M}_t)}_{=\hat{\mathbb{E}}_t[a_t]} dt]\} \;\text{ and }\; \dot{\gamma}_t = -\left(\frac{\beta_{3t}\gamma_t}{\sigma_Y}\right)^2, \tag{16}$$

i.e., $\hat{M}$ responds to unanticipated changes in the signal Y, with an intensity that increases with both the strength of the leader's signaling, β_3, and the degree of uncertainty of the follower, γ. The leader's problem is then to maximize her expected payoff when she controls $\hat{M}$ (via Y) as above—we look for a quadratic value function that features an optimal linear policy whose coefficients match those conjectured by the follower in (16).

Proposition 1. *When $\sigma_X = 0$, an LME exists for all $T > 0$ and $r \geq 0$. In any such LME, $a_t = (1 - \beta_{3t})\hat{M}_t + \beta_{3t}\theta$ for some strictly decreasing deterministic function β_{3t} satisfying $\beta_{3t} \in (1/2, 1)$, $t \in [0, T)$, and $\beta_{3T} = 1/2$.*

Under full information, the leader would choose $a_t = \theta$ at all times—from this perspective, the leader shifts weight from the state of the world toward the follower's belief because she cares about coordinating with an uninformed agent. The key, however, is that the leader does not lower the weight to the static counterpart of $1/2$ on the type, except at the end of the interaction. Indeed, stronger signaling, by generating higher signals on average, enables the leader to steer the follower's belief—and hence, his actions—towards θ faster, thereby allowing her to enjoy future coordination at a lower adaptation cost. In other words, the leader's revelation motive manifests in gradually investing in the follower's adaptation.

An important takeaway from Proposition 1 is that such signaling incentives fall over time (β_3 is decreasing) partly due to there being less time to enjoy the future benefits as time progresses.[23] In economic terms, a model with public signals yields the prediction that organizations facing new environments would engage in more novel changes in the beginning of their transition: leadership's actions are more sensitive to its superior information early on, decaying over time as the static coordination motive becomes stronger.

Remark 2 (Demonstrating by doing)**.** *A strategy of the form "place a large weight on the type at time zero, and play the type otherwise" also exhibits a decreasing weight on*

[23] There is a second force in the same direction: as learning progresses, i.e., γ falls, the belief becomes less responsive to changes in Y, so it is more difficult to steer the belief in any direction.

the type, and is a feasible deviation for the leader. However, it is not optimal when the follower conjectures a linear strategy: it is costly to deviate from the type today, and it risks miscoordination tomorrow due to placing excessive importance on a single signal conveying the information at once. Our equilibrium smooths out that logic, and it is more in line with the idea of "demonstrating by doing:" that learning is better achieved via a gradual progression of challenges, rather than exposing agents to completely foreign problems.

Higher-order uncertainty The previous finding on the evolution of the leader's signaling motive is, however, non-generic. In fact, as long as Y is private, any finite amount of volatility σ_X in the public signal will generate a *non-monotonic* signaling coefficient $\alpha_3 = \beta_3 + \beta_1\chi$.

Proposition 2. *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 $\alpha_{3t} > 0$. If, moreover, $r > 0$, then α_{3t} is non-monotonic and eventually decreasing.*[24]

Typical signaling coefficients α_3 are given by the hump-shaped dashed curves in Figure 1; the strictly decreasing signaling coefficient β_3 in the public case is depicted in black.

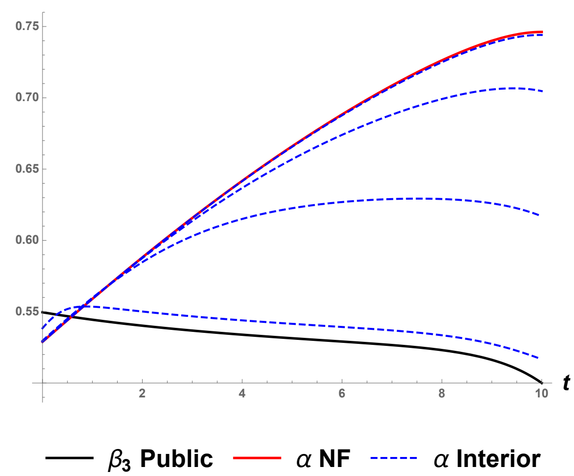

Figure 1: Signaling coefficients for $\sigma_X \in \{0, .1, .75, 2, 10, +\infty\}$ and $r > 0$; "α" denotes α_3.

To rationalize the initial increasing signaling phase, we now move to the polar opposite case of the previous public benchmark: when the public signal is simply pure noise, i.e., $\sigma_X = \infty$. This *no-feedback case* offers the cleanest set-up for understanding how the presence of higher-order uncertainty affects outcomes.

As argued in Section 3, the leader will now rely on her second-order belief M. And since the public signal is uninformative, the public belief L coincides with the prior mean μ at all

[24]Time horizons for which we can guarantee the existence of an LME are presented in the next section.

times. Using that $\beta_0 \equiv 0$ in equilibrium, a linear Markov strategy for the leader becomes

$$a_t = \beta_{1t} M_t + \beta_{2t}\mu + \beta_{3t}\theta.$$

Conjecturing a representation $M_t = \chi_t\theta + (1-\chi_t)\mu$, the follower expects $a_t = (\beta_{2t} + \beta_{1t}(1-\chi_t))\mu + (\beta_{3t} + \beta_{1t}\chi_t)\theta$ in equilibrium. The signaling coefficient $\alpha_{3t} = \beta_{3t} + \beta_{1t}\chi_t$ is then used by this player to construct $\hat{M}$. One can then show that the leader's belief satisfies

$$dM_t = \frac{\alpha_{3t}\gamma_t}{\sigma_Y^2}[a_t' - \underbrace{(\alpha_{2t} + \alpha_{3t}M_t)}_{\mathbb{E}_t[\hat{\mathbb{E}}_t[a_t]]}]dt, \quad \text{with} \quad \dot{\gamma}_t = -\left(\frac{\alpha_{3t}\gamma}{\sigma_Y}\right)^2 \quad \text{and} \quad \chi = 1 - \frac{\gamma_t}{\gamma^o}.$$

The controlled dynamic of M is now deterministic, reflecting that the leader's sole source of information to forecast $\hat{M}$ is her actions. This dynamic is again coupled with the follower's posterior variance, γ_t, but the latter's evolution now depends on the total extent of signaling $\alpha_{3t} = \beta_{3t} + \beta_{1t}\chi_t$ that accounts for the history-inference effect $\beta_1\chi$. Finally, if the above linear Markov strategy is followed, there is a one-to-one mapping between χ and γ.

