# Convergence in Misspecified Learning Models with Endogenous Actions*

Paul Heidhues
DICE

Botond Kőszegi
Central European University

Philipp Strack
UC Berkeley

December 5, 2018

**Abstract**

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. The crucial assumptions of our model are that the state and action spaces are continuous, the state has a unidirectional effect on output, and the prior and noise are normal. These assumptions imply that the agent's posterior admits a one-dimensional summary statistic, allowing us to apply tools from stochastic approximation theory to establish convergence. Applications of our framework include learning by a person who has an incorrect model of a technology she uses, is subject to confirmatory bias, conservatism, or base-rate neglect, or is overconfident about herself, learning by a representative agent who misunderstands macroeconomic outcomes, as well as learning by a firm that has an incorrect parametric model of demand.

*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*, whose hospitality is gratefully acknowledged.

# 1 Introduction

A significant literature in statistics studies inferences by an observer who has a misspecified model of the world and receives exogenous signals. It has been shown that in this setting, the observer's beliefs converge under weak conditions (Berk, 1966, Shalizi, 2009). In many or most economic applications, however, a person is not 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. For this type of learning with misspecification, little is known about the convergence of beliefs, and previous work has identified situations in which beliefs do not converge (e.g., Nyarko, 1991, Fudenberg, Romanyuk and Strack, 2017).

In this paper, we establish convergence of beliefs to a point belief, and convergence of actions, in any one-dimensional misspecified model with endogenous actions in which (i) the state and action spaces are continuous; (ii) output is increasing in the state; and (iii) all actions are equally informative in the agent's mind. This framework is sufficiently general to cover a number of economically relevant settings. In addition, we argue that if any of our central assumptions is removed, reasons for non-convergence described in the literature can arise, so that our convergence result is in a sense (albeit in a rather weak sense) tight.

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 $\Theta$. The agent does not directly observe $\Theta$ (or $b_t$), but attempts to learn about it from her observed outputs and to adjust her action optimally in response. Crucially, the agent has a misspecified model: she believes that output is determined according to $\tilde{Q}(a_t, b_t)$. We assume that $Q$ and $\tilde{Q}$ are strictly monotonic in $b_t$, and — to guarantee that the agent can always find an explanation for her observations — we also suppose that fixing $a_t$, the range of $\tilde{Q}(a_t, b_t)$ is at least as large as the range of $Q(a_t, b_t)$.

In Section 4, we identify a few potential economic applications of our abstract framework. The agent might be learning about a technological parameter $\Theta$ — 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 $\Theta$ and set optimal prices $a_t$ with a misspecified functional form for demand. The agent might be a representative consumer seeking to adjust her behavior to macroeconomic conditions she misunderstands. And through a rewriting, our model can also be used to capture confirmatory bias, conservatism, and base-rate neglect — updating mistakes in which the agent does not choose actions, but misperceives signals in a way that depends on her current beliefs.

In Section 5, we apply tools from stochastic approximation theory to show that under economically weak technical conditions on $Q$, the agent's beliefs converge with probability one to a point belief, and her actions also converge. Within our setting, this convergence result substantially supersedes previous ones in the literature, including those of Berk (1966) and Shalizi (2009) — where the agent does not make 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, Kőszegi and Strack (2018) — where the convergence proof requires much more structure on the problem. As a preliminary observation, note that for any output $q_t$, there is a unique state $\tilde{b}_t$ that the agent believes must have generated $q_t$. The agent's updating therefore 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 variable. 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 does the agent's belief. We conclude that the agent's belief converges to a point where the external conditions $\tilde{b}_t$ that she thinks she has observed on average equal to her belief about the fundamental. Based on this logic, we identify several plausible variants of our model under which convergence also obtains.

While beliefs converge with probability one, we give a simple example featuring multiple possible limiting beliefs such that the limiting belief the agent converges to depends on her initial observations. This is in contrast to the correctly specified model, where under our assumptions beliefs always converge to the true state and thus have a unique limit point. In other cases, the limiting belief is unique and false. And we show that in the specific cases of conservatism and base-rate neglect, the agent always ends up with correct long-run beliefs despite her misspecified model.

