Theoretical Economics 16 (2021), 73–99 1555-7561/20210073

Convergence in models of misspecified learning

Paul Heidhues
Düsseldorf Institute for Competition Economics, Heinrich-Heine Universität Düsseldorf

Botond Kőszegi
Department of Economics and Business, Central European University

Philipp Strack
Department of Economics, Yale University

We establish convergence of beliefs and actions in a class of one-dimensional learning settings in which the agent's model is misspecified, she chooses actions endogenously, and the actions affect how she misinterprets information. Our stochastic-approximation-based methods rely on two crucial features: that the state and action spaces are continuous, and that the agent's posterior admits a one-dimensional summary statistic. Through a basic model with a normal–normal updating structure and a generalization in which the agent's misinterpretation of information can depend on her current beliefs in a flexible way, we show that these features are compatible with a number of specifications of how exactly the agent updates. Applications of our framework include learning by a person who has an incorrect model of a technology she uses or is overconfident about herself, learning by a representative agent who may misunderstand macroeconomic outcomes, and learning by a firm that has an incorrect parametric model of demand.

Keywords. Misspecified model, Bayesian learning, convergence, Berk–Nash equilibrium.

JEL classification. D83, D90.

1. Introduction

A significant literature in statistics studies inferences by an observer who has a misspecified model of the world and receives exogenous—and in most papers independent—signals. Some counterexamples notwithstanding, it has been shown that in this setting, the observer's beliefs converge under weak conditions (Berk 1966, Bunke and Milhaud 1998, Shalizi 2009). In many or most economic applications, however, a person is not

Paul Heidhues: heidhues@dice.hhu.de
Botond Kőszegi: botondkoszegi@gmail.com
Philipp Strack: philipp.strack@yale.edu
We are grateful to three anonymous referees for excellent suggestions. Heidhues and Kőszegi thank the European Research Council for financial support under Grants 313341 and 788918. Part of this research was carried out while all authors visited briq Institute on Behavior and Inequality, whose hospitality is gratefully acknowledged.

© 2021 The Authors. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at https://econtheory.org. https://doi.org/10.3982/TE3558

only a passive observer of her economic environment, but she also chooses actions based on her beliefs. Whenever this is the case, the action could affect what information she observes, making signals endogenous and typically nonindependent. For this type of learning with misspecification, little is known about the convergence of beliefs, and previous work has identified economically natural conditions under which beliefs do not converge (e.g., Nyarko 1991, Fudenberg et al. 2017).

In this paper, we establish convergence of beliefs and actions in models of misspecified learning with endogenous actions in which the state and action spaces are continuous and the agent's beliefs admit a one-dimensional summary statistic. As we illustrate through a few examples, this class of models is sufficiently general to cover a number of economically relevant settings.

After discussing related literature in Section 2, we present our framework in Section 3. In each period $t \in \{1, 2, 3, \dots\}$, the agent produces observable output $q_t = Q(a_t, b_t)$, which depends on her action $a_t \in (\underline{a}, \overline{a})$ and external conditions b_t beyond her control. The states b_t are independent normally distributed random variables with mean equal to a fixed fundamental Θ, about which the agent's prior is normal. The agent attempts to learn about Θ from her outputs and to adjust her action optimally in response, aiming to maximize discounted expected output. Crucially, she has a misspecified model: she believes that output is determined according to $\tilde{Q}(a_t, b_t)$. To guarantee that the agent can always find an explanation for her observations, we assume that Q and $\tilde{Q}$ are strictly monotonic in b_t, and that fixing a_t, the range of $\tilde{Q}(a_t, b_t)$ includes the range of $Q(a_t, b_t)$.

In Section 4, we apply tools from stochastic approximation theory to show that under economically weak conditions, the agent's beliefs converge with probability 1 to a point belief and her actions also converge. Within our setting, this convergence result supersedes previous ones in the literature, including those of Berk (1966), Bunke and Milhaud (1998), and Shalizi (2009), where the agent does not take decisions based on her beliefs, Esponda and Pouzo (2016a), where convergence to stable beliefs is established only for nearby priors and in expectation approximately optimal actions, and Heidhues et al. (2018), where the convergence proof requires strong supermodularity restrictions on the environment. Because of our assumption that noise is inside the output function, for any output q_t, there is a unique state $\tilde{b}_t$ that the agent believes must have generated q_t. This implies that the agent's updating follows a normal–normal structure, and her posterior belief in period t can be described by her mean belief $\tilde{\theta}_t$, a one-dimensional summary statistic. Since the agent puts less and less weight on her new observations as time goes by, changes in her beliefs slow down, so the same change requires observing more and more signals. Due to a version of the law of large numbers, this means that changes in her beliefs can in the limit be well approximated by a deterministic process governed by an ordinary differential equation (ODE). As the solution to this ODE converges, so do the agent's beliefs. We conclude that the agent's beliefs converge to a point where the external conditions $\tilde{b}_t$ that she thinks she has observed on average equal her belief about the fundamental.

In Section 5, we identify several variants of our model to which we can extend our stochastic-approximation-based methods. For convergence as well as the determination of limiting beliefs, what matters is that we can reduce the agent's beliefs to a one-dimensional summary statistic and that his belief moves in the direction of her perceived signal $\tilde{b}_t$ by an amount on the order of $1/t$. This condition is satisfied by Bayesian updating if the b_t are exponentially distributed, since this also leads to a linear updating rule. Furthermore, it is sufficient for the agent's subjective model of the b_t to be normal or exponential. The true model can be any distribution with mean zero and finite variance, and a variant of our convergence result also obtains if the agent's objective differs from what she observes.

In Section 6, we identify a few economic applications of our abstract framework. The agent might be a representative consumer seeking to adjust her behavior to macroeconomic conditions she misunderstands. The agent might be learning about a technological parameter Θ—such as the usefulness of a new fertilizer—and choosing its use a_t with a misspecified model. The agent might be a firm trying to learn about a demand parameter Θ and set optimal prices a_t with a misspecified functional form for demand. As in a correctly specified model, beliefs converge with probability 1 in each case, but they also display properties that could not occur in a correctly specified model. First, in all of our examples, the limiting belief is false and the limiting action is, therefore, suboptimal. In our price-setting model, for instance, the firm's unique limiting belief about the level of demand is correct, but its belief about the elasticity is not, so the price it chooses is suboptimal. Second, in a variant of this model inspired by Fudenberg et al. (2017) and Nyarko (1991), we demonstrate that without assuming continuous state and action spaces, convergence can easily fail. Third, we illustrate that there may be multiple possible limiting beliefs. In the macroeconomic context, the agent becomes—depending on her initial observations—either irrationally exuberant or irrationally pessimistic about the economy. Similarly, the agent may under- or overestimate the effectiveness of fertilizer, but in either case, she underestimates its return net of usage costs.

Section 7 concludes by discussing other settings in which our methods can help to characterize long-run beliefs.

2. Related literature

The statistics literature on learning with misspecified models, such as Berk (1966), Bunke and Milhaud (1998), and Shalizi (2009), identifies conditions under which beliefs converge. In these models, the observer does not take endogenous actions. There is a growing literature on learning with misspecified models in which actions are endogenous to beliefs and at the same time actions affect how the agent (mis)interprets information. Most previous work establishes convergence under very restrictive conditions, and there are examples in which convergence does not obtain. Indeed, it seems widely recognized among researchers that proving convergence of beliefs is in most cases notoriously difficult.

In Heidhues et al. (2018), we study an overconfident agent's often self-defeating learning process and behavior when she chooses actions endogenously based on her

beliefs. Our earlier paper assumes that noise is outside the production function ($q_t = Q(a_t, \Theta) + \epsilon_t$), while in the current paper, noise is inside the production function ($q_t = Q(a_t, b_t) = Q(a_t, \Theta + \epsilon_t)$). Our earlier paper establishes convergence of beliefs through a different method that (i) requires restricting attention to situations in which there can only be one limiting belief and (ii) relies heavily on the particular structure of Q and $\tilde{Q}$ (e.g., that overconfidence distorts the subjectively optimal action away from the optimal one in the same direction as an underestimation of the state). By assuming that noise is inside the production function, we can establish convergence for a far more general class of technologies and misspecifications. In particular, we show convergence even when the limiting belief is not unique, which is a natural and economically relevant feature, for instance, in the examples on equilibrium feedback and fertilizer use.

Given the above difference, the question arises as to which specification of noise is more appropriate. While most previous researchers assumed that noise is outside the production function, they did so mostly for tractability or by habit. In our case, having the noise inside the production function leads to a more tractable model. Furthermore, this specification may be more realistic in some situations. For instance, if the production function depends on the agent's action and the productivity of his teammate, then any noise in the productivity of the teammate must go inside the production function.

Esponda and Pouzo (2016a) develop a general framework for studying games in which players have misspecified models, and Esponda and Pouzo (2016b) extend the framework to general dynamic single-agent Markov decision problems with nonmyopic agents. Building on Berk (1966), Esponda and Pouzo (2016a) establish that if actions converge, then beliefs converge to a limit at which a player's predicted distribution of outcomes is closest to the actual distribution. Our limiting beliefs have a similar property, but we derive significantly stronger results on the convergence of beliefs.

Fudenberg et al. (2017) completely characterize the conditions under which beliefs converge when the agent has a two-point prior and the signal is Brownian. Due to the discrete nature of the state space, beliefs need not converge.[1] For environments with a finite action space, Esponda et al. (2019) use stochastic approximation techniques to study the convergence of frequencies of actions (rather than beliefs) when the agent learns with a misspecified model. Frick et al. (2020) establish sufficient conditions for (non-)convergence of beliefs for environments with a finite, but not binary, state space, and Fudenberg et al. (2020) provide such conditions for a finite action space. Finally, in the context of optimal stopping problems, He (2018) models the gambler's fallacy as a dogmatic belief in negative correlation between consecutive signals and derives long-run beliefs about the signal-generating distribution.

3. Learning environment

In this section, we introduce our framework, and perform a few preliminary steps of analysis.

[1] More distantly related, Bohren (2016) and Bohren and Hauser (2020) study long-run beliefs in misspecified social learning environments with a two-point prior and short-lived agents. Also in the context of social learning, Frick et al. (2019) analyze the robustness of long-run beliefs with respect to small amounts of misspecification about others' preferences.

3.1 *Setup*

The objective environment In each period $t \in \{1, 2, 3, \dots\}$, the agent produces observable output $q_t \in \mathbb{R}$ according to the twice differentiable output function $Q(a_t, b_t)$, which depends on her action $a_t \in (\underline{a}, \overline{a}) = A$ and an unobservable external state $b_t \in \mathbb{R}$ beyond her control. We allow $\underline{a}$ and $\overline{a}$ to be finite or infinite. We assume that

$$b_t = \Theta + \epsilon_t,$$

where $\Theta \in \mathbb{R}$ is an underlying fixed fundamental and the ϵ_t are independent normally distributed random variables with mean zero and variance σ^2. We denote by $\phi(\cdot; \theta, \eta^2)$ the density of the normal distribution with mean θ and variance η^2.

The agent's subjective beliefs The agent's prior is that Θ is distributed normally with mean $\tilde{\theta}_0$ and variance v_0. While the agent understands the basic environment correctly, she has a misspecified model regarding output: she believes that output is being produced according to the twice differentiable function $\tilde{Q}(a_t, b_t)$. Given her model, the agent updates her beliefs about the fundamental in a Bayesian way and chooses her action in each period to maximize perceived discounted expected output.