Proposition 3. *When $\sigma_X = +\infty$, an LME exists for all $T > 0$ and $r \geq 0$. In any such LME, $a_t = (1-\alpha_{3t})\mu + \alpha_{3t}\theta$, where $\alpha_{3t} \in (1/2, 1)$, $t \in [0, T]$. Also, α_{3t} is strictly increasing for all $r > 0$ (and constant for $r = 0$), and $\alpha_T \nearrow 1$ as $T \nearrow \infty$.*

The leader's signaling here exhibits a completely different pattern: generically, the coefficient is strictly increasing, approaching 1 as the length of the interaction grows. To see why, note that in the public case higher types take higher actions because they have both higher (static) adaptation motives and higher (dynamic) steering motives, as they must drive their organizations to higher states of the world. These two forces are present in the no-feedback case, but the history-inference effects is also at play: higher types develop higher second-order beliefs, inducing them to take even higher actions due to the desire to coordinate. Moreover, since the coordination motive (β_1) becomes stronger as the end-game approaches, and χ grows due to a stronger reliance on past play for forecasting the continuation game, the history-inference effect strengthens over time, overcoming any fall in β_3.

We can now interpret the robust finding of a non-monotone signaling coefficient in the leader's strategy. Specifically, the leader starts "small" relative to the public case (the coefficients in Figure 1 cross at $t = 0$, addressed shortly), and the history-inference effect dominates in the beginning: as the leader expects the follower to gradually "get the message," she implements changes that better reflect the environment—the increasing phase. As time goes by, however, the coordination motive becomes stronger. While in the no-feedback case this further strengthened the history-inference effect, the public signal available can be used to coordinate: weight is shifted from β_1 to β_2 attached to L, thus dampening the history-inference effect and, combined with a decreasing steering incentive, α eventually decreases.

This is a much better story of adaptation to change, in that the leadership does not begin with drastic changes that the organization is unfamiliar with at the outset (the public case). Rather, the leadership slowly familiarizes the organization with the new environment first, and only implements more novel changes—as measured by the sensitivity of the leader's action to her private information—after more common understanding has been developed.

Outcomes: learning and payoffs Figure 1 suggests that there can be considerably more information transmission in a setting with higher-order uncertainty than if beliefs are public. Will this be the case, and are there any deep connections between learning and performance?

To formalize this conjecture, we contrast the public benchmark only with the no-feedback counterpart, the latter approximated by a large σ_X. Further, we set $r = 0$, a case in which we obtain unique analytic solutions. The use of superscripts *Pub* and *NF* should be clear.

Proposition 4. *Suppose that $r = 0$. For all $T > 0$,*

(i) The leader's ex ante payoff is larger in the public case;

(ii) $\beta_{30}^{Pub} > \alpha_{30}^{NF}$ and $\gamma_T^{Pub} > \gamma_T^{NF}$.

The result says two things. First, the organization is better off in the public case. Second, there is always more total information transmission in the no-feedback case — the follower's terminal belief has lower variance ($\gamma_T^{Pub} > \gamma_T^{NF}$). That is, worsening the feedback to a leader leads to more knowledge being transmitted by the end of the game. This latter finding is non-trivial because at the very beginning of the game, there is temporarily *less* information transmission in the no-feedback case ($\beta_{30}^{Pub} > \alpha_{30}^{NF}$) driven by an intertemporal substitution effect: anticipating that her future actions will be highly informative due to the history-inference effect, the leader optimally signals less aggressively early on.

Thus, this application uncovers a novel tension between learning and performance: organizations with a better understanding of the economic environment can underperform their lesser informed counterparts. Indeed, at the core of this result is that learning is a measure of miscoordination in that ex ante (undiscounted) coordination costs satisfy

$$\mathbb{E}_0 \int_0^T (a_t - \hat{a}_t)^2 dt = \int_0^T \alpha_t^2 \underbrace{\mathbb{E}_0(\theta - \hat{M}_t)^2}_{=\gamma_t} dt = \int_0^T -\sigma_Y^2 \frac{\dot{\gamma}_t}{\gamma_t} dt = \sigma_Y^2 \ln(\gamma^o/\gamma_T),$$

which falls in γ_T.[25] To grasp some intuition, consider the public case. There, the leader could

[25]This expression for undiscounted coordination costs holds for all $r \geq 0$ and all $\sigma_X \in [0, +\infty]$.

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 no information is transmitted. It is only when the leadership introduces changes for which the organization does not know how to respond that information then gets transmitted, but this creates miscoordination. A dichotomy between learning and performance arises in that an organization's better understanding of the economic environment reflects a painful struggle to coordinate.

Our comparison of signaling coefficients is useful for understanding payoffs. Clearly, the direct effect of shutting down the public signal is negative—increased uncertainty makes coordination more difficult. The strategic effect is that, in response to more information being transmitted, the follower's belief is more volatile in the no-feedback case—but this generates even more miscoordination. Anticipating this extra responsiveness, a patient leader signals less aggressively in the beginning, explaining why the signaling coefficients cross.

4.2 Reputation: Sustaining a Reputation for Neutrality

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

$$\text{“politician”:} \ -\int_0^T e^{-rt}(a_t-\theta)^2dt - e^{-rT}\psi\hat{a}_T^2,\ \psi>0;\ \text{“media outlet”:} \ -(\hat{a}_t-\theta)^2.$$

The story behind these was that the politician/expert has a long-term career concern reflected in the lump-sum terminal payoff $-e^{-rT}\psi\hat{a}_T^2$: she wants to appear as neutral from the outlet's perspective at T due to $\hat{a}_t=\hat{M}_t$ at all times and $\mu=0$ capturing the unbiased type. To get feedback about her reputation, the politician relies on a public reporting process $dX_t=\hat{M}_tdt+\sigma_X dZ_t^X$. Will more accurate reporting—i.e., low σ_X—help the politician? How does the politician's effort to manage her reputation vary over time?

Having acquired intuition from the previous application, we forego a detailed reporting of analogous results for each corner case and instead offer a more streamlined exposition of the key economics using the interior case; we exploit the corner cases at the end of this section when we analyze the effect of feedback on learning and payoffs. We begin with a full characterization of the LMEs that arise in this setting for arbitrary discount rates and precision of the public feedback X, demonstrating the tractability of the setting studied:

Proposition 5. *Suppose that $r\geq 0$ and $\sigma_X\in(0,\infty)$. In any LME, the politician'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 $[0,T)$. Moreover, $\alpha_{3t}:=\beta_{3t}+\beta_{1t}\chi_t\in(0,1)$, and there exists $\bar{r}>0$ such that for all $r\in(0,\bar{r})$, α_3 has an interior minimum and is initially decreasing (hence non-monotonic).*

A politician with a higher bias prefers to take higher actions, all else equal ($\beta_3>0$).

But in the interest of her reputation, the politician places lower weight on her type than is myopically optimal ($0 < \beta_3 < 1$). Further, the more she believes her reputation to be biased (i.e., a large $|M_t|$), the more she wants to drive her reputation back to the neutral level $\mu = 0$; therefore, the weight β_1 on M_t is negative. And as with traditional signal jamming, the politician gets trapped into taking lower actions in response to public information that her type is high ($\beta_2 < 0$).

Toward understanding the payoff implications of higher order uncertainty in this setting, note that the ex ante expectation of $\hat{M}_T^2$ in the politician's terminal loss is exactly the amount of learning by the outlet, defined as $\gamma^o - \gamma_T$, the reduction in the posterior variance of its belief about the politician. All else equal, the politician is better off on average when the outlet learns less about her type. Now to quantify equilibrium learning and its relationship with higher order uncertainty, one must first analyze the information transmission that takes place in equilibrium.