We conclude our introduction by arguing that all central features of our framework described above are necessary for convergence of beliefs and actions to obtain in general. First, our setup posits continuous state and action spaces. If this was not the case, convergence could easily fail. Fudenberg et al. (2017) describe an example (similar to Nyarko, 1991) in which a decisionmaker conceives of two states, 0 and 1, and can take two actions, -1 and 1. While -1 is optimal under state 0 and 1 is optimal under state 1, taking -1 leads the agent to accumulate evidence in favor of state 1 and taking 1 leads her to accumulate evidence in favor of state 0. Hence, if she has no other action available, her beliefs cannot converge to point beliefs, and her actions must fluctuate forever. We provide a similar example with a discrete prior but continuous actions in which beliefs as well as actions fail to converge, and non-convergence clearly relies on the discreteness. And in a variant of the example, we show that if actions are discrete but the prior is continuous, then actions do not converge, although we conjecture that beliefs do.

Second, along with the other assumptions, the feature of our model that noise is inside the output function ensures that in the agent's view all actions are equally informative. This eliminates the experimentation motive. It is well known from basic bandit problems that if not all actions are equally informative, and in particular some action is not fully informative even if taken infinitely many times, then the agent's belief may not converge to a point belief. Furthermore, Fudenberg et al. (2017) also show that an experimentation motive in itself can generate non-convergence of beliefs. Consider again the previous example, but suppose that the agent has an additional action, 0, that is both optimal for intermediate beliefs and uninformative. If she is myopic, she eventually chooses this action, as with the other two actions her beliefs tend to drift to the middle. Hence, her beliefs converge in this case (albeit to non-point beliefs). But if she is extremely patient, then she would like to experiment and figure out the state, and therefore she takes one of the informative actions. Then, her beliefs must fluctuate forever.

Third, if an increase in the state could have a non-monotonic effect on output, then the agent's inferences from output could confound different states, potentially leading to limited learning. As an obvious example, if $\tilde{Q}$ was completely flat in $b$, then the agent's beliefs of course do not converge to a point belief.

## 2 Related Literature

The classical statistics literature on learning with misspecified models, such as Berk (1966) 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 as we have discussed above, 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 an earlier paper (Heidhues et al., 2018), we study an overconfident agent's learning process and behavior when she chooses actions endogenously based on her beliefs, showing that overconfidence leads her to make mistaken inferences about the fundamental that are often self-defeating. Our earlier paper establishes convergence of beliefs through a different method and in a different setting, in particular restricting attention to situations in which there can only be one limiting belief and heavily relying on the particular structure that overconfidence puts on $Q$ and $\tilde{Q}$. The current paper allows for more general technologies and misspecifications.

Esponda and Pouzo (2016a) develop a general framework for studying repeated games in which players have misspecified models, and Esponda and Pouzo (2016b) extend the framework to general dynamic single-agent Markov decision problems with non-myopic agents. Building on Berk (1966), Esponda and Pouzo (2016a) establish that if actions converge, 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. In contrast to our results, due to the discrete nature of the state space, beliefs need not converge.

## 3 Learning Environment

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

### 3.1 Setup

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 $a_t \in (\underline{a}, \overline{a})$ and an unobservable external state $b_t \in \mathbb{R}$ beyond her control. We assume that $b_t = \Theta + \epsilon_t$, where $\Theta \in \mathbb{R}$ is an underlying fixed fundamental and the $\epsilon_t$ are independent normally distributed random variables with mean zero and variance $\sigma_\epsilon^2$, precision $h_\epsilon = 1/\sigma_\epsilon^2$. The agent's prior is that $\Theta$ is distributed normally with mean $\tilde{\theta}_0$ and variance $\sigma_0^2$, precision $h_0 = 1/\sigma_0^2$. While the agent understands the basic environment correctly, she has a misspecified model regarding output: she believes that output is being produced according to $\tilde{Q}(a_t, b_t)$. Given her model, the agent updates her belief 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 is 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: $\tilde{Q}_b > 0$.[1] 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:

**Assumption 1** (Technical Assumptions on Misspecified Model)**.** (i) $\tilde{Q}_{aa} < 0$, and $\lim_{a \to \underline{a}} \tilde{Q}_a(a, b) > 0 > \lim_{a \to \overline{a}} \tilde{Q}_a(a, b)$ for all $b$; (ii) $|\tilde{Q}_{ab}|$ and $|\tilde{Q}_{aa}|$ are integrable with respect to any normal distribution over the second component;[2] and (iii) there exists a $\kappa$ such that $|\tilde{Q}_{ab}(a, b)|/|\tilde{Q}_{aa}(a, b)| \le \kappa$.