We impose a few important conditions on the misspecified and true models. First, for any action $a \in (\underline{a}, \overline{a})$ and any state $b \in \mathbb{R}$, there exists a subjective state $\tilde{b}$ such that $Q(a, b) = \tilde{Q}(a, \tilde{b})$. This guarantees that the agent can find an explanation for any output she observes. Without such an assumption, Bayes' rule does not specify beliefs after some histories. Second, the agent believes that an increase in the state always affects output in the same direction, and we normalize this direction to be positive: $\tilde{Q}_b > 0$.[2] This ensures that the agent always infers a unique external state. Third, output also changes monotonically with the state (i.e., $Q_b > 0$ everywhere or $Q_b < 0$ everywhere). This implies that an agent with a correctly specified model can infer the realized state for any action she took.

We also make economically weak technical assumptions on the misspecified model. First, $\tilde{Q}_{aa} < 0$ and $\lim_{a \to \underline{a}} \tilde{Q}_a(a, b) > 0 > \lim_{a \to \overline{a}} \tilde{Q}_a(a, b)$ for all b. This guarantees that there is always a unique optimal action when the agent is subjectively certain about the state. Second, for any action a, the functions $\tilde{Q}(a, b)$, $\tilde{Q}_a(a, b)$, $\tilde{Q}_{ab}(a, b)$, and $\tilde{Q}_{aa}(a, b)$ are integrable with respect to admissible distributions over b.[3] This implies that the agent's problem is well defined, and we can analyze the agent's choices using first-order methods and the implicit function theorem.

3.2 *Preliminaries*

We begin the analysis of our model by noting a few basic properties. After observing output q_t generated by the realized state b_t, the agent believes that the realized state was $\tilde{b}_t$ satisfying

$$\tilde{Q}(a_t, \tilde{b}_t) = q_t = Q(a_t, b_t). \tag{1}$$

[2]Throughout we use sub-indices to denote partial derivatives.

[3]That is, we have that $\int_{\mathbb{R}} |\tilde{Q}(a, b)| \phi(b; \theta, \eta^2)\, db < \infty$ for any θ and η^2, and the identical condition for the other functions.

By assumption, there is a $\tilde{b}_t$ satisfying (1), and since $\tilde{Q}_b > 0$, it is unique. Therefore, the agent believes that whatever action she chooses, she can infer a unique signal $\tilde{b}_t$ about Θ. Since she believes that $\tilde{b}_t$ is normally distributed with mean Θ and variance σ^2 independent of her action, she believes that all actions are equally informative. This implies that she chooses her action in each period to maximize that period's perceived expected output. In addition, given her prior that the state is distributed according to $\mathcal{N}(\tilde{\theta}, v_0)$ and her belief that $\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_t$ are independent and identically distributed (i.i.d.) according to $\mathcal{N}(\Theta, \sigma^2)$, at the end of period $t \geq 1$, her posterior belief is that Θ is normally distributed with mean[4]

$$\tilde{\theta}_t = \frac{\dfrac{\sigma^2}{v_0}\tilde{\theta}_0 + \displaystyle\sum_{s=1}^{t} \tilde{b}_s}{\dfrac{\sigma^2}{v_0} + t} \tag{2}$$

and variance

$$v_t = \frac{1}{v_0^{-1} + t\sigma^{-2}}.$$

Hence, when choosing her action in period t, the agent believes that Θ is normally distributed with mean $\tilde{\theta}_{t-1}$ and variance v_{t-1}.

We denote by $M : A \times \mathbb{R} \to \mathbb{R}$ the agent's subjective expected payoff when she is sure about the fundamental:

$$M(a, \tilde{\theta}) = \int \tilde{Q}(a, \tilde{b})\phi\big(\tilde{b}; \tilde{\theta}, \sigma^2\big)\, d\tilde{b}.$$

Lemma 1 (Optimal actions). *The optimal action taken by the agent in period t is unique, depends only on t and the mean of her posterior belief $\tilde{\theta}_{t-1}$, and is given by*

$$a^*(t, \tilde{\theta}_{t-1}) = \operatorname*{argmax}_{a \in A} \int M(a, x)\phi(x; \tilde{\theta}_{t-1}, v_{t-1})\, dx.$$

Furthermore, there is a unique action that the agent perceives as optimal if she is certain that the fundamental is $\tilde{\theta}_{t-1}$, which is given by

$$a^*(\tilde{\theta}_{t-1}) = \operatorname*{argmax}_{a \in A} M(a, \tilde{\theta}_{t-1}).$$

Both $a^*(t, \tilde{\theta}_{t-1})$ and $a^*(\tilde{\theta}_{t-1})$ are differentiable with respect to $\tilde{\theta}_{t-1}$.

To keep track of how the agent misperceives the signals, we define $\tilde{b}(b, a)$ as the state $\tilde{b}$ that solves (1), i.e., the state the agent perceives as a function of the true state b and her action a. We make three weak assumptions that bound the agent's misinference and the sensitivity of her misinference and behavior.

[4]See, for example, https://en.wikipedia.org/wiki/Conjugate_prior.

Assumption 1. (i) There exists a constant $\Delta > 0$ such that $|b - \tilde{b}(b, a)| \leq \Delta$ for all b, a.

(ii) The function $\tilde{b}_a$ is bounded.

(iii) There exist constants $d, k > 0$ such that for any t and any $\tilde{\theta}$, we have $|a^*(t, \tilde{\theta}) - a^*(\tilde{\theta})| \leq \frac{1}{t^k} d$.

Part (i) imposes that the agent's misinference about the realized external state is bounded by Δ, guaranteeing that in the long run her beliefs are in a bounded interval. Part (ii), a technical condition, bounds the derivative of the misinference with respect to the chosen action. Part (iii) bounds the effect of the agent's uncertainty about the fundamental—which in period t is of order $1/t$—on her action—that it should be of order $1/t^k$ and hence vanish over time. This assumption, easy to check in specific applications, allows us to approximate the agent's time-sensitive optimal action $a^*(t, \tilde{\theta})$ with her confident action $a^*(\tilde{\theta})$. The following lemma identifies sufficient conditions on primitives for Assumption 1 to be satisfied. As our fertilizer example in Section 6 shows, these conditions are not necessary.

Lemma 2 (Sufficient conditions on primitives). *(i) If $|\tilde{Q}(a, b) - Q(a, b)|$ is uniformly bounded and $\tilde{Q}_b$ is uniformly bounded away from zero, then part (i) of Assumption 1 holds.*

(ii) If $\tilde{Q}_{\tilde{b}}(a, \tilde{b})$ is uniformly bounded from below, and $|Q_a(a, b)|$ and $|\tilde{Q}_a(a, \tilde{b})|$ are uniformly bounded, then part (ii) of Assumption 1 holds.

(iii) If $\tilde{Q}_{aa}(a, \tilde{b})$ is uniformly bounded away from zero and $|\tilde{Q}_{ab}(a, \tilde{b})|$ is uniformly bounded, then part (iii) of Assumption 1 holds.

4. Convergence of the agent's beliefs

Establishing convergence is technically challenging because of the endogeneity of actions: as the agent updates her belief, she changes her action, thereby changing the objective distribution of the perceived signal she uses to update her belief. This means that we cannot apply results from the statistical learning literature, such as those of Berk (1966) and Shalizi (2009), where the observer does not choose actions based on her beliefs. At the same time, several features of our model facilitate a convergence argument. The assumption of a normal prior and normal signals ensures that (i) beliefs concentrate for any sequence of signals and (ii) the expected changes in the agent's beliefs vanish on the order of $1/t$. Property (i) implies that eventually the agent's mean beliefs are sufficient to describe her beliefs and actions arbitrarily well, so it suffices to analyze the dynamics of mean beliefs. Property (ii) then allows us to use results from stochastic approximation theory to establish that the agent's mean beliefs converge.[5]

[5] Kushner and Yin (2003) introduce and discuss classic results in stochastic approximation theory, which dates back to Robbins and Monro (1951) and Kiefer and Wolfowitz (1952), as well as modern developments in the field.

We now turn to the formal argument. Using (2), the dynamics of the agent's beliefs can be written as

$$\tilde{\theta}_{t+1} = \tilde{\theta}_t + \gamma_t[\tilde{b}_{t+1} - \tilde{\theta}_t], \quad \text{where } \gamma_t = \frac{1}{t+1+\dfrac{\sigma^2}{v_0}}.$$

We define the function $g : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ as the objective expectation of $\tilde{b}_{t+1} - \tilde{\theta}_t$:

$$g(t,x) = \mathbb{E}\big[\tilde{b}\big(b_{t+1}, a^*(t+1,x)\big)\big] - x. \tag{3}$$

The above expectation is the one of an outside observer who knows the true fundamental Θ and the agent's subjective belief $\tilde{\theta}_t$, so that, being able to deduce the action $a^*(t+1, \tilde{\theta}_t)$, she knows the distribution of $\tilde{b}_{t+1}$. The function g can be thought of as the agent's mean surprise regarding the fundamental.

As we are interested in how the agent's subjective beliefs are updated in the limit, we define

$$g(x) = \lim_{t\to\infty} g(t,x) = \mathbb{E}\big[\tilde{b}\big(b_{t+1}, a^*(x)\big)\big] - x.$$

We impose that g does not touch the horizontal axis without crossing.[6]

Definition 1 (Stationary point). A point $\tilde{\theta}$ is *stationary* if $g(\tilde{\theta}) = 0$.

Intuitively, the agent cannot have a nonstationary long-run belief $\tilde{\theta}$, as her limiting action $a^*(\tilde{\theta})$ would generate subjective signals that push her systematically away from $\tilde{\theta}$. Specifically, when $g(\tilde{\theta}_t) > 0$, then the agent eventually generates signals that are on average above $\tilde{\theta}_t$, so that her mean beliefs drift upward, and if $g(\tilde{\theta}_t) < 0$, then the agent's mean beliefs drift downwards.

Furthermore, intuition suggests that not every stationary belief is a possible limiting belief: whenever g is negative on the left of $\tilde{\theta}$ and positive on the right of $\tilde{\theta}$, the agent's noisy perceived signals would still push her away from $\tilde{\theta}$ in one of the two directions. This motivates the following definition.

Definition 2 (Stable point). A stationary point $\tilde{\theta}$ is *stable* if g crosses zero from above at $\tilde{\theta}$.[7] We denote by H the set of stable points.

The main result of the paper shows that the set of stable beliefs completely describes the set of possible long-run beliefs.

Theorem 1. *Suppose that Assumption 1 holds. Then almost surely the agent's mean beliefs* $(\tilde{\theta}_t)$ *converge and the limit point is stable:* $\lim_{t\to\infty} \tilde{\theta}_t \in H$.

[6]Formally, for any x such that $g(x) = 0$, we have that $g(x-\delta)g(x+\delta) < 0$ for any sufficiently small $\delta > 0$.

[7]Formally, $g(\tilde{\theta} - \delta) > 0$ and $g(\tilde{\theta} + \delta) < 0$ for any sufficiently small $\delta > 0$.