Recall that in the public benchmark case, the signaling coefficient is given by the equilibrium weight on the type β_3 exclusively: all types would agree on $\hat{M}$ for a given public history of Y, so β_1 (the weight on $\hat{M}_t = M_t = L_t$ in this case) is not linked to information transmission. This is akin to the dismissive reaction "everyone would backtrack this way" when seeing someone doing some form of damage control. From the lens of our model, therefore, underlying this common reaction is the presence of public information that induces different politician types to take the same corrective actions, thereby conveying no information.

With higher-order uncertainty, however, the signaling coefficient is $\alpha_3 = \beta_3 + \beta_1\chi$—in contrast to the coordination game, the history inference effect now has a negative sign, introducing a new force that dampens information transmission. Intuitively, higher types reflect on their higher past actions and expect their reputations to be more biased upward, in turn forcing them to take more drastic corrective actions in a way that reduces separation.

Proposition 5 also speaks to rich dynamics of the politician's reputation management, consequent to this negative history-inference effect: as long as the discount rate is not too high, there is an intermediate point at which separation of types—captured by the sensitivity of the politician's actions to her type, α_3—is *minimized* (a phenomenon that occurs at all public histories due to $a_t = \alpha_{3t}\theta + \alpha_{2t}L_t$ in equilibrium). Indeed, as the politician begins taking actions, her past actions begin haunting her: an increasing fear of being perceived as biased, manifested in an increasingly strong history-inference effect, develops early on. As time progresses, however, and there is less time to manage the reputation, the politician beings catering to her bias more and more, and the signaling coefficient starts to increase.

Figure 2 illustrates this phenomenon for different values of the discount rate, an obviously important parameter when studying reputation effects. Naturally, low (high) values of r

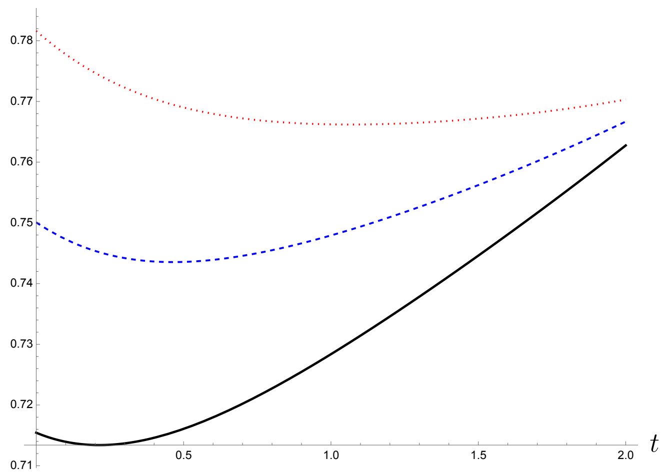

Figure 2: Nonmonotonic equilibrium signaling coefficient α_3 for varying discount rates.

frontload (backload) the incentives to moderate, so the minimum shifts to the left (right)—our condition $0 < r < \bar{r}$ in Proposition 5 simply guarantees that this minimum is interior. The robust prediction, however, is that moderation is maximized strictly *after* the beginning of the politician's career, despite there being less time to invest in the reputation.

In this game, the prior is that the politician is unbiased. Thus, if all types tied their hands, or pooled into some action, the outlet's belief would remain at the prior because no information is conveyed—the long-term goal would be achieved, but this is costly. The negative pressure on separation resulting from the history-inference effect then suggests that the presence of higher-order uncertainty potentially alleviates this form of commitment problem due to less information being transmitted. To explore these conjectures, we again exploit the analytical solutions for the public and no-feedback cases without discounting.[26]

Proposition 6. *Suppose that $r = 0$ and $\psi < \sigma_Y^2/\gamma^o$. Then for all $T > 0$, there exists a unique LME for the public case and no-feedback case. In the no-feedback case, the outlet learns less about the politician's bias, and the politician's ex ante payoff is higher.*

Ignorance can be bliss: the politician can be better off if she does not know her reputation due to imperfect reporting. A sufficient condition for this to happen is that the direct (negative) effect of being unable to tailor her actions to her reputation is not too large, which occurs when the terminal payoff is not too concave (small ψ) or the outlet's belief is not too responsive (resulting from either low initial uncertainty or a noisy private signal).[27]

[26]Note that the learning comparison in Proposition 6 does not follow immediately from the negative history-inference effect, since β_3 varies across cases.

[27]This payoff comparison would hold verbatim if instead the outlet reports on what it observes at any

5 The Main Existence Result

In this section, we transform the problem of existence of LME to a boundary value problem (BVP), and we provide time horizons for which such a problem admits a solution.

To this end, we postulate a quadratic value function for the sender

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

where $v_{i\cdot}$, $i = 0, ..., 9$ depend on time only. The Hamilton-Jacobi-Bellman (HJB) equation is

$$rV = \sup_{a'} \left\{ \tilde{u}(a', \mathbb{E}_t[\hat{a}_t], \theta) + V_t + \mu_M(a')V_m + \mu_L V_\ell + \frac{\sigma_M^2}{2}V_{mm} + \sigma_M\sigma_L V_{m\ell} + \frac{\sigma_L^2}{2}V_{\ell\ell} \right\}, \quad (17)$$

where $\tilde{u} := u + \frac{1}{2}u_{\hat{a}\hat{a}}\delta_{1t}^2\gamma_t\chi_t$, $\mu_M(a')$ and μ_L (respectively, σ_M and σ_L) denote the drifts (respectively, volatilities) in (12) and (13), and $\hat{a}_t$ is determined via (15).

Letting $a(\theta, m, \ell, t)$ denote the maximizer of the right-hand side in the HJB equation, the first-order condition (FOC) reads

$$\frac{\partial u}{\partial a}(a(\theta, m, \ell, t), \delta_{0t} + \delta_{1t}m + \delta_{2t}\ell, \theta) + \underbrace{\frac{\gamma_t\alpha_{3t}}{\sigma_Y^2}}_{dM_t/da_t} \underbrace{[v_{2t} + 2v_{5t}m + v_{7t}\theta + v_{9t}\ell]}_{V_m(\theta,m,\ell,t)} = 0. \quad (18)$$

Solving for $a(\theta, m, \ell, t)$ in (18), the equilibrium condition becomes $a(\theta, m, \ell, t) = \beta_{0t} + \beta_{1t}m + \beta_{2t}\ell + \beta_{3t}\theta$, which is a linear equation. We can then solve for (v_2, v_5, v_7, v_9) directly in terms of $\vec{\beta}$ and (γ, χ) (see (C.1)-(C.4)); the associated mapping is well defined provided that α_3 and γ never vanish, which will be the case in equilibrium. Next, we insert the resulting expressions 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 the $(\beta_0, \beta_1, \beta_2, \beta_3)$ coefficients and remaining value function coefficients. Because the pair (γ, χ) affects the law of motion of (M, L), these ODEs are coupled with (9)–(10). 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.