Part (i) guarantees that there is always a unique myopically optimal action. Part (ii) is a very weak technical condition that implies that we can take subjective expectations over the second derivatives. Part (iii) implies a bound on the influence of small changes in the agent's mean belief on the perceived optimal action.

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

[2] For any $a, b' \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$ we have that $\int_{\mathbb{R}} (|\tilde{Q}_{aa}(a, x)| + |\tilde{Q}_{ab}(a, x)|) \, d\Phi_{b', \sigma^2}(x) < \infty$ where $\Phi_{b', \sigma^2}$ is the Normal distribution with mean $b'$ and variance $\sigma^2$.

### 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 a $\tilde{b}_t$ satisfying

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

By assumption, there is a $\tilde{b}_t$ satisfying Equation (1), and since $\tilde{Q}_b > 0$, it is unique. Therefore, the agent believes that whatever action she chooses, she will infer the same signal $\tilde{b}_t$ about $\Theta$. This implies that she chooses her action in each period to maximize that period's perceived expected output. In addition, since the agent believes that $\tilde{b}_s \sim \mathcal{N}(\Theta, h_\epsilon^{-1})$, at the end of period $t \geq 1$ she believes that $\Theta$ is distributed with mean

$$\tilde{\theta}_t = \sum_{s=1}^{t} \frac{h_\epsilon \tilde{b}_s + h_0 \tilde{\theta}_0}{t h_\epsilon + h_0} \tag{2}$$

and precision $h_0 + t h_\epsilon$. Hence, at the beginning of period $t$, the agent believes that $\tilde{b}_t$ is normally distributed with mean $\tilde{\theta}_{t-1}$ and variance $h_\epsilon^{-1} + [h_0 + (t-1)h_\epsilon]^{-1}$. Let $\tilde{\mathbb{E}}_{t-1,\tilde{\theta}_{t-1}}$ be her subjective expectation with respect to this distribution.

**Lemma 1** (Agent Chooses Myopic Decision Rule)**.** *Suppose Assumption 1 holds. In each period $t$, there is a unique optimal action given by*

$$a^*(t, \tilde{\theta}_{t-1}) = \arg\max_a \; \tilde{\mathbb{E}}_{t-1,\tilde{\theta}_{t-1}} \left[\tilde{Q}(a, \tilde{b})\right]. \tag{3}$$

*Furthermore, the optimal action is differentiable in the agent's mean belief $\tilde{\theta}_{t-1}$, and the corresponding derivative is bounded.*

The proof of Lemma 1 also implies that there is a unique confident action

$$a^*(\tilde{\theta}) = \arg\max_a \tilde{\mathbb{E}}_{\infty,\tilde{\theta}} \left[\tilde{Q}(a, \tilde{b})\right]$$

that the agent perceives as optimal if she is confident that the fundamental is $\tilde{\theta}$, and that this action is differentiable in $\tilde{\theta}$ with a bounded derivative.

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

**Assumption 2** (Relationship Between True and Misspecified Models)**.** (i) $|b - \tilde{b}(b,a)| \leq \Delta$; and (ii) $\tilde{b}_a(b,a)$ is bounded.