The mathematical intuition behind Theorem 1 is the following. Because the agent's posterior can be fully described by its mean and variance and the latter depends only on time t, we study the dynamics of the agent's mean beliefs $\tilde{\theta}_t$. Although—as we explained above—the agent's subjective signals are not independently and identically distributed over time, we use a result from stochastic approximation that in our context requires the perceived signals to be approximately independent over time conditional on the current subjective belief. To use this result, for $s \geq t$, we approximate the agent's time-dependent action $a^*(s, \tilde{\theta}_{s-1})$ by the time-independent confident action $a^*(\tilde{\theta}_{s-1})$. By Assumption 1, the mistake we make in approximating $a^*(s, \tilde{\theta}_{s-1})$ in this way is of order $1/s^k$, and since the perceived signal $\tilde{b}_s$ is Lipschitz continuous in a_s, the mistake we make in approximating $\tilde{b}_s$ is also of order $1/s^k$. When updating, since the agent's newest signal gets a weight of order $1/s$, this means that the error in approximating the change in her beliefs is of order $1/s^{1+k}$, and, hence, the total error in approximating beliefs from period t onward is of order $\sum_{s=t}^{\infty} 1/s^{1+k}$, i.e., finite. Furthermore, as $t \to \infty$, the approximation error from period t onward goes to zero.

Given the above considerations, on the tail we can think of the dynamics as driven by those that would prevail if the agent chose $a^*(\tilde{\theta}_t)$ in period $t+1$. The expected change in the mean belief in period $t+1$ is, therefore, a function of $g(\tilde{\theta}_t)$. The expected change, however, is a time-dependent function of $g(\tilde{\theta}_t)$: as the agent accumulates observations, she puts less and less weight on a single observation. Stochastic approximation theory, therefore, defines a new time scale τ_t and a new process $z(\tau)$ that "speeds up" $\tilde{\theta}_t$ to keep its expected steps constant over time, also making $z(\tau)$ a continuous-time process by interpolating between points. It then follows that on the tail, the realization of z can be approximated by the ordinary differential equation

$$z'(\tau) = g\big(z(\tau)\big), \tag{4}$$

which is a deterministic equation. Intuitively, since $z(\tau)$ is a sped-up version of $\tilde{\theta}_t$, as τ increases, any given length of time in the process $z(\tau)$ corresponds to more and more of the agent's observations. Applying a version of the law of large numbers to these many observations, the change in the process is close to deterministic. In this sense, on the tail, the noisy model reduces to a noiseless model. Now because the solution to (4) converges to a stable point, the agent's beliefs converge to a stable point.

In line with the above intuition, the first step in the proof of Theorem 1 is to eliminate actions from the formalism, simply approximating the agent's signal $\tilde{b}_t$ with a time-independent signal. Then we show that if the agent's beliefs depend only on time and a one-dimensional statistic in a way that admits a time-independent approximation with an error on the order of $1/t^{1+k}$, then the statistic converges. Because the second argument allows us to cover additional economic applications, we formally prove it in a separate Theorem 2 below.

We conclude this section by relating our theorem to Esponda and Pouzo's (2016a) concept of Berk–Nash equilibrium. Denote the *log-likelihood ratio* $L^a(b, \tilde{\theta})$ between the objective probability measure associated with realizations of b and the agent's subjective

probability measure over realizations of $\tilde{b}(b, a)$ when she takes the action a as

$$L^a(b, \tilde{\theta}) = \log\left(\frac{\phi(b - \Theta)}{\phi\big(\tilde{b}(b, a) - \tilde{\theta}\big)\tilde{b}_b(b, a)}\right),$$

where $\phi(x) = \phi(x; 0, \eta^2)$ is the density of a normally distributed random variable with variance η^2.[8] The *Kullback–Leibler* divergence between the objective distribution of signals and the subjective distribution when the agent believes the state equals $\tilde{\theta}$ and randomly takes an action a according to the mixed strategy $\nu \in \Delta(A)$ equals

$$K^\nu(\tilde{\theta}) = \int \mathbb{E}\big[L^a(b, \tilde{\theta})\big]\, d\nu(a).$$

Definition 3 (Berk–Nash equilibrium). A distribution over actions $\nu \in \Delta(A)$ constitutes a *Berk–Nash equilibrium* if there exists a probability distribution over states $\zeta \in \Delta(\mathbb{R})$ such that the following statements hold:

(i) The distribution over actions is optimal given the subjective belief ζ:

$$\int\int M(a, \tilde{\theta})\, d\zeta(\tilde{\theta})\, d\nu(a) = \max_a \int M(a, \tilde{\theta})\, d\zeta(\tilde{\theta}).$$

(ii) The support of the belief ζ is contained in the set of Kullback–Leibler minimizers:

$$\operatorname{supp} \zeta \subseteq \operatorname*{argmin}_{\tilde{\theta}} K^\nu(\tilde{\theta}).$$

Intuitively, Berk–Nash equilibrium requires that (i) the agent randomizes over actions in an optimal way given her subjective beliefs and (ii) her subjective belief is the belief that "best" explains the data she is observing given her actions. Furthermore, we say that a Berk–Nash equilibrium is *strict* if any other distribution over actions is strictly suboptimal in (i).

Proposition 1 (Characterization of Berk–Nash equilibria). *(i) A distribution over actions is a Berk–Nash equilibrium if and only if it assigns probability* 1 *to a single action a such that $a = a^*(\tilde{\theta})$ for some stationary belief $\tilde{\theta}$.*

(ii) All Berk–Nash equilibria are strict.

(iii) Play never converges to a Berk–Nash equilibrium $a = a^*(\tilde{\theta})$ associated with a stationary belief $\tilde{\theta}$ that is not stable.

Combined with Theorem 1's result that the agent's beliefs converge with probability 1 to a stable stationary belief, the proposition says that the agent's beliefs converge with

[8]Formally, L denotes the logarithm of the Radon–Nykodym derivative between the objective and the subjective measure. The density of the objective measure follows as b_t is normally distributed with mean Θ and variance η^2. The density of the subjective signal distribution follows as the agent misinterprets the signal b as $\tilde{b}(b, a)$ when taking the action a, and thinks that $\tilde{b}$ is normally distributed with mean $\tilde{\theta}$ and variance η^2.

probability 1 to a Berk–Nash equilibrium, but only the subset of Berk–Nash equilibria corresponding to stable beliefs are possible limit points. Specifically, part (i) says that in our model, the set of Berk–Nash equilibria coincides with the set of stationary beliefs, i.e., beliefs at which the agent's mean surprise about the fundamental is zero. Furthermore, part (ii) says that all Berk–Nash equilibria are strict, independently of whether the associated stationary beliefs are stable or not. Using Theorem 1, we can conclude that the agent's beliefs converge with probability 1 to a strict Berk–Nash equilibrium. Part (iii), however, implies that when there is an unstable stationary belief, then that belief corresponds to a strict Berk–Nash equilibrium to which the agent's beliefs do not converge with positive probability. Hence, in our model these strict Berk–Nash equilibria do not have a learning foundation.

5. Generalizations

While our model above assumes that the agent has a normal prior and observes normally distributed signals, weaker assumptions are sufficient to establish convergence using our methodology. In this section, we outline a general, abstract learning model and identify a few applications of it.

We consider a sequence of i.i.d. signals that are distributed according to some distribution $F : (\underline{b}, \overline{b}) \to [0, 1]$ with (potentially unbounded) support $(\underline{b}, \overline{b})$, strictly positive density $f > 0$, and finite expectation and variance. We suppose that if the agent perceives to have observed the sequence of signals $\tilde{b}_1, \dots, \tilde{b}_t$, then her posterior subjective belief can be summarized by a single-dimensional "sufficient statistic"

$$\tilde{\theta}_t = \frac{\alpha\tilde{\theta}_0 + \sum_{s=1}^{t} \tilde{b}_s}{\alpha + t}, \tag{5}$$

where α is a free parameter. Furthermore, the subjective signal $\tilde{b}_t$ is only a function of calendar time t, the agent's last-period belief $\tilde{\theta}_{t-1}$, and the true signal b_t:

$$\tilde{b}_t \equiv \tilde{b}(t, b_t, \tilde{\theta}_{t-1}).$$

We assume that the difference between the objective and the subjective signals is bounded ($|\tilde{b}(t, b, \tilde{\theta}) - b| \leq \Delta < \infty$) and that the subjective signal $\tilde{b}(t, b, \tilde{\theta})$ is differentiable, strictly increasing in the objective signal b and continuous in the belief $\tilde{\theta}$.

Let

$$g(x) = \lim_{t\to\infty} \mathbb{E}\big[\tilde{b}(t, b_t, x) - x\big] \tag{6}$$

be the long-run limit of the objective expectation of the difference between the subjective signal and the agent's subjective belief x. We assume that g is well defined (i.e., the above limit exists) and there exist constants $d_1, k > 0$ such that for all $x \in \mathbb{R}$ and all $t \geq 1$,

$$\big|g(x) - \mathbb{E}\big[\tilde{b}(t, b_t, x) - x\big]\big| \leq d_1 \frac{1}{t^k}. \tag{7}$$

We impose that g does not touch the horizontal axis without crossing.[9] We define stationary and stable points analogous to Definitions 1 and 2. We then have the following theorem.

THEOREM 2. *Almost surely the sufficient statistic* $(\tilde{\theta}_t)$ *converges and the limit point is stable:* $\lim_{t\to\infty} \tilde{\theta}_t \in H$.

Despite differences in appearance, our main model with normally distributed signals and endogenous actions (Section 3) is a special case of the above setup. Namely, the optimal action is only a function of the posterior belief in the previous period, which is normal with mean $\tilde{\theta}_{t-1}$ and variance v_{t-1}. Hence, the action in period t is only a function of $\tilde{\theta}_{t-1}$ and t. As the action a_t and the true signal b_t determine the perceived signal, the perceived signal $\tilde{b}_t$ is only a function of $(t, b_t, \tilde{\theta}_{t-1})$. We use this insight to prove Theorem 1 from Theorem 2 in Appendix A.2.

We next discuss how Theorem 2 can help establish convergence in variants of our model in Section 3. Throughout, we assume that unless stated otherwise, any assumption we have made in Section 3 holds.

5.1 *Exponentially distributed signals*

Consider a modification of our model in Section 3 in which the b_t are i.i.d. exponentially distributed random variables with mean Θ, and the agent's prior about Θ is inverse-gamma distributed with shape $\alpha_0 > 2$ and scale $\lambda_0 > 0$. We denote by $\psi(x; \lambda, \alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)}(1/x)^{\alpha+1}\exp(-\lambda/x)$ the density of the inverse-gamma distribution with shape α and scale λ.[10] The agent's posterior belief about Θ in period t is then inverse-gamma distributed with scale $\lambda_t = \lambda_0 + \sum_{s=1}^{t} \tilde{b}_s$ and shape $\alpha_t = \alpha_0 + t$.[11] The mean of the agent's posterior belief equals

$$\tilde{\theta}_t = \frac{\tilde{\theta}_0(\alpha_0 - 1) + \sum_{s=1}^{t} \tilde{b}_s}{\alpha_0 - 1 + t}, \tag{8}$$

where $\tilde{\theta}_0 = \frac{\lambda_0}{\alpha_0 - 1}$ equals the mean of the prior.[12] The variance of the agent's posterior belief in period t is given by $\frac{\tilde{\theta}_t^2}{\alpha_0 - 2 + t}$. Hence, the scale and shape parameter can be uniquely identified from calendar time t and the posterior mean $\tilde{\theta}_t$, i.e., $(\alpha_t, \lambda_t) = (\alpha_0 + t, (\alpha_0 - 1 + t)\tilde{\theta}_t)$.