This procedure yields a system of ODEs for $(\beta_1, \beta_2, \beta_3, v_6, v_8, \gamma, \chi)$, to which we need to add *boundary conditions*. First, γ and χ satisfy exogenous initial conditions $\gamma_0 = \gamma^o > 0$ and $\chi_0 = 0$. Second, there are *endogenous* terminal values for the remaining variables that are determined by the static (Bayes) Nash equilibrium played at time T. To simplify expressions,

instant (i.e., $dX_t = dY_t + \sigma_X dZ_t^X$), since the information structure would be unchanged for $\sigma_X \in \{0, +\infty\}$.

we provide these for the case in which there are no terminal payoffs, $\psi \equiv 0$:

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

We now introduce our technical conditions for the existence of LME. First, we require a static equilibrium to exist for all possible histories of play, i.e., for all $\chi_T \in [0,1]$. This holds as long as $u_{a\hat{a}}\hat{u}_{a\hat{a}} < 1$ in the denominators, which is the basic requirement that myopic best replies in the static game of two-sided incomplete information intersect. Second, we will require $\alpha_{3T} \propto u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_T$ to never vanish, which is guaranteed by $u_{a\theta}(u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}) > 0$ since $\chi_T \in [0,1)$. This, in turn, will ensure that α_3 never vanishes. (And since since γ never vanishes as long as α_3 is finite, we can recover (v_2, v_5, v_7, v_9) as promised.) These assumptions are for convenience, as they allow us to discard the pathological possibility of a potential history in which the signaling game is ill-posed. We collect them next.

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

Hence, we have reduced the task of finding an LME to solving a boundary value problem (BVP) for $(\beta_1, \beta_2, \beta_3, v_6, v_8, \gamma, \chi)$. The ODEs for this boundary value problem, after a change of variables that facilitates our analysis that follows, are displayed in Appendix C.

Establishing the existence of a solution to the BVP is nontrivial because there are multiple ODEs in both directions: $(\vec{\beta}, v_6, v_8)$ is traced backward from its terminal values, while (γ, χ) is traced forward using its initial values. In BVPs where only one variable has an initial condition and the remaining variables have terminal conditions, a traditional approach is a one-dimensional shooting argument: guess the terminal value of the remaining variable, trace all variables backward, and argue via the intermediate value theorem that there is some guess for which the target (i.e., the exogenous initial condition) is hit. In our problem, however, two variables have initial conditions, so we must develop a new approach.[28]

Our approach is motivated by the observation that the problem of equilibrium existence is fundamentally a fixed-point problem: the evolution of the learning coefficients (γ, χ) depends on the signaling that takes place during the game, but the latter depends on the path of the learning coefficients because these are taken as given in the best response problem. Therefore, we translate our BVP into a fixed point equation in the space of functions (γ, χ).

Our fixed point argument is infinite dimensional. First, we choose an arbitrary pair $\lambda = (\gamma, \chi)$ in a suitable domain Λ; this domain nests all functions (γ, χ) that can be obtained as solutions to their coupled ODEs (9)–(10) 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*

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

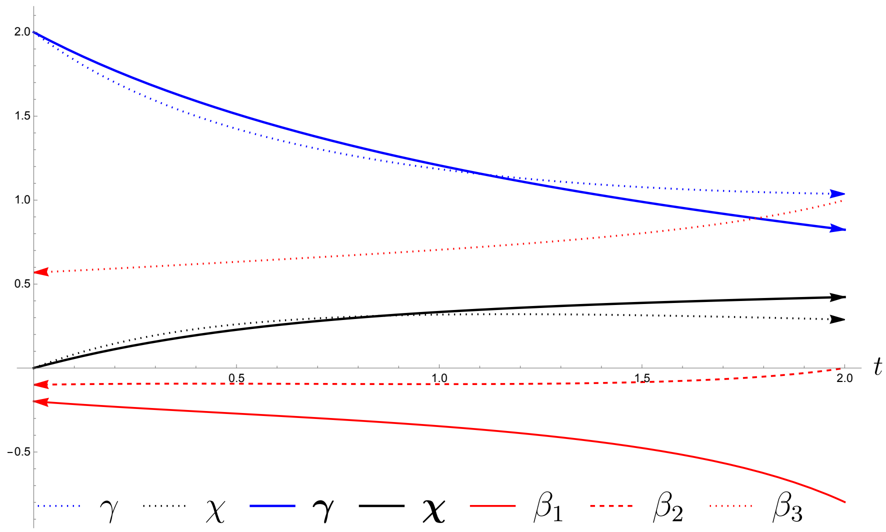

Figure 3: One iteration under our fixed-point method for the reputation game. Take as given candidate learning coefficients (dotted curves pointing to the right). Next, working backwards, generate candidate equilibrium coefficients (curves pointing to the left). Finally, use the latter functions to generate solutions to the learning ODEs (solid curves pointing to the right).

for the coefficients in the strategy consisting of their ODEs (which depend on λ) in time-reversed form and, for initial conditions, the endogenous time-T conditions (which depend on λ_T). We then derive a sufficient condition on the time horizon such that: this initial value problem has a (unique) solution for all λ in the domain; the solution satisfies the uniform bound referred to above; and the solution is continuous in λ. We then "shoot forward:" we feed the resulting (β_1, β_3) into the ODEs for (γ, χ) to get an "output" pair denoted $\boldsymbol{\lambda}$. We show that the mapping from input pairs λ to $\boldsymbol{\lambda}$ is continuous, and $\boldsymbol{\lambda}$ lies in Λ; we then apply Schauder's infinite dimensional fixed-point theorem. By construction, the fixed point and the remaining variables obtained from the procedure constitute a solution to the BVP. Figure 3 illustrates one iteration of this procedure.

We now state our main theorem for the whole class of games studied. 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 ψ is linear, or if ψ is not too concave, there exists a strictly positive and decreasing function $\gamma^o \mapsto T(\gamma^o)$ of order $\Omega(1/\gamma^o)$ such that for all $r \geq 0$ and horizons less than $T(\gamma^o)$ there exists an LME with non-trivial signaling at all times.*

The theorem states that existence is guaranteed when the curvature of the terminal payoff and time horizon are not too large, or when there is not too much initial uncertainty, for times that hold irrespective of the discount rate. (We note that if ψ is linear—which nests the case of no terminal payoffs, $\psi \equiv 0$—there are no additional restrictions on its slope.)

This is a powerful and insightful result. First, the presence of terminal payoff makes the static Nash equilibrium arising at T more complex due to "last minute" incentives; a purely technical curvature condition on ψ (that depends on parameters) 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 our fixed-point argument. Second, as the initial uncertainty γ^o increases, beliefs are naturally more responsive to new information, and hence there is more scope for manipulating beliefs; mathematically, the ODEs for the equilibrium coefficients are proportional to γ, so the uniform bounds for the associated solutions become tighter.