Given that $(t, \tilde{\theta}_t)$ is, therefore, a sufficient statistic for the agent's beliefs and that the updating rule (8) is linear, our methods are easily adapted to establish convergence of beliefs. To do so, we redefine the expected output $M(a, \tilde{\theta}) = \int \tilde{Q}(a, \tilde{b})\frac{1}{\tilde{\theta}}e^{-\tilde{b}/\tilde{\theta}}\, d\tilde{b}$ when

[9] Formally, for any x such that $g(x) = 0$, $g(x - \delta)g(x + \delta) < 0$ for any sufficiently small $\delta > 0$.
[10] The equality $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\, dx$ is the gamma function.
[11] See https://en.wikipedia.org/wiki/Conjugate_prior.
[12] See https://en.wikipedia.org/wiki/Inverse-gamma_distribution.

the agent is certain of the state. It follows from the same argument as in Lemma 1 that the optimal action is continuous in the state and satisfies

$$a_t = a^*(t, \tilde{\theta}_{t-1}) = \operatorname*{argmax}_{a \in A} \int M(a, x)\psi\big(x; \tilde{\theta}_{t-1}(\alpha_{t-1} - 1), \alpha_{t-1}\big)\, dx.$$

By exactly the same proof as for Theorem 1, we obtain the following corollary.

Corollary 1. *Suppose that Assumption 1 holds. Then almost surely the agent's mean beliefs* $(\tilde{\theta}_t)$ *converge and the limit point is stable:* $\lim_{t\to\infty} \tilde{\theta}_t \in H$.

5.2 *Other objective signal distributions*

The general model immediately implies another point: that it is sufficient for the agent's subjective model to be normal or exponential; the true model generating the signals b_t can be something else. This does not modify the agent's updating rule or behavior given her beliefs; it only modifies the definition of g in (3). Since we do not use the true distribution of b in the proof of Lemma 1 or Theorem 1, the next corollary follows immediately.

Corollary 2. *Suppose that the true distribution of* b_t *is i.i.d. with mean* Θ *and finite variance, and that Assumption 1 holds. Then almost surely the agent's mean beliefs* $(\tilde{\theta}_t)$ *converge and the limit point is stable:* $\lim_{t\to\infty} \tilde{\theta}_t \in H$.

5.3 *Decoupling the agent's objective from her observations*

Consider also decoupling the agent's objective from her observations. Suppose that in choosing a_t, the agent aims to maximize the expectation of some function $\tilde{Q}(a_t, b_t)$, she observes $Q(a_t, b_t)$, and she believes that she observes $\hat{Q}(a_t, b_t)$. For notational simplicity, our model imposes $\hat{Q} = \tilde{Q}$, but this is not necessary. In particular—supposing that $\hat{Q}$ also satisfies all assumptions previously stated for $\tilde{Q}$—we can generalize the definition of $\tilde{b}_t$ in (1) by replacing $\tilde{Q}$ with $\hat{Q}$:

$$\hat{Q}(a_t, \tilde{b}_t) = Q(a_t, b_t).$$

This generalization, thus, changes only the definition of $\tilde{b}$ and a^*. But since Theorem 1 relies solely on the properties of $\tilde{b}$ and a^* imposed through Assumption 1, we have the following result.

Corollary 3. *Suppose that Assumption 1 holds. Then almost surely the agent's mean beliefs* $(\tilde{\theta}_t)$ *converge and the limit point is stable:* $\lim_{t\to\infty} \tilde{\theta}_t \in H$.

6. Economic applications

Developing detailed insights in specific economic settings is not the goal of this paper. Nevertheless, to demonstrate the usefulness of our framework in a range of situations, we identify a few applications and use our convergence result to make some simple points in them. Throughout this section, we assume that the agent's prior and the signals are normally distributed.

Mis-estimating equilibrium feedback Our framework can be used to analyze representative–agent models in which the agent interprets macroeconomic observations incorrectly. As a simple example, suppose that the fundamental is a measure of macroeconomic conditions, a_t is an action (such as one's consumption or housing choice) that individuals look to align with macroeconomic conditions, and the true fundamental Θ is normalized to zero. Each period t fundamental influences the economy through a short-run shock b_t, which equals the fundamental plus normal noise. The representative agent chooses a_t to maximize the expectation of $\tilde{Q}(a_t, b_t) = b_t - L(a_t - b_t)$, and observes a macroeconomic outcome (such as growth) given by $Q(a_t, b_t) = b_t + m(a_t) - L(a_t - (b_t + m(a_t)))$, where L is a symmetric, twice differentiable loss function satisfying $L(0) = 0$, $L(x) \geq 0$, $L''(x) > 0$, and $|L'(x)| < k < 1$ for all x, and m is a twice differentiable function. While in reality the aggregate outcome depends on the level of individuals' choices through $m(a)$, the representative agent does not understand this and believes it depends only on how well his action a_t matches the short-run shock b_t.[13] We suppose that $m(a)$ and $m'(a)$ are bounded from above and below, and $m(0) = 0$, $m'(0) > 1$, $m'(a) > 0$ and $m''(a) < 0$ for $a > 0$, and $m(-a) = -m(a)$.

We argue that the agent's beliefs almost always converge and the limiting belief is ambiguous. Given that the agent's posterior is symmetric and L is symmetric, she chooses the action corresponding to her mean belief: $a_t = \tilde{\theta}_{t-1}$. Thus, $a^*(t, \tilde{\theta})$ is independent of t and, hence, Assumption 1 part (iii) holds trivially. Furthermore, since $a_t = \tilde{\theta}_{t-1}$, we have $\tilde{b}(b_t, a_t) = b_t + m(a_t)$, so Assumption 1 part (i) holds because $m(a)$ is bounded. Part (ii) of Lemma 2 implies that Assumption 1 part (ii) also holds in this application.[14] In addition, since $\tilde{b}(b_t, a_t) = b_t + m(a_t)$ and $\mathbb{E}[b_t] = \Theta = 0$, we have that $g(\tilde{\theta}) = m(\tilde{\theta}) - \tilde{\theta}$. This implies that there is a unique $\tilde{\theta}_\infty > 0$ such that $g(\tilde{\theta}_\infty) = 0$,[15] and, therefore, the set of stationary points is given by $\{-\tilde{\theta}_\infty, 0, \tilde{\theta}_\infty\}$. But 0 is not a stable stationary point of g, so the set of stable points is given by $H = \{-\tilde{\theta}_\infty, \tilde{\theta}_\infty\}$. By Theorem 1, the agent's beliefs converge with probability 1 to one of $\tilde{\theta}_\infty$ and $-\tilde{\theta}_\infty$; in fact, by the symmetry of the problem, it must be the case that the agent's beliefs converge to each with probability 1/2. Accordingly, the economy converges to either an overly high or an overly low level of activity relative to what is justified by the fundamental. Which one the economy and the agent's beliefs converge to depends on early observations. If by chance output happens to be high in initial periods, the representative agent chooses high actions, and as she misinterprets the resulting observations as indicating strong fundamentals, she develops optimistic, "irrationally exuberant" beliefs about the economy. The converse happens if by chance her initial observations of output are low.

[13]In an alternative interpretation motivated by Dal Bó et al. (2018), a_t is a policy chosen by voters and Q is an economic outcome affected by the policy. While in reality the policy generates a general-equilibrium feedback effect $m(a_t)$, voters do not understand this.

[14]Since $\tilde{Q}_{\tilde{b}}(a, \tilde{b}) \geq 1 - k$, it is uniformly bounded from below, and since $|\tilde{Q}_a(a, \tilde{b})| \leq k$ and $|Q_a(a, b)| \leq |m'(0)| + k + k|m'(0)|$, the absolute values of these derivatives are uniformly bounded, and hence part (ii) of Lemma 2 applies.

[15]Since $m(a)$ is bounded and its derivative is positive and convex on the positive domain, it must be that $\lim_{a\to\infty} m'(a) = 0$, implying the existence of a $\tilde{\theta}_\infty > 0$ such that $g(\tilde{\theta}_\infty) = 0$.

A priori, it was not obvious that the agent's beliefs converge. It may seem possible, for instance, that the agent periodically receives signals that disconfirm her current direction of belief and that are sufficiently strong for her to start developing an opposite bias. Our framework says that this can happen only finitely many times.

Mislearning new technology Our basic model in Section 3 can capture learning about a technology over time. As a specific example, consider fertilizer use in developing countries. It has long been hypothesized that many farmers use fertilizer to a lower extent than optimal, and they seem to forego substantial returns by doing so (Duflo et al. 2008). While one reason for this phenomenon is likely to be the lack of experimentation with the new technology, it is possible that misspecified learning also plays a role. To model this, suppose that Θ is a measure of the average effectiveness of fertilizer, $a_t > 0$ is the amount of fertilizer the farmer uses, c is the unit cost of the fertilizer, and profit is determined according to $Q(a_t, b_t) = k(a_t)\exp b_t - ca_t$. While the farmer understands the general form of output, she misunderstands k: $\tilde{Q}(a_t, b_t) = \tilde{k}(a_t)\exp b_t - ca_t$. One possible reason for this misspecification is that $\tilde{Q}$ is the correct functional form for output with the old technology, which the farmer is very familiar with. She understands that she is currently facing a new technology and that she needs to assess its potential. But in doing so, she neglects that she should reestimate $\tilde{k}$ and, instead, reestimates only the fundamental value.

The functions k and $\tilde{k}$ are illustrated in Figure 1. The actual technology is quite concave, so there is a small range of usage amounts that are optimal for a large range of Θ and c. Because she is familiar with more traditional technologies, however, the agent expects a less stark response pattern to usage. To complement the figure, we impose that $\Theta + \sigma^2/2 = 0$ and $c = 1$, $\tilde{k}, k > 0$, $\lim_{a\to 0}\tilde{k}'(a) = \lim_{a\to 0} k'(a) = \infty$, $\lim_{a\to\infty}\tilde{k}'(a) = 0$, and $k'' < \tilde{k}'' < 0$. These assumptions imply that at the true fundamental Θ, the optimal and perceived-optimal actions are identical and satisfy $k'(a) = \tilde{k}'(a) = 1$.[16] Therefore, if the agent knew or could learn the true fundamental, then despite her misspecification, she would choose the optimal action. To ensure that the example satisfies our assumptions, we furthermore suppose that $(k'(a)/k(a)) - (\tilde{k}'(a)/\tilde{k}(a))$ is bounded from above as well as that $\tilde{k}'(a)/|\tilde{k}''(a)|$ is bounded from above.

From (1), we have $\tilde{k}(a^*(\tilde{\theta}))\exp(\tilde{b}(b, a^*(\tilde{\theta}))) = k(a^*(\tilde{\theta}))\exp(b)$, so that $\tilde{b}(b, a^*(\tilde{\theta})) - b = \log(k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta})))$. Using this fact, one can verify that our example satisfies Assumption 1, which we do in Appendix A.4. Substituting $\tilde{b}(b, a^*(\tilde{\theta}))$ into (3), we obtain $g(\tilde{\theta}) = \mathbb{E}\,[b + \log(k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta})))] - \tilde{\theta} = \Theta - \tilde{\theta} + \log(k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta})))$. Furthermore, the subjectively optimal action $a^*(\tilde{\theta})$ satisfies the first-order condition $\tilde{k}'(a^*(\tilde{\theta}))\exp(\tilde{\theta} + \sigma^2/2) = 1$. This can be rewritten as $\tilde{k}'(a^*(\tilde{\theta}))\exp(\Theta + \sigma^2/2 + \tilde{\theta} - \Theta) = 1$, so using $\Theta + \sigma^2/2 = 0$ yields $\Theta - \tilde{\theta} = \log(\tilde{k}'(a^*(\tilde{\theta})))$. Hence, we get $g(\tilde{\theta}) = \log[\tilde{k}'(a^*(\tilde{\theta}))k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta}))]$. The first-order condition also implies that the perceived optimal action $a^*(\tilde{\theta})$ is monotonically increasing in the agent's mean belief $\tilde{\theta}$, goes to zero as $\tilde{\theta} \to -\infty$, and goes to plus infinity as $\tilde{\theta} \to \infty$.

[16]To see why, note that the mean of a log-normal distribution with parameters μ and σ^2 is $\exp(\mu + \sigma^2/2)$, so that the expectation of $\exp(b_t)$ is $\exp(\Theta + \sigma^2/2) = 1$.

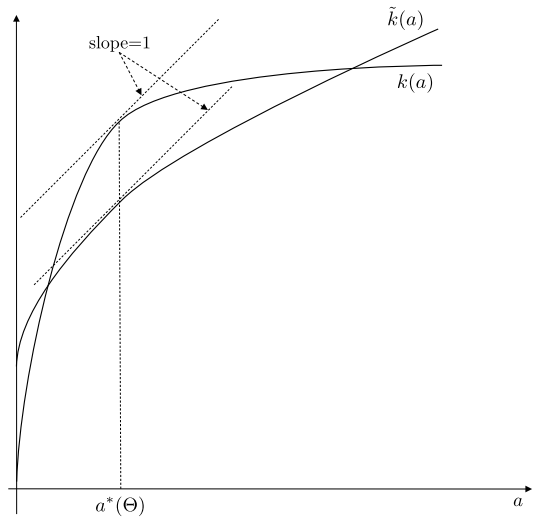

FIGURE 1. Actual and estimated returns to fertilizer use.

Now we can characterize the agent's possible limiting beliefs. Observe that since $\tilde{k}'(a^*(\Theta)) = 1$ and $k(a^*(\Theta)) > \tilde{k}(a^*(\Theta))$, we have $g(\Theta) > 0$. Furthermore, since $\tilde{k}'(a^*(\tilde{\theta})) < 1$ for any $\tilde{\theta} > \Theta$, at any belief $\tilde{\theta} > \Theta$ such that $k(a^*(\tilde{\theta})) \leq \tilde{k}(a^*(\tilde{\theta}))$, we have $g(\tilde{\theta}) < 0$. Hence, g has a stable zero above Θ. There could also be a stable zero below Θ.[17] These two types of long-run beliefs arise through the following mechanisms. If the agent begins to conclude that the technology is not very productive, then she chooses low usage. This choice leads to low productivity, but because she underestimates how sensitive productivity is to usage, she attributes the low productivity in a large part to the technology. As a result, she develops overly pessimistic beliefs about the technology ($\tilde{\theta}_\infty < \Theta$), and underuses it. If, instead, the agent begins to conclude that the fertilizer is effective, she uses it heavily. This leads to high productivity, which she attributes too much to the technology. As a result, she develops overly favorable views of the technology ($\tilde{\theta}_\infty > \Theta$) and overuses it. Interestingly, even in this case she underestimates the technology *net of usage costs*, as she believes that the high productivity requires high usage. Hence, if she has an alternative technology she could use, then with either type of limiting belief, she is too likely to use the alternative.

Estimating parameters of wrong demand model In a simple application of our model, the decisionmaker is a firm. The firm chooses price $a_t > 0$ in each period and obtains profits $Q(a_t, b_t)$, where b_t is the stochastic demand state. The firm, however, believes that profit is determined according to $\tilde{Q}(a_t, b_t)$.

[17]Assuming that g has finitely many zeros, a necessary and sufficient condition for this is that there is an $a' < a^*(\Theta)$ such that $\tilde{k}'(a')k(a')/\tilde{k}(a') < 1$. This can happen, for instance, if $\tilde{k}'(a) \approx 1$ (that is, k is almost linear) in a sufficiently large range around $a^*(\Theta)$ to include $a < a^*(\Theta)$ for which $k(a)$ is nontrivially less than $\tilde{k}(a)$. Then since $\lim_{a\to 0} \tilde{k}'(a)k(a)/\tilde{k}(a) > 1$, there must be an $a'' < a'$ such that $\tilde{k}'(a'')k(a'')/\tilde{k}(a'') = 1$, and the function $\tilde{k}'(a)k(a)/\tilde{k}(a)$ crosses 1 at a'' from above. The fundamental $\tilde{\theta}$ that satisfies $\Theta - \tilde{\theta} = \log(\tilde{k}'(a''))$ is a stable zero of g.

As a specific case, we consider a simplified version of the main example in Nyarko (1991) and the nonconvergence result in Fudenberg et al. (2017). Suppose that true demand is $b_t - m(a_t)$ with $\Theta = 6$, whereas the agent believes that demand is determined according to $b_t - a_t$. Positing that marginal cost is zero, this implies that $Q(a_t, b_t) = a_t(b_t - m(a_t))$ and $\tilde{Q}(a_t, b_t) = a_t(b_t - a_t)$. Hence, if the agent believes that the fundamental is $\tilde{\theta}$, then she chooses a price of $\tilde{\theta}/2$. We suppose that $m(a_t) = 3a_t$ for $a_t \in [0, 6]$, that $m(a_t)$ is twice differentiable, that $m'(a_t) > 1$ and bounded from above, and that $m(a_t) - a_t$ is bounded from above and below.

Here, $\tilde{b}(b_t, a_t)$ solves the equation $b_t - m(a_t) = \tilde{b}_t - a_t$, so $\tilde{b}(b_t, a_t) = b_t - (m(a_t) - a_t)$ and it is straightforward to verify that Assumption 1 holds.[18] Using $a^*(\tilde{\theta}) = \tilde{\theta}/2$ yields $g(\tilde{\theta}) = \mathbb{E}[b_t - (m(a^*(\tilde{\theta})) - a^*(\tilde{\theta}))] - \tilde{\theta} = \Theta - m(\frac{\tilde{\theta}}{2}) - \frac{\tilde{\theta}}{2}$, which (since $\Theta = 6$) implies that $H = \{3\}$. Hence, the agent comes to believe that the fundamental is 3, which is exactly the belief at which the distribution of demand she expects to obtain with the perceived optimal price of 3/2 equals the true distribution of demand. The agent understands the level of demand correctly, but underestimates the responsiveness of demand to price, so she sets an overly high price.

This convergence result relies on the continuity of the state and action spaces. Suppose instead that (as in Nyarko) the agent entertains only two possible levels of the fundamental: $\tilde{\theta} = 2$ and $\tilde{\theta} = 4$. If she believes the former, then she chooses a price of 1, generating an average demand of 3, which according to her model is consistent with $\tilde{\theta} = 4$. Conversely, if she believes the latter, then she sets a price of 2, generating an average demand of 0, which according to her model is consistent with $\tilde{\theta} = 2$. Hence, her beliefs cannot converge to either point belief. Furthermore, a single signal can move any intermediate belief by a nontrivial amount, so her beliefs also cannot converge to a nondegenerate belief.[19] As a result, her beliefs and prices fluctuate forever. By the same argument, if the agent's prior is normal, but she can choose only two price levels, 1 and 2, then choosing one price eventually convinces the agent that the other price is optimal, so her prices must fluctuate forever.

Overconfidence A large literature in psychology documents and recent economic research explore the implications of the idea that individuals have unrealistically positive views of their traits and prospects. This hypothesis can easily be captured in our framework by assuming that $\tilde{Q}(a, b) > Q(a, b)$, that is, for any action and any state, the agent

[18]Part (i) holds since $m(a_t) - a_t$ is bounded, part (ii) since $\tilde{b}_a(b_t, a_t) = 1 - m'(a_t)$ is bounded, and part (iii) since $a^*(t, \tilde{\theta}) = a^*(\tilde{\theta})$.

[19]To see this, suppose that the agent assigns probability p to the fundamental being 2 and probability $(1 - p)$ to the fundamental being 4. For notational simplicity, denote the associated optimal action by a, and denote the probability density function of a normal distribution with mean zero and variance σ^2 by $\phi(\cdot)$. Upon deducing $\tilde{b}(b, a)$, Bayes' rule yields the new belief

$$p \cdot \frac{1}{p + (1-p)\dfrac{\phi\big(\tilde{b}(b,a) - 4\big)}{\phi\big(\tilde{b}(b,a) - 2\big)}}.$$

Clearly, there is a $\underline{b}$ such that if $b > \underline{b}$, then $\phi(\tilde{b}(b, a) - 4)/\phi(\tilde{b}(b, a) - 2) > 2$. For such b, the new beliefs are below $p/(2 - p)$. Hence, for any p, the agent's beliefs drop from p to below $p/(2 - p)$ with a nonvanishing probability, contradicting that beliefs converge.

expects higher output than is realistic. For instance, the agent's productivity may be an average of different skills, and in taking the average, the agent includes her strengths, but neglects her weaknesses. The external state can be any other variable that influences output and to which the agent is trying to adjust her action. This model is very similar to that in Heidhues et al. (2018), but has a different functional form for what determines output. Our earlier paper explores the consequences of such a misspecification in settings with a unique limiting belief. Our current model allows for multiple possible limiting beliefs as well.

7. Conclusion

At the cost of additional notation, our methods can likely be extended to situations in which the agent's action and/or information are multidimensional. This does, however, require that the agent observes not only output, but also other information, so that she believes she can recover the relevant multidimensional signal. In an industrial-organization setting, for instance, the firm may model demand as linear, so that she seeks to learn the intercept and slope. We conjecture that if she observes signals of the level and variance of demand (e.g., because she has multiple stores), then a version of our methods can be used to establish convergence of her beliefs.

While we have focused on situations in which the agent's signals are endogenous and nonindependent due to the endogenous actions she takes, our general methods apply to models of mistaken learning in which endogenously dependent signals arise for other reasons. These include situations in which the agent has a correct prior, but does not use Bayes' rule to update beliefs. She may, for example, be subject to an updating bias, such as conservatism or confirmatory bias, in which her current belief colors how she sees information. Such settings fit in the general framework of Theorem 2, so under our conditions, the agent's beliefs converge. In fact, we show in our working paper that when the agent has conservatism or recency bias, her beliefs converge to the truth with probability 1. Intuitively, for the agent to learn the true state, it is sufficient for her beliefs to move in the direction of her newest signal: the exact speed at which this happens is irrelevant.

Appendix: Proofs

Throughout, we denote by C the set of stationary points

$$C = \{x \in \mathbb{R} \colon g(x) = 0\} \tag{9}$$

and denote by H the set of stable points

$$H = \{x \in C \colon g(x - \epsilon) > 0 \text{ and } g(x + \epsilon) < 0 \ \forall \epsilon \text{ small enough}\}.$$

A.1 *Proof of Theorem 2*

We begin by showing two auxiliary lemmas, which are useful in the proof of Theorem 2. Throughout, we fix a state Θ and denote by $\mu = \mathbb{E}[b_t]$ the expected signal conditional on the state.