A natural question that arises is, why this infinite-dimensional approach? First, it is only when the ODEs for the equilibrium 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 (for continuous coefficients (β_1, β_3)) if traced forward, but not necessarily backwards starting from generic values. Thus, the approach fully exploits the basic economics of the system in each direction. Moreover, note that any bidirectional fixed-point argument necessarily employs only a subset of the ODEs at any "shooting" step, and therefore requires candidate solutions of the remaining ODEs as inputs, so any such argument must be infinite dimensional. Ours is the best avenue.[29,30]

Finally, all the steps that we have taken can be refined. We can include more general terminal payoffs, obtain better uniform bounds depending on the game at hand (we only use the degree of the polynomials involved), and potentially find horizons of existence that increase with the discount rate: this is because behavior must closer to myopic as r increases, and the equilibrium for myopic players is well defined for all T. In the next section we discuss further properties of this method in light of the existing literature, and areas for future applicability.

6 Discussion and Concluding Remarks

Receiver's myopia Allowing for a forward-looking receiver has minimal impact on our analysis. First, no additional states beyond $(t, \hat{M}, L)$ are necessary to construct LMEs. Second, the same equilibrium found in each application would arise in a non-myopic case.

[29]In an earlier version, we developed a finite-dimensional—hence, unidirectional—fixed-point argument. This method guesses the boundary values for the ODEs at one end, and shoots all the ODEs only once in the same direction. Hence, the time horizons for existence can sometimes be substantially smaller. Cetemen (2020) applies that method to an asymmetric LQG game with multi-sided private information where no states beyond first-order beliefs are needed.

[30]To exploit discounting in our proof, we perform two further modifications to the BVP before applying our fixed point argument. 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).

Consider our reputation example and suppose the outlet deviates from $\hat{a}_t = \hat{M}_t$, incurring a cost today. While the public signal is affected, the coefficient α_3 is unchanged (it is deterministic and the deviation hidden), so the future informativeness of signals is unaffected. Thus, there is no future benefit. And in the coordination game, any change in the leader's action would be offset in the next period due to the coordination motive. More generally, this logic holds for prediction problems of the form $-\frac{1}{2}(c_0 + c_1\theta + c_2 a_t - \hat{a}_t)^2$.[31]

Of course, there are settings beyond this class in which non-trivial dynamic incentives for the receiver can arise. We note two things. First, because these incentives involve affecting a public signal, they are well understood (e.g., signal jamming). Second, our methods are still equipped to handle the question of existence of LME these cases: the only difference is that the new BVP incorporates ODEs for the coefficients $(\delta_0, \delta_1, \delta_2)$ in the receiver's strategy. But our fixed point method can handle any number of ODEs in any direction.

Private-value environments and one-dimensional Shooting As argued, the presence of two learning dynamics γ and χ severely complicates the question of finding a solution to the final BVP. Economically, this is the reflection of the players potentially signaling at very different rates, so it is natural to examine environments with some symmetry. We say that an environment is of *private values* if $\hat{u}_{\hat{a}\theta} = 0$, i.e., the receiver strategically cares about the sender's action only. The players then signal at proportional rates ($\delta_1 = \hat{u}_{\hat{a}a}\alpha_3$) due to the receiver's best-response $\hat{a}_t$ being an affine function of $\hat{\mathbb{E}}[a_t]$. The environment is, therefore, strategically symmetric.

In this case, our online appendix shows that there is a one-to-one mapping between γ and χ of the form $\chi_t = \frac{c_1 c_2(1-[\gamma_t/\gamma^o]^d)}{c_1+c_2[\gamma_t/\gamma^o]^d} \in [0,1)$ for some positive constants c_1, c_2 and d. Using this relationship, the shooting problem becomes one-dimensional going backwards, and continuity arguments apply—see, for instance, Bonatti et al. (2017). A remarkable aspect of our approach is that, while the multidimensional case is both conceptually and technically considerably more challenging, the horizons for which we can guarantee the existence of LME in Theorem 1 are of the same order as in the simpler one-dimensional case. The reason is that the horizons found are pinned down, in both settings, by uniformly bounding the ODEs associated with the equilibrium coefficients exclusively (i.e., the dependence of the learning ODEs is only implicit). Consequently, our infinite-dimensional method establishes itself as the "right" extension of the one-dimensional shooting case when it comes to LQG games.

[31] Formally, it suffices to show that the myopic coefficients in our analysis solve the ODEs that govern the dynamic versions of such coefficients when the receiver is forward looking. See `spm.nb` in our websites.

Further applications of the fixed-point argument Our fixed-point technique for establishing the existence of LME accommodates multiple ODEs in each direction. This technique therefore provides a way of studying settings with asymmetries or complex information structures where multiple learning ODEs must be confronted — even if private monitoring is absent. For instance, it can be applied to asymmetric games of multi-sided incomplete information with public signals exclusively, such as dynamic oligopolies where firms' costs are drawn from different distributions. Similarly, it can be applied to reputation models—i.e., one-sided signaling—with multidimensional types. Or it can be even extended to situations in which several players affect a commonly observed signal, such as in trading models.[32]

Other signal structures Our mixed public-private information structure is useful in that it allows us to close the state space at the level of a second-order belief. Variations allowing for stochastic types, or even private signals for all players will certainly require more states, and whether an infinite regress problem is at play remains to be determined—we leave exploration of such models for future work.

Beliefs, and in particular, players' beliefs about what others believe, are at the center of game theory. When it comes to incomplete information, a large body of work has restricted to settings in which only first-order beliefs are sufficient statistics, rendering environments in which higher-order beliefs matter much less explored—and even less so if dynamics are allowed, as the further complexity arises of those beliefs themselves evolving due to both ongoing learning and strategic manipulation effects. From this perspective, our paper has uncovered a complex, yet still tractable, class of games through which we can understand how higher-order uncertainty affects strategic information transmission through actions.

To accomplish this goal, we have exploited the tractability of a linear-quadratic-Gaussian structure. While static LQG models have been exploited in many areas due to their well-known tractability, it is far less obvious what to expect in dynamic settings involving rich information structures like the ones we study. This paper demonstrates that, while there is a substantial gap in terms of difficulty when transitioning to the latter world, it is still possible to get answers, and new economic insights arise. Moreover, it is our belief that the stylized nature of these games is an asset for uncovering forces that are robust to other, more nonlinear, settings.

[32]For example, multiple learning variables arise due to a nested information structure in the trading model with two traders of Foster and Viswanathan (1994), who confront the problem numerically.

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 $x(s)<y(s)\ \forall s\in[t,T)$.

In what follows, and in the Online Appendix, we often abbreviate α_{3t} to α_t.