Lemma 3. *The function* $g(x) = \lim_{t\to\infty} \mathbb{E}[\tilde{b}(t, b_t, x) - x]$ *is continuous and satisfies*

$$\mu - x - \Delta \le g(x) \le \mu - x + \Delta.$$

Furthermore, $C = \{x\colon g(x) = 0\}$ *is finite and* H *is nonempty.*

Proof. Define $g(t, x) = \mathbb{E}[\tilde{b}(t, b_{t+1}, x) - x]$. By the definition of $g(t, x)$, we have that

$$g(t, x) + x - \mathbb{E}[b_t] = \mathbb{E}\big[\tilde{b}(t, b_{t+1}, x) - b_t\big].$$

Taking the absolute value, applying the triangle inequality, and using that $|\tilde{b} - b| \le \Delta$ yields

$$\big|g(t, x) - (\mu - x)\big| \le \mathbb{E}\big[\big|\tilde{b}(t, b_{t+1}, x) - b_t\big|\big] \le \Delta.$$

This implies that $g(t, \mu + 2\Delta) \le \mu - (\mu + 2\Delta) + \Delta = -\Delta < 0$ and that $g(\mu - 2\Delta) \ge \mu - (\mu - 2\Delta) - \Delta = +\Delta > 0$. Taking the limit $t \to \infty$ yields that $g(\mu + 2\Delta) \le -\Delta < 0$ and $g(\mu - 2\Delta) \ge +\Delta > 0$. As $x \mapsto \tilde{b}(t, b, x)$ is continuous in x, $x \mapsto g(t, x)$ is continuous, and as $g(t, x)$ converges to $g(x)$ uniformly, the limiting function g is continuous. Thus, g crosses zero at least once from above in the interval $(\mu - 2\Delta, \mu + 2\Delta)$.

To see that C is finite, observe that $g(x) = 0$ implies that $x \in [\mu - 2\Delta, \mu + 2\Delta]$. Suppose that $|C|$ is infinite. Then there exists a converging sequence of points $(x_k)_k$ with $x_k \in C$. As g is continuous, we have that the point $y = \lim_{k\to\infty} x_k$ satisfies $g(y) = \lim_{k\to\infty} g(x_k) = 0$ and, thus, $g(y) \in C$. We, thus, know that for all δ small enough, we have that $g(y - \delta)g(y + \delta) < 0$. This implies that $g(x_k) \ne 0$ for all k large enough, which contradicts $x_k \in C$. Thus, C is finite. As $g(\mu + 2\Delta) < 0$ and $g(\mu - 2\Delta) > 0$, and g is continuous, g crosses zero at least once from above and H is nonempty. □

Define the stochastic process $(\theta_t)_t$ by

$$\theta_t = \frac{\alpha\tilde{\theta}_0 + \sum_{s=1}^{t} b_s}{\alpha + t}. \tag{10}$$

Lemma 4. *We have that* $\lim_{t\to\infty} \theta_t = \mu$ *a.s. and* $|\tilde{\theta}_t - \theta_t| \le \Delta$.

Proof. The process $(\theta_t)_t$ converges almost surely to μ by the law of large numbers. We note that

$$\theta_{t+1} = \theta_t + \gamma_t(b_{t+1} - \theta_t)$$
$$\tilde{\theta}_{t+1} = \tilde{\theta}_t + \gamma_t(\tilde{b}_{t+1} - \tilde{\theta}_t),$$

where $\gamma_t = \frac{1}{\alpha+t+1}$. This implies that $\theta_{t+1} - \tilde{\theta}_{t+1} = (1 - \gamma_t)(\theta_t - \tilde{\theta}_t) + \gamma_t(b_{t+1} - \tilde{b}_{t+1})$. Since $\tilde{\theta}_0 = \theta_0$ and $|b_t - \tilde{b}_t| \le \Delta$, it follows by induction that $|\tilde{\theta}_t - \theta_t| \le \Delta$. □

Lemma 5. *The process* $(\tilde{\theta}_t)$ *is almost surely bounded:*

$$\mathbb{P}\Big[-\infty < \liminf_t \tilde{\theta}_t\Big] = \mathbb{P}\Big[\limsup_t \tilde{\theta}_t < \infty\Big] = 1.$$

Proof. As $|\tilde{\theta}_t - \theta_t| \le \Delta$ and θ_t converges to μ a.s. by Lemma 4, $\limsup_{t\to\infty} \tilde{\theta}_t \le \Theta + \Delta$ and $\liminf_{t\to\infty} \tilde{\theta}_t \ge \Theta - \Delta$ almost surely, which implies the result. □

Proof of Theorem 2. The set of *stationary points* of the ODE

$$x'(t) = g(x) \tag{11}$$

is given by $C = \{x\colon g(x) = 0\}$ as defined in (9). As argued in Lemma 3, C is finite and we denote the stationary points by $C = \{s_1, s_2, \dots, s_{|C|}\}$. As g is continuous, and $\lim_{x\to-\infty} g(x) = +\infty$ and $\lim_{x\to\infty} g(x) = -\infty$ due to Lemma 3, it follows that the number of crossing points $|C|$ must be odd and g crosses zero from above at the odd points $H = \{s_1, s_3, \dots, s_{|C|}\}$. The domain of attraction for each of these points s_k for the ODE (11) is the interval

$$A_k = (s_{k-1}, s_{k+1}),$$

where we set $s_0 = -\infty$ and $s_{|C|+1} = +\infty$, i.e., $x(0) \in A_k$ implies $\lim_{t\to\infty} x(t) = s_k$. We observe for every $s_k \in H$ and for every $\delta > 0$, that if the initial value of the ODE (11) satisfies

$$\big|x(0) - s_k\big| \le \min\big\{\delta, |s_k - s_{k-1}|, |s_{k+1} - s_k|\big\}$$

then the path satisfies $|x(t) - s_k| < \delta$ for all $t > 0$ and $\lim_{t\to\infty} x(t) = s_k$. Thus, all the stationary points in $H \subseteq C$ are *asymptotically stable* in the sense of Liapunov.[20]

First, note that we can rewrite (5) as

$$\tilde{\theta}_{t+1} = \tilde{\theta}_t + \gamma_t[\tilde{b}_{t+1} - \tilde{\theta}_t], \tag{12}$$

where $\gamma_t = \frac{1}{\alpha+1+t}$. The convergence follows from Theorem 2.1 of Kushner and Yin (2003, p. 127) applied to dynamics given in (12). These dynamics are a special case of the dynamics given in Kushner and Yin (2003, equation (1.2), p. 120). As we established in Lemma 5 that the process $(\tilde{\theta}_t)$ is bounded with probability 1, we can apply the theorem for the case in which the dynamics are bounded with probability 1 and there is no constraint set. To do so, we have to verify that the assumptions for the theorem apply.

Below we follow the enumeration of assumptions used in Kushner and Yin and verify them one by one:

(1.1) The theorem requires $\sum_{t=1}^{\infty} \gamma_t = \infty$ and $\lim_{t\to\infty} \gamma_t = 0$. This is immediate from the definition of $\gamma_t = \frac{1}{1+t+\alpha}$.

[20]A set A is said to be asymptotically stable in the sense of Liapunov if for each $\delta > 0$, there exists $\delta_1 > 0$ such that all trajectories starting in the δ_1 neighborhood of A never leave the δ neighborhood of A and all trajectories ultimately go to A (see the definition in Kushner and Yin 2003, p. 104).

(A2.1) The theorem requires that $\sup_t \mathbb{E}[|\tilde{b}_{t+1} - \tilde{\theta}_t|^2] < \infty$, where the expectation conditions on the true state and is taken at time 0. We have that

$$\tilde{b}_{t+1} - \tilde{\theta}_t = (\tilde{b}_{t+1} - b_{t+1}) + (b_{t+1} - \mu) + (\mu - \theta_t) + (\theta_t - \tilde{\theta}_t),$$

where θ_t was defined in (10). By the triangle inequality for the L_2 norm, we have that

$$\begin{aligned}\sqrt{\mathbb{E}\big[|\tilde{b}_{t+1} - \tilde{\theta}_t|^2\big]} &\le \sqrt{\mathbb{E}\big[|\tilde{b}_{t+1} - b_{t+1}|^2\big]} + \sqrt{\mathbb{E}\big[|b_{t+1} - \mu|^2\big]} \\ &\quad + \sqrt{\mathbb{E}\big[|\mu - \theta_t|^2\big]} + \sqrt{\mathbb{E}\big[|\theta_t - \tilde{\theta}_t|^2\big]} \\ &\le 2\Delta + \operatorname{var}(b_1) + \sqrt{\mathbb{E}\big[|\mu - \theta_t|^2\big]}.\end{aligned}$$

The second to last line follows as $|\theta_t - \tilde{\theta}_t| \le \Delta$ and $|b_t - \tilde{b}_t| \le \Delta$ by Lemma 4, and $\mu = \mathbb{E}[b_{t+1}]$ and $\operatorname{var}(b_{t+1}) = \operatorname{var}(b_1)$. It thus suffices to show that

$$\sup_t \mathbb{E}\big[|\mu - \theta_t|^2\big] < \infty.$$

To show this, we can use Young's inequality and the fact that $\mathbb{E}[b_t] = \mu$,

$$\begin{aligned}\mathbb{E}\big[|\mu - \theta_t|^2\big] &= \mathbb{E}\left[\left|\frac{\alpha\tilde{\theta}_0 + t\mu}{\alpha + t} - \theta_t + \frac{\alpha(\mu - \tilde{\theta}_0)}{\alpha + t}\right|^2\right] \\ &\le 2\mathbb{E}\left[\left|\frac{\alpha\tilde{\theta}_0 + t\mu}{\alpha + t} - \theta_t\right|^2\right] + 2\left(\frac{\alpha(\mu - \tilde{\theta}_0)}{\alpha + t}\right)^2 \\ &\le 2\operatorname{var}(\theta_t) + 2(\mu - \tilde{\theta}_0)^2 = 2\frac{t}{(\alpha + t)^2}\operatorname{var}(b_1) + 2(\mu - \tilde{\theta}_0)^2 \\ &\le 2\operatorname{var}(b_1) + 2(\mu - \tilde{\theta}_0)^2 < \infty.\end{aligned}$$

(A2.2) The theorem requires the existence of a function $g : \mathbb{R} \to \mathbb{R}$ and a sequence of random variables $(\beta_t)_t$ such that $\mathbb{E}[\tilde{b}_{t+1} - \tilde{\theta}_t \mid \tilde{\theta}_t] = g(\tilde{\theta}_t) + \beta_t$. Hence, we define

$$\beta_t = \mathbb{E}[\tilde{b}_{t+1} - \tilde{\theta}_t \mid \tilde{\theta}_t] - g(\tilde{\theta}_t) = g(t, \tilde{\theta}_t) - g(\tilde{\theta}_t),$$

where $g(t, x) = \mathbb{E}[\tilde{b}(t, b_t, x) - x]$ and $g(x) = \lim_{t\to\infty} g(t, x)$ as defined in (6).

(A2.3) The theorem requires $g(\tilde{\theta})$ to be continuous in $\tilde{\theta}$, which follows as $x \mapsto \tilde{b}(t, b, x)$ is continuous.

(A2.4) The theorem requires $\sum_{t=1}^{\infty}(\gamma_t)^2 < \infty$ which is immediate from the definition of $\gamma_t = \frac{1}{1+t+\alpha}$.

(A2.5) The theorem requires $\sum_{t=1}^{\infty} \gamma_t|\beta_t| < \infty$ with probability 1. By assumption (7), we have that $|\beta_t| \le d_1 \frac{1}{t^k}$ for some $k > 1$. We, thus, have that

$$\sum_{t=1}^{\infty} \gamma_t|\beta_t| \le d_1 \sum_{t=1}^{\infty} t^{-(k+1)} < \infty.$$

(A2.6) The theorem requires the existence of a real-valued function f that satisfies $f'(\tilde{\theta}) = -g(\tilde{\theta})$ and that f is constant on each connected subset of $C = \{x\colon g(x) = 0\}$. We thus define f by

$$f(\tilde{\theta}) = -\int_{\tilde{\theta}_0}^{\tilde{\theta}} g(z)\,dz.$$

As g is bounded from above and below by a linear function, and as g is continuous by Lemma 3, f is well defined. Whereas C is finite, H is finite and f trivially satisfies all conditions of the theorem.

Theorem 2.1 in Kushner and Yin (2003, p. 127) is, thus, applicable to the process and (in our notation) can be stated as follows:

> If $s_k \in C$ is locally asymptotically stable in the sense of Liapunov for (11) and $\tilde{\theta}_t$ is in some compact set in the domain of attraction A_k of s_k infinitely often with probability $\geq \rho$, then $\tilde{\theta}_t \to s_k$ with at least probability ρ. Suppose that (A2.6) holds. Then, for almost all $(\tilde{\theta}_t)$ converges to a unique point in C.

By the second part of the theorem, we know that $\tilde{\theta}_t$ converges almost surely to a point in C. We show that it cannot converge to a point in $C \setminus H$. Suppose, toward a contradiction, that $\lim_{t\to\infty} \tilde{\theta}_t = s_k \in C \setminus H$. To begin, we show that either $\tilde{\theta}_t \in A_{k-1}$ infinitely often or $\tilde{\theta}_t \in A_{k+1}$ infinitely often; either this is the case or $\tilde{\theta}_t = s_k$ infinitely often. Now observe that as the distribution of b_t admits a density, and $b \mapsto \tilde{b}_t(t, b, \tilde{\theta})$ is strictly increasing and differentiable, the objective distribution of $\tilde{b}_t = \tilde{b}(t, b, \tilde{\theta})$ admits a density and an interval that contains more than a single point as support. Thus, the distribution of $\tilde{\theta}_{t+1}$ admits a density conditional on every t, $\tilde{\theta}_t$.

Hence, for every unstable stationary point $s_k \in C \setminus H$, we have that the probability that $\tilde{\theta}_{t+1}$ equals s_k is zero ($\mathbb{P}[\tilde{\theta}_{t+1} = s_k | \tilde{\theta}_t = s_k] = 0$) and, hence, we have that whenever $\lim_{t\to\infty} \tilde{\theta}_t = s_k$, the mean belief $\tilde{\theta}_t$ is either in A_{k-1} or in A_{k+1} infinitely often with probability 1. Theorem 2.2 in Kushner and Yin (2003, p. 131) makes the following statement under the conditions of Theorem 2.1:

> Let the set A be locally asymptotically stable in the sense of Liapunov. Suppose that $(\tilde{\theta}_t)_t$ visits a compact set in the domain of attraction of A infinitely often. Then $\tilde{\theta}_t \to A$ with probability 1.

Thus, as we have shown that $\lim_{t\to\infty} \tilde{\theta}_t = s_k$ implies that $\tilde{\theta}_t$ is infinitely often in the domain of attraction of either s_{k-1} or s_{k+1} we have that $\lim_{t\to\infty} \tilde{\theta}_t = s_{k+1}$ or $\lim_{t\to\infty} s_{k-1}$ with probability 1, which contradicts $\lim_{t\to\infty} \tilde{\theta}_t = s_k$. Hence, $(\tilde{\theta}_t)$ cannot converge to a point in $C \setminus H$ with positive probability. □

A.2 *Proving Theorem 1*

Proof of Lemma 1. We established in the text that the agent's posterior belief at the beginning of period t is that Θ is normally distributed with mean $\tilde{\theta}_{t-1}$ and variance v_{t-1}. As the agent believes that the information she receives is independent of her action, she behaves myopically and maximizes the expected flow payoff in each period. As $M(a, \Theta)$

is the expected payoff of the agent when the state equals Θ and as the agent believes the state to be normally distributed with mean $\tilde{\theta}_{t-1}$ and variance v_{t-1}, her action in period t satisfies

$$a_t \in \operatorname*{argmax}_{a\in A} \int M(a,x)\phi(x;\tilde{\theta}_{t-1},v_{t-1})\,dx.$$

As $\tilde{Q}$ is strictly concave in a, so is the expectation M, and, thus, there is a unique optimal action $a^*(t,\tilde{\theta}_{t-1})$ whenever it exists. Since for all b_t, $\lim_{a\to\underline{a}} \tilde{Q}_a(a,b_t) < 0$ and $\lim_{a\to\overline{a}} \tilde{Q}_a(a,b_t) > 0$, an optimal action exists and is characterized by the first-order condition

$$0 = \int M_a(a,x)\phi(x;\tilde{\theta}_{t-1},v_{t-1})\,dx. \tag{13}$$

As $|\tilde{Q}_{ab}|$ and $|\tilde{Q}_{aa}|$ are integrable with respect to a normal distribution over the external state, we can apply the dominated convergence theorem to show that the right-hand of (13) is differentiable in a and θ. As $\tilde{Q}$ is strictly concave, it follows that $M_{aa} < 0$. We can, thus, apply the implicit function theorem to get that the optimal action is differentiable in the mean belief of the agent $\tilde{\theta}$ and that $a^*_{\tilde{\theta}}(t,\tilde{\theta})$ is given by

$$a^*_{\tilde{\theta}}(t,\tilde{\theta}) = -\frac{\int M_{a\tilde{\theta}}(a,x)\phi(x;\tilde{\theta}_{t-1},v_{t-1})\,dx}{\int M_{aa}(a,x)\phi(x;\tilde{\theta}_{t-1},v_{t-1})\,dx}.$$

The result for $a^*(\tilde{\theta})$ follows from an identical argument. □

Proof of Lemma 2. We first prove that condition (i) is a sufficient condition for part (i) of Assumption 1. Let Δ_Q be a bound such that $|\tilde{Q}(a,b) - Q(a,b)| \le \Delta_Q$ and let $\underline{\tilde{Q}}_b$ be a bound such that $\tilde{Q}_b \ge \underline{\tilde{Q}}_b > 0$. For the sake of a contradiction, suppose that $|b - \tilde{b}(b,a)| > \Delta_Q/\underline{\tilde{Q}}_b$. Then using that $|\tilde{Q}(a,\tilde{b}) - \tilde{Q}(a,b)| \ge \underline{\tilde{Q}}_b|\tilde{b} - b|$ and that $|\tilde{Q}(a,b) - Q(a,b)| \le \Delta_Q$, one has

$$\left|\tilde{Q}(a,\tilde{b}) - Q(a,b)\right| = \left|\tilde{Q}(a,\tilde{b}) - \tilde{Q}(a,b) + \tilde{Q}(a,b) - Q(a,b)\right| \ge \underline{\tilde{Q}}_b|\tilde{b} - b| - |\Delta_Q| > 0,$$

contradicting (1).

We next establish that condition (ii) implies part (ii) of Assumption 1. That $\tilde{b}(b,a)$ is differentiable in a follows from applying the implicit function theorem to (1), yielding

$$\frac{\partial \tilde{b}(b,a)}{\partial a} = \frac{Q_a(a,b) - \tilde{Q}_a(a,\tilde{b})}{\tilde{Q}_{\tilde{b}}(a,\tilde{b})}.$$

Part (ii) of Assumption 1 follows since $\tilde{Q}_{\tilde{b}}(a,\tilde{b})$ is uniformly bounded from below, and $Q_a(a,b)$ and $\tilde{Q}_a(a,\tilde{b})$ are uniformly bounded from above and below.

We next establish that condition (iii) implies part (iii) of Assumption 1. We established in the text that the agent believes at the end of period $t-1$ that Θ is normally

distributed with mean $\tilde{\theta}_{t-1}$ and variance

$$v_{t-1} = \frac{1}{v_0^{-1} + (t-1)\sigma^{-2}}.$$

As conditional on Θ, b_t has mean zero and variance σ^2, the agent believes that $\tilde{b}_t$ is normally distributed with variance $\sigma^2 + v_{t-1}$ and mean $\tilde{\theta}_{t-1}$. Thinking of the optimal action as a function of the mean belief $\tilde{\theta}$ and variance v of the agent's belief about Θ, we can use the fact that a^* is interior and the output $\tilde{Q}$ is differentiable with respect to the action to rewrite (13) as

$$0 = \mathbb{E}\big[\tilde{Q}_a\big(a^*(v, \tilde{\theta}), \tilde{\theta} + \eta\sqrt{\sigma^2 + v}\big) \mid \eta \sim \mathcal{N}(0,1)\big].$$

Because by assumption $|\tilde{Q}_{ab}|$ is bounded and $|\tilde{Q}_{aa}|$ is integrable with respect to a normal distribution over the external state, we can apply the dominated convergence theorem to show that the right-hand side of (13) is differentiable in a and η. We can, thus, apply the implicit function theorem to get

$$\frac{\partial a^*(v, \tilde{\theta})}{\partial v} = -\frac{\mathbb{E}\Big[\tilde{Q}_{ab}\big(a^*(v, \tilde{\theta}), \tilde{\theta} + \eta\sqrt{\sigma^2 + v}\big) \times \dfrac{\eta}{2\sqrt{\sigma^2 + v}} \;\Big|\; \eta \sim \mathcal{N}(0,1)\Big]}{\mathbb{E}\big[\tilde{Q}_{aa}\big(a^*(v, \tilde{\theta}), \tilde{\theta} + \eta\sqrt{\sigma^2 + v}\big) \mid \eta \sim \mathcal{N}(0,1)\big]}.$$

Using the fact that $\tilde{Q}_{aa}$ is uniformly bounded away from zero and denoting this bound by $\underline{c}_{aa}$, we get that

$$-\mathbb{E}\big[\tilde{Q}_{aa}\big(a^*(v, \tilde{\theta}), \tilde{\theta} + \eta\sqrt{\sigma^2 + v}\big) \mid \eta \sim \mathcal{N}(0,1)\big] \geq |\underline{c}_{aa}|.$$

As there exists a constant $\overline{c}_{a,b} > 0$ such that $|\tilde{Q}_{ab}| \leq \overline{c}_{ab}$, we get that

$$\begin{aligned}
&\left|\mathbb{E}\left[\tilde{Q}_{ab}\big(a^*(v, \tilde{\theta}), \tilde{\theta} + \eta\sqrt{\sigma^2 + v}\big) \times \frac{\eta}{2\sqrt{\sigma^2 + v}} \;\Big|\; \eta \sim \mathcal{N}(0,1)\right]\right| \\
&\quad \leq \overline{c}_{ab}\mathbb{E}\left[\frac{|\eta|}{2\sqrt{\sigma^2 + v}} \;\Big|\; \eta \sim \mathcal{N}(0,1)\right] \\
&\quad = \overline{c}_{ab}\frac{\sqrt{\dfrac{2}{\pi \mathrm{e}}}}{2\sqrt{\sigma^2 + v}} \leq \overline{c}_{ab}\frac{\sqrt{\dfrac{2}{\pi \mathrm{e}}}}{2\sqrt{\sigma^2}}.
\end{aligned}$$