Proof of Lemma 1. Let L in (6) denote a process that is measurable with respect to X. Inserting (6) into (4) 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

$$\begin{aligned} dM_t &= \underbrace{(\kappa_{0t} + \kappa_{1t}a_t + \kappa_{2t}M_t)dt}_{=\mathbb{E}_t[(\kappa_{0t}+\kappa_{1t}a_t+\kappa_{2t}\hat{M}_t)dt]} + \frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2}[dX_t - (\delta_{0t} + \delta_{1t}M_t + \delta_{2t}L_t)dt] \quad (A.4) \\ \dot{\gamma}_{2t} &= 2\kappa_{2t}\gamma_{2t} + (B_t^Y)^2 - (\gamma_{2t}\delta_{1t}/\sigma_X)^2, \quad (A.5) \end{aligned}$$

with $dZ_t := [dX_t - (\delta_{0t} + \delta_{1t}M_t + \delta_{2t}L_t)dt]/\sigma_X$ is a Brownian motion from the sender's standpoint.[33] 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, (i) $\hat{\kappa}_{0t} = -\alpha_{3t}\gamma_{1t}\alpha_{0t}/\sigma_Y^2 + \alpha_{3t}\gamma_{1t}\beta_{0t}/\sigma_Y^2 - \delta_{0t}\frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2}$, (ii) $\hat{\kappa}_{1t} = \alpha_{3t}\gamma_{1t}\beta_{1t}/\sigma_Y^2 - \alpha_{3t}^2\gamma_{1t}/\sigma_Y^2 - \delta_{1t}\frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2}$, (iii) $\hat{\kappa}_{2t} = -\alpha_{3t}\gamma_{1t}\alpha_{2t}/\sigma_Y^2 + \alpha_{3t}\gamma_{1t}\beta_{2t}/\sigma_Y^2 - \delta_{2t}\frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2}$, (iv) $\hat{\kappa}_{3t} = \left[\frac{\alpha_{3t}\gamma_{1t}}{\sigma_Y^2}\right]\beta_{3t}$ and (v) $\hat{B}_t = \frac{\gamma_{2t}\delta_{1t}}{\sigma_X^2}$.

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 (6) yields the equations $\chi_t = \int_0^t R(t,s)\hat{\kappa}_{3s}ds$ and $L_t = [R(t,0)\mu + \int_0^t R(t,s)[\hat{\kappa}_{0s} + \hat{\kappa}_{2s}L_s]ds + \int_0^t R(t,s)\hat{B}_s dX_s]/[1-\chi_t]$. The validity of the construction boils down to finding a solution to the previously stated equation for χ that takes values in $[0,1)$. Indeed, when this is the case, it is easy to see that

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

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:

$$\begin{aligned} \dot{\gamma}_{1t} &= -\gamma_{1t}^2(\beta_{3t} + \beta_{1t}\chi_t)^2/\sigma_Y^2 \\ \dot{\gamma}_{2t} &= -2\gamma_{2t}\gamma_{1t}(\beta_{3t} + \beta_{1t}\chi_t)^2/\sigma_Y^2 + \gamma_{1t}^2(\beta_{3t} + \beta_{1t}\chi_t)^2/\sigma_Y^2 - (\gamma_{2t}\delta_{1t})^2/\sigma_X^2 \\ \dot{\chi}_t &= \gamma_{1t}(\beta_{3t} + \beta_{1t}\chi_t)^2(1-\chi_t)/\sigma_Y^2 - (\delta_{1t}\chi_t)(\gamma_{2t}\delta_{1t})/\sigma_X^2. \end{aligned}$$

In the proof of Lemma 3 we establish that $\chi = \gamma_2/\gamma_1 \in [0,1)$ taking the system above as a primitive. Setting $\gamma_2 = \chi\gamma_1$ in the third ODE, and writing γ for γ_1, the first and

[33] 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 $\gamma_{2t} = \gamma_{2t}^\dagger$.

third ODEs become (9)–(10). Using (i)–(v) that define $(\vec{\hat{\kappa}}, \hat{B})$ yields that (A.6) becomes $dL_t = (\ell_{0t} + \ell_{1t}L_t)dt + B_t dX_t$ where

$$(l_{0t}, l_{1t}, B_t) = [\sigma_X^2(1-\chi_t)]^{-1} \times (-\gamma_t\chi_t\delta_{0t}\delta_{1t}, -\gamma_t\chi_t\delta_{1t}(\delta_{1t}+\delta_{2t}), \gamma_t\chi_t\delta_{1t}). \tag{A.7}$$

That L_t coincides with $\mathbb{E}[\theta|\mathcal{F}_t^X]$ is proved in the Online 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 one follows from (11) using (A.7) 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 (14), (A.4)–(A.5), and the proof of Lemma 3 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. □

Proof of Lemma 3. Consider the system in $(\gamma_1, \gamma_2, \chi)$ from the proof of Lemma 1, and let $\delta_{1t} := \hat{u}_{\hat{a}\theta} + \hat{u}_{a\hat{a}}\alpha_{3t}$.[34] The local existence of a solution follows from Peano's Theorem. Suppose that the maximal interval of existence is $[0, \tilde{T})$, with $\tilde{T} \leq T$. Since the system is locally Lipschitz continuous in $(\gamma_1, \gamma_2, \chi)$ uniformly in $t \in [0, T]$, its solution over $[0, \tilde{T})$ is unique (Picard-Lindelöf). Applying the comparison theorem to the pairs $\{\gamma_1, 0\}$ and $\{\gamma_1, \gamma^o\}$, we get $\gamma_{1t} \in (0, \gamma^o]$ over $[0, \tilde{T})$. Hence, γ_2/γ_1 is well-defined, and since it solves the χ-ODE, $\chi = \gamma_2/\gamma_1$ by uniqueness. Replacing $\gamma_2 = \chi\gamma_1$ in the $\chi-$ODE then yields (10). A second application of the comparison theorem to $\{\chi, 0\}$ and $\{\chi, 1\}$ then implies $\chi \in [0, 1)$, and in turn $\gamma_2 = \chi\gamma_1 \in [0, \gamma^o)$, over $[0, \tilde{T})$. Since the solution is bounded, if $\tilde{T} < T$, it can be extended to $\tilde{T}$ by the continuity of the RHS of the system; and then subsequently extended beyond $\tilde{T}$ by Peano's theorem, a contradiction. But if $\tilde{T} = T$, it can be extended to T—the first part of the lemma holds. If $\beta_{30} \neq 0$, then $\dot{\gamma}_{10} < 0$ and $\dot{\chi}_0 > 0$, so 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

[34] All the results in this proof extend to δ_1 being a generic continuous function over $[0, T]$, the latter case arising when the receiver becomes forward looking.

theorem, these strict inequalities hold up to time T. □

Appendix B: Proofs for Section 4

In this section, we highlight the main steps for proving Proposition 1; the no-feedback case and the corner cases of the reputation game follow similar steps. We also prove Proposition 2. All other results and assertions made in Section 4 are proved in the online appendix.