We, thus, have that

$$\left|\frac{\partial a^*(v, \tilde{\theta})}{\partial v}\right| \leq \frac{\overline{c}_{ab}}{|\underline{c}_{aa}|}\frac{\sqrt{\dfrac{2}{\pi \mathrm{e}}}}{2\sqrt{\sigma^2}} =: \kappa.$$

Observing that $\lim_{t\to\infty} v_t = 0$, we, thus, have that

$$\big|a^*(v_t, \theta) - a^*(0, \theta)\big| = \left|\int_0^{v_t} \frac{\partial a^*(z, \tilde{\theta})}{\partial z}\,\mathrm{d}z\right| \leq \int_0^{v_t}\left|\frac{\partial a^*(z, \tilde{\theta})}{\partial z}\right| dz$$

$$\leq \kappa v_t = \frac{\kappa}{v_0^{-1} + t\sigma^{-2}} \leq \frac{1}{t} d$$

for $d = \frac{\kappa}{v_0 + \sigma^{-2}}$. □

Proof of Theorem 1. We prove Theorem 1 by showing that the conditions of Theorem 2 are satisfied. For this purpose, we define $\tilde{b}(t, b, \tilde{\theta})$ as the signal the agent believes to have observed when the true signal equals b and the agent took the optimal action in period t:[21]

$$\tilde{b}(t, b, \tilde{\theta}) = \tilde{b}\big(b, a^*(t, \tilde{\theta})\big).$$

We note that $|\tilde{b}(t, b, \tilde{\theta}) - b| \leq \Delta$ by Assumption 1. As a^* is differentiable by Lemma 1 and $\tilde{b}_a$ is bounded by Assumption 1, it follows that $x \mapsto \tilde{b}(t, b, x)$ is continuous. This implies that

$$g(t, x) = \mathbb{E}\big[\tilde{b}(t, b_t, x) - x\big]$$

is continuous in x. Define $\hat{g}(x) = \lim_{t\to\infty} g(t, x)$. We want to show that $\hat{g}$ is well defined and that $\hat{g}(x) = g(x) = \mathbb{E}[\tilde{b}(b, a^*(x)) - x]$ Observe that by parts (ii) and (iii) of Assumption 1, there exist constants $d_1, d_2 > 0$:

$$\begin{aligned}\big|g(t, x) - g(x)\big| &= \big|\mathbb{E}\big[\tilde{b}\big(b, a^*(t, \tilde{\theta})\big) - \tilde{b}\big(b, a^*(\tilde{\theta})\big)\big]\big| \leq \mathbb{E}\big[\big|\tilde{b}\big(b, a^*(t, \tilde{\theta})\big) - \tilde{b}\big(b, a^*(\tilde{\theta})\big)\big|\big] \\ &\leq d_1 \mathbb{E}\big[\big|a^*(t, \tilde{\theta}) - a^*(\tilde{\theta})\big|\big] \leq d_1 d_2 \frac{1}{t^k}.\end{aligned}$$

Hence, $\hat{g} = g$ and, furthermore, (7) holds. □

A.3 *Proof of Proposition 1*

Proof of Proposition 1. As Q is strictly concave in a, it follows that for any belief ζ there is a unique strict best action. Thus, there can be no Berk–Nash equilibria in mixed strategies and all Berk–Nash equilibria are strict. We first note that the log-likelihood ratio simplifies to

$$L^a(b, \tilde{\theta}) = -\frac{[b - \Theta]^2}{2\eta^2} + \frac{\big[\tilde{b}(b, a) - \tilde{\theta}\big]^2}{2\eta^2} - \log\big(\tilde{b}_b(b, a)\big).$$

The Kullback–Leibler divergence for a given action a equals

$$K^{\delta_a}(\tilde{\theta}) = \mathbb{E}\big[L^a(b, \tilde{\theta})\big] = \mathbb{E}\bigg[-\frac{[b - \Theta]^2}{2\eta^2} + \frac{\big[\tilde{b}(b, a) - \tilde{\theta}\big]^2}{2\eta^2} - \log\big(\tilde{b}_b(b, a)\big)\bigg].$$

As only the second part of the sum depends on $\tilde{\theta}$ and as the square is strictly convex, we get that the set of Kullback–Leibler minimizers is given by

$$\operatorname*{argmin}_{\tilde{\theta}} K^{\delta_a}(\tilde{\theta}) = \big\{\tilde{\theta} \colon \tilde{\theta} = \mathbb{E}\big[\tilde{b}(b, a)\big]\big\}.$$

[21] Recall here that we defined $\tilde{b}(b, a)$ as the signal the agent believes to have received when taking the action a.

Hence, $a \in A$ is a Berk–Nash equilibrium if and only if $\tilde{\theta} = \mathbb{E}[\tilde{b}(b, a)]$ for $a = a^*(\tilde{\theta})$, which by the definition of C is equivalent to $a = a^*(\tilde{\theta})$ for some $\tilde{\theta} \in C$. As we have established in Theorem 1 that beliefs never converge to the points in $C \setminus H$, it follows that the agent's action never converges to the corresponding Berk–Nash equilibria. □

A.4 *Additional calculations for the mislearning new technology example*

We verify that Assumption 1 holds in the mislearning new technology example. Let $\bar{a}$ be the action above $a^*(\theta)$ at which $k(\bar{a}) = \tilde{k}(\bar{a})$. Since for all $a \geq \bar{a}$, $\log(k(a)/\tilde{k}(a)) \leq 0$, setting $\Delta = \max_{a\in[0,\bar{a}]} \log(k(a)/\tilde{k}(a))$ verifies that part (i) of Assumption 1 holds. Because $\tilde{b}_a(b, a) = (k'(a)/k(a)) - (\tilde{k}'(a)/\tilde{k}(a))$, it is bounded and, thus, part (ii) of Assumption 1 is satisfied. To verify Assumption 1 part (iii), note that the subjectively optimal action at time t is implicitly defined through

$$\tilde{k}'\big(a^*(t, \tilde{\theta})\big) \exp\left\{\tilde{\theta} + \frac{\sigma^2 + v_{t-1}}{2}\right\} - 1 = 0.$$

Equivalently thinking of the subjectively optimal action as a function of the mean belief $\tilde{\theta}$ and variance v_{t-1}, and applying the implicit function theorem, we have

$$\frac{\partial a^*(v, \tilde{\theta})}{\partial v} = -\frac{1}{2}\frac{\tilde{k}'\big(a^*(v, \tilde{\theta})\big)}{\tilde{k}''\big(a^*(v, \tilde{\theta})\big)}.$$

Let κ be a bound so that $\tilde{k}'(a)/|\tilde{k}''(a)| \leq \kappa$. Then

$$\big|a^*(t, \tilde{\theta}) - a^*(\tilde{\theta})\big| = \big|a^*(v_{t-1}, \tilde{\theta}) - a^*(0, \tilde{\theta})\big| = \left|\frac{1}{2}\int_{\sigma^2}^{\sigma^2+v_{t-1}} \frac{\tilde{k}'\big(a^*(v, \tilde{\theta})\big)}{\tilde{k}''\big(a^*(v, \tilde{\theta})\big)}\, dv\right| \leq v_{t-1}\frac{\kappa}{2}.$$

Hence, since v_t goes to zero on the order of $1/t$, so does $|a^*(t, \tilde{\theta}) - a^*(\tilde{\theta})|$, which implies that Assumption 1 part (iii) also holds.

References

Berk, Robert (1966), "Limiting behavior of posterior distributions when the model is incorrect." *Annals of Mathematical Statistics*, 37, 51–58. [73, 74, 75, 76, 79]

Bohren, Aislinn and Daniel Hauser (2020), "Learning with model misspecification: Characterization and robustness." Unpublished Paper, University of Pennsylvania. [76]

Bohren, J. Aislinn (2016), "Informational herding with model mispecification." *Journal of Economic Theory*, 163, 222–247. [76]

Bunke, Olaf and Xavier Milhaud (1998), "Asymptotic behavioir of Bayes estimates under possibly incorrect models." *Annals of Statistics*, 26, 617–644. [73, 74, 75]

Dal Bó, Ernesto, Pedro Dal Bó, and Erik Eyster (2018), "The demand for bad policy when voters underappreciate equilibrium effects." *Review of Economic Studies*, 85, 964–998. [86]

Duflo, Esther, Michael Kremer, and Jonathan Robinson (2008), "How high are rates of return to fertilizer? Evidence from field experiments in Kenya." *American Economic Review: Papers & Proceedings*, 98, 482–488. [87]

Esponda, Ignacio and Demian Pouzo (2016a), "Berk-Nash equilibrium: A framework for modeling agents with misspecified models." *Econometrica*, 84, 1093–1130. [74, 76, 81]

Esponda, Ignacio and Demian Pouzo (2016b), "Equilibrium in misspecified Markov decision process." Unpublished Paper, arXiv:1502.06901. [76]

Esponda, Ignacio, Demian Pouzo, and Yuichi Yamamoto (2019), "Asymptotic behavior of Bayesian learners with misspecified models." Unpublished Paper, arXiv:1904.08551. [76]

Frick, Mira, Ryota Iijima, and Yuhta Ishii (2019), "Misinterpreting others and the fragility of social learning." Working Paper, Cowles Foundation Discussion Paper No. 2160. [76]

Frick, Mira, Ryota Iijima, and Yuhta Ishii (2020), "Stability and robustness in misspecified learning models." Working Paper, Cowles Foundation Discussion Paper No. 2235. [76]

Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack (2020), "Limits points of endogenous misspecified learning." Unpublished Paper, SSRN 3553363. [76]

Fudenberg, Drew, Gleb Romanyuk, and Philipp Strack (2017), "Active learning with a misspecified prior." *Theoretical Economics*, 12, 1155–1189. [74, 75, 76, 89]

He, Kevin (2018), "Mislearning from censored data: The gambler's fallacy in optimal-stopping problems." Unpublished Paper, arXiv:1803.08170. [76]

Heidhues, Paul, Botond Kőszegi, and Philipp Strack (2018), "Unrealistic expectations and misguided learning." *Econometrica*, 86, 1159–1214. [74, 75, 90]

Kiefer, J. and J. Wolfowitz (1952), "Stochastic estimation of the maximum of a regression function." *Annals of Mathematical Statistics*, 23, 462–466. [79]

Kushner, Harold J. and G. George Yin (2003), *Stochastic Approximation and Recursive Algorithms and Applications*. Springer. [79, 92, 94]

Nyarko, Yaw (1991), "Learning in mis-specified models and the possibility of cycles." *Journal of Economic Theory*, 55, 416–427. [74, 75, 89]

Robbins, Herbert and Sutton Monro (1951), "A stochastic approximation method." *Annals of Mathematical Statistics*, 22, 400–407. [79]

Shalizi, Cosma Rohilla (2009), "Dynamics of Bayesian updating with dependent data and misspecified models." *Electronic Journal of Statistics*, 3, 1039–1074. [73, 74, 75, 79]

Co-editor Ran Spiegler handled this manuscript.

Manuscript received 10 December, 2018; final version accepted 2 June, 2020; available online 8 June, 2020.