Proof of Proposition 1. 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 $\beta_{1t} + \beta_{3t} \neq 0$, $t \in [0, T]$. 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 $rV = \sup_{a\in\mathbb{R}}\{\frac{1}{4}[-(a-\theta)^2 - (a-\hat{a}_t)^2] + \frac{\beta_{3t}\gamma_t}{\sigma_Y^2}[a - \beta_{0t} - (\beta_{1t}+\beta_{3t})m]V_m + \frac{\beta_{3t}^2\gamma_t^2}{2\sigma_Y^2}V_{mm} + V_t\}$. We guess a quadratic solution $V(\theta, m, t) = v_{0t} + v_{1t}\theta + v_{2t}m + v_{3t}\theta^2 + v_{4t}m^2 + v_{5t}\theta m$, from which the FOC in the HJB reads $0 = -\frac{1}{2}(\beta_{0t} + \beta_{1t}m + \beta_{3t}\theta - \theta) - \frac{1}{2}\beta_{3t}(\theta - m) + (\beta_{3t}\gamma_t/\sigma_Y^2)[v_{2t} + 2mv_{4t} + \theta v_{5t}]$ when the maximizer is $a^* := \beta_{0t} + \beta_{1t}m + \beta_{3t}\theta$. Provided $\beta_{3t}, \gamma_t > 0$ (as we verify later), $(v_{2t}, v_{4t}, v_{5t}) = \left(\frac{\sigma_Y^2\beta_{0t}}{2\beta_{3t}\gamma_t}, \frac{\sigma_Y^2(\beta_{1t}-\beta_{3t})}{4\beta_{3t}\gamma_t}, \frac{\sigma_Y^2(2\beta_{3t}-1)}{2\beta_{3t}\gamma_t}\right)$, due to the FOC holding for all $(\theta, m, t) \in \mathbb{R}^2 \times [0, T]$. And since $v_{iT} = 0$ for $i \in \{0, \ldots, 5\}$, we deduce that $(\beta_{0T}, \beta_{1T}, \beta_{3T}) = (0, 1/2, 1/2)$.

Inserting a^* into the HJB equation, and using the previous expressions for (v_{2t}, v_{4t}, v_{5t}) to replace $(v_{2t}, v_{4t}, v_{5t}, \dot{v}_{2t}, \dot{v}_{4t}, \dot{v}_{5t})$, yields an equation in $\vec{\beta} := (\beta_0, \beta_1, \beta_3)$ and $\dot{\vec{\beta}}$. Grouping by coefficients (θ, m, θ^2,..., etc.) in the latter yields the ODEs

$$(\dot{\beta}_{0t}, \dot{\beta}_{1t}, \dot{\beta}_{3t}) = \left(2r\beta_{0t}\beta_{3t}, \beta_{3t}\left[r(2\beta_{1t} - 1) + \frac{\beta_{1t}\beta_{3t}\gamma_t}{\sigma_Y^2}\right], \beta_{3t}\left[r(2\beta_{3t} - 1) - \frac{\beta_{1t}\beta_{3t}\gamma_t}{\sigma_Y^2}\right]\right). \quad \text{(B.1)}$$

The existence of an LME thus reduces to the BVP defined by $\dot{\gamma}_t = -\gamma_t^2\beta_{3t}^2/\sigma_Y^2$ and (B.1), with $\gamma_0 = \gamma^o$ and $(\beta_{0T}, \beta_{1T}, \beta_{3T}) = (0, 1/2, 1/2)$. Solving this BVP delivers the remaining v_i, as their ODEs are uncoupled and linear in themselves (see Online Appendix).

To show existence, we transform this BVP 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 online appendix. □

Proof of Proposition 2. 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 $\beta_1+\beta_2+\beta_3 = 1$.
Non-monotonicity: Note that $\alpha_T = \frac{1}{2-\chi_T} > 0$, $\beta_{1T} = \frac{1}{2(2-\chi_T)}$, and $v_{8T} = 0$. The α-ODE is

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

Applying the comparison theorem to α (backward), we have $\alpha > 0$ and thus $\beta_{30} = \alpha_0 > 0$. Using the strict inequalities in Lemma 3, $\dot{\alpha}_T = -\frac{2\alpha_T^3\gamma_T\chi_T}{\sigma_X^2\sigma_Y^2(1-\chi_T)}\left\{\sigma_Y^2\chi_T\frac{1-\chi_T}{2(2-\chi_T)}\right\} < 0$ (i.e. α is eventually decreasing) and $\alpha_T > 1/2$. Now at $t = 0$, we have $\chi_0 = 0$ and thus $\dot{\alpha}_0 = r\alpha_0(2\alpha_0 - 1)$; it follows that $\dot{\alpha}_0 > 0$ iff $\alpha_0 > \frac{1}{2}$. Consider two cases: (i) $\alpha_0 > \frac{1}{2}$ and (ii) $\alpha_0 \leq \frac{1}{2}$. In case (i), we have $\dot{\alpha}_0 > 0$. In case (ii), we have $\alpha_T > \frac{1}{2} \geq \alpha_0$, so by the mean value theorem, $\dot{\alpha}_t > 0$ for some $t \in (0,T)$. In either case, since $\dot{\alpha}_T < 0$, α is non-monotonic. □

Appendix C: Proofs for Section 5

Overview of approach Our overall proof strategy consists of reducing the HJB equation (17) subject to the equilibrium condition (18) 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 3. 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).[35] Part (iii) reflects that the dynamic for L carries a denominator of that form; by Lemma 3, 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).[36]

[35] 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.

[36] 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.

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)\left\{\beta_{1t}^2\hat{u}_{a\hat{a}}[2u_{a\hat{a}} + u_{\hat{a}\hat{a}}\hat{u}_{a\hat{a}}]\right. \\
&\quad \left.+\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]\right\} \\
\dot{\tilde{v}}_{8t} &= \tilde{v}_{8t}[r + \delta_{1t}^2\gamma_t\chi_t/\sigma_X^2] - \gamma_t\left\{(\tilde{\beta}_2 + \beta_{1t})[u_{a\theta} + u_{\hat{a}\theta}\hat{u}_{a\hat{a}}] - \beta_{1t}\beta_{3t}\right\} \\
\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 \left\{\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]\right. \\
&\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^2u_{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 \left.-\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}]\right\}. \\
\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 \left\{\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}]\right. \\
&\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^2u_{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 \left.+\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_tu_{a\hat{a}}u_{a\theta}\hat{u}_{a\hat{a}}(1 - 2u_{a\hat{a}}\hat{u}_{a\hat{a}})]\right\} \\
\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 \\
&\quad \left\{\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}})]\right. \\
&\quad -\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) \\
&\quad +\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) \\
&\quad +\delta_{1t}^2\alpha_t\chi_t\tilde{v}_{8t}(\beta_{3t} + \chi_tu_{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 \\
&\quad \left.+\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}])]\right\} \\
\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].
\end{aligned}$$

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 $(\beta^m_{1T}, \tilde{\beta}^m_{2T}, \beta^m_{3T}, 0, 0)$—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* $(\gamma, \chi, \beta^c_1, \tilde{\beta}^c_2, \beta^c_3, \tilde{v}_6, \tilde{v}_8)$, *where* $(\beta^c_1, \tilde{\beta}^c_2, \beta^c_3)$ *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 $-r\frac{\alpha}{\alpha^m}(\beta_i - \beta^m_i)$ 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., $x^c_{it} := x_{it} - x^m_{it}$ for $x \in \{\beta_1, \tilde{\beta}_2, \beta_3\}$, and work with the ODEs of $(\beta^c_1, \tilde{\beta}^c_2, \beta^c_3, \tilde{v}_6, \tilde{v}_8)$ in backward form.[37]

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 $(\beta_1, \tilde{\beta}_2, \beta_3) = (\beta^c_1 + \beta^m_1, \tilde{\beta}^c_2 + \tilde{\beta}^m_2, \beta^c_3 + \beta^m_3)$, (ii) $(\beta^m_1, \tilde{\beta}^m_2, \beta^m_3)$ are functions of χ and are independent of r, (iii) $(\dot{\beta}^m_1, \dot{\tilde{\beta}}^m_2, \dot{\beta}^m_3)$ carry a factor of γ through $\dot{\chi}$, and (iv) $\alpha^m_t = \frac{u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t}{1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t}$. (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 $(\beta^c_1, \tilde{\beta}^c_2, \beta^c_3, \tilde{v}_6, \tilde{v}_8)$ (and unless otherwise stated, we always refer to the backward system when invoking this tuple). Also, let $\vec{\beta}^c := (\beta^c_1, \tilde{\beta}^c_2, \beta^c_3)$.

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*

$$\dot{x}^c_t = -r x^c_t \frac{\alpha_t}{\alpha^m_t} + \frac{\gamma_t h_x(\vec{\beta}^c, \tilde{v}_{6t}, \tilde{v}_{8t}, \chi_t)}{\sigma^2_X \sigma^2_Y (u_{a\theta} + u_{a\hat{a}}\hat{u}_{\hat{a}\theta}\chi_t)^{n_{1,x}} (1 - u_{a\hat{a}}\hat{u}_{a\hat{a}}\chi_t)^{n_{2,x}} (1 - u_{a\hat{a}}\hat{u}_{a\hat{a}})^{n_{3,x}}}$$

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

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.*[38] *The initial conditions are* $(\vec{\beta}^c_0, \tilde{v}_{60}, \tilde{v}_{80}) = (0, 0, 0, 0, 0)$.

In particular, notice that (i) the terms not containing r continue scaling with γ, (ii)

[37]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 1 and 3. 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.

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

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 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 (9)-(10); and the ODE (C.6) of $\tilde{\alpha}$.[39] 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.[40]

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 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

[39] For consistency, the α_t in the r-term in (C.6) and in (9)-(10) must be written as $(\beta_{1t}^c + \beta_{1t}^m)\chi_t + \beta_{3t}^c + \beta_{3t}^m$.

[40] 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.

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

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

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 $(\mathrm{IVP}^{\mathrm{bwd}}(\hat{\lambda}))$ 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.[41] For arbitrary $K > 0$, we now establish sufficient conditions for $(\mathrm{IVP}^{\mathrm{bwd}}(\hat{\lambda}))$ 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 $(\mathrm{IVP}^{\mathrm{bwd}}(\hat{\lambda}))$ exists over $[0, T]$ and satisfies $||\mathbf{b}_i(\cdot;\lambda)||_\infty < K$ for all $i \in \{1, \dots, 5\}$. 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 $||\mathbf{b}_i(\cdot;\lambda)||_\infty < K$ 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

[41] 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 $\Lambda(\rho + K)$.

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 $j \in \{4,5\}$,

$$|\mathbf{b}_{it}| \leq \int_0^t e^{-r\int_s^t \frac{\tilde{\alpha}_u}{\alpha_u^m}du}\gamma^o h_i(K)ds \leq \int_0^t \gamma^o h_i(K)ds = t\gamma^o h_i(K)$$

$$|\mathbf{b}_{jt}| \leq \int_0^t e^{-\int_s^t (r+\hat{\gamma}_u R_j(\mathbf{b}_u,\hat{\chi}_u))du}\gamma^o h_j(K)ds \leq \int_0^t \gamma^o h_j(K)ds = t\gamma^o h_j(K),$$

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 $r \geq 0$; also $T(\gamma^o; K) \in \Omega(1/\gamma^o)$.[42] □

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 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 Online 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*

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

[42] 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 $\mathbf{b}_i$, $i \in \{1,\dots,5\}$.

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

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

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 $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$

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

consisting of the two (forward) learning ODEs (9)-(10) 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 3 gives existence and uniqueness of a solution to ($\mathrm{IVP}^{\mathrm{fwd}}(q(\lambda))$) 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, $\boldsymbol{\lambda}_0 := (\boldsymbol{\lambda}_{10}, \boldsymbol{\lambda}_{20}) = (\gamma^o, 0)$, 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, we have that $|\dot{\boldsymbol{\lambda}}_{1t}| = |-\frac{\boldsymbol{\lambda}^2_{1t}([q_2(\lambda)]_t + [q_1(\lambda)]_t\boldsymbol{\lambda}_{2t})^2}{\sigma^2_Y}| \leq (\gamma^o)^2(2[\rho+K])^2/\sigma^2_Y$ for all $t \in [0, T]$. Similarly, from the χ-ODE, $|\dot{\boldsymbol{\lambda}}_{2t}| \leq \gamma^o\left[(2[\rho+K])^2/\sigma^2_Y + (|\hat{u}_{\hat{a}\theta}| + |\hat{u}_{a\hat{a}}|(2[\rho+K]))^2/\sigma^2_X\right]$. 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 ($\mathrm{IVP}^{\mathrm{fwd}}(q(\lambda))$) 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 ($\mathrm{IVP}^{\mathrm{bwd}}(\hat{\lambda})$) 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 is proportional to $1/\gamma^o$. □

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 $\alpha_t := \beta_{1t}\chi_t + \beta_{3t}$, $\beta_{1t} := \beta^c_{1t} + \beta^m_{1t}$ and $\beta_{3t} := \beta^c_{1t} + \beta^m_{1t}$—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^c_{1t}, and β^c_{3t} yields that α in backward form satisfies

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

Relative to (C.6), therefore, the r-term as well as the last fraction multiplying α coincide. Call this last term C_t—a continuous function of time—and observe that $p := \alpha - \tilde{\alpha}$ satisfies the ODE $\dot{p}_t = p_t C_t$ with initial condition $p_0 = 0$ due to $\alpha_0 = \tilde{\alpha}_0 = \alpha^m_0$ (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 online 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* $\underline{\psi} \in [-\infty, 0)$ *and* $T(\gamma^o) \in \Omega(1/\gamma^o)$ *such that if* $\psi_{\hat{a}\hat{a}} \in (\underline{\psi}, 0]$ *and* $T < T(\gamma^o)$*, a linear Markov equilibrium exists for all* $r \geq 0$.

References

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.

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. 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.

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.

Grant, R. (1996): "Toward a knowledge-based theory of the firm," *Strategic Management Journal*, 17, 109–122.

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

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

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.

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

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

Marschak, J. (1955): "Elements for a theory of teams," *Management Science*, 1, 127–137.

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.

Milgrom, P. R. and J. D. Roberts (1992): *Economics, organization and management*, Prentice-Hall.

Nonaka, I. (1991): "The knowledge-creating company," *Harvard Business Review*, 69, 96–104.

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.

Sugaya, T. (2021): "The Folk Theorem in Repeated Games with Private Monitoring," *The Review of Economic Studies*.

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 (2019): "Back-Running: Seeking and Hiding Fundamental Information in Order Flows," *Review of Financial Studies*, forthcoming.

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