Part (i) imposes that the agent's misinference about the realized external state is bounded by $\Delta$, guaranteeing that in the long run her beliefs are in a bounded interval. A sufficient condition for Part (i) is that $|\tilde{Q}(a,b) - Q(a,b)|$ is bounded and $\tilde{Q}_b$ is bounded from above and below by positive numbers.[3] But as our fertilizer example below shows, neither of these conditions is necessary. Part (ii), a technical condition, bounds the derivative of the misinference with respect to the chosen action.[4]

## 4 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 below to make some simple points in them.

*Mislearning New Technology.* Our model can capture learning about a technology over time. The agent's output is determined by the fundamental value ($\Theta$) of the technology, and she must decide in each period how intensively to use the technology ($a_t > 0$). But she may well have a misspecified model of how the inputs lead to output. 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, Kremer and Robinson, 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 $\Theta$ 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

[3] To see this formally, let $\Delta_Q$ be a bound such that $|\tilde{Q}(a,b) - Q(a,b)| \leq \Delta_Q$ and let $\underline{\tilde{Q}}_b$ be a bound such that $\tilde{Q}_b \geq \underline{\tilde{Q}}_b > 0$. For the sake of a contradiction, suppose that $|b - \tilde{b}(b,a)| > \Delta_Q / \underline{\tilde{Q}}_b$. Then, using that $|\tilde{Q}(a,\tilde{b}) - \tilde{Q}(a,b)| \geq \underline{\tilde{Q}}_b |\tilde{b} - b|$ and that $|\tilde{Q}(a,b) - Q(a,b)| \leq \Delta_Q$, one has

$$|\tilde{Q}(a,\tilde{b}) - Q(a,b)| = |\tilde{Q}(a,\tilde{b}) - \tilde{Q}(a,b) + \tilde{Q}(a,b) - Q(a,b)| \geq \underline{\tilde{Q}}_b |\tilde{b} - b| - |\Delta| > 0,$$

contradicting Equation (1).

[4] That $\tilde{b}(b,a)$ is differentiable in $a$ follows from applying the implicit function theorem to Equation (1), yielding $\partial\tilde{b}/\partial a = (Q_a(a,b) - \tilde{Q}_a(a,\tilde{b}))/\tilde{Q}_{\tilde{b}}(a,\tilde{b})$.

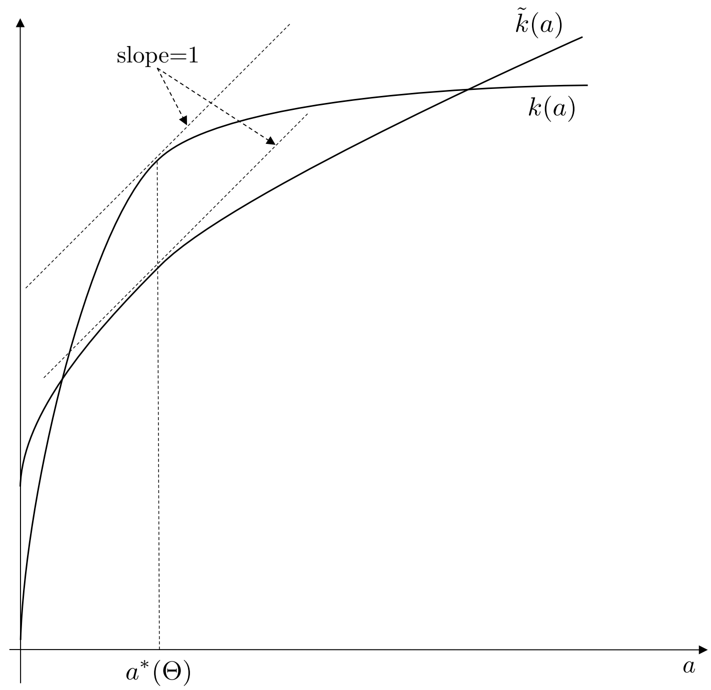


Figure 1: Actual and Estimated Returns to Fertilizer Use.
Note: since $\Theta + \sigma_\epsilon^2 = 0$ and $c = 1$, the optimal action $a^*(\Theta)$ satisfies $k'(a^*(\Theta)) = 1$.

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 $k$: $\tilde{Q}(a_t, b_t) = \tilde{k}(a_t)\exp b_t - ca_t$. The functions $k$ and $\tilde{k}$ are illustrated in Figure 1. The actual technology is quite concave, so that there is a small range of usage amounts that are optimal for a large range of $\Theta$ and $c$. Perhaps because she is familiar with more traditional technologies, however, the agent expects a less stark response pattern to usage. Formally, we suppose that $\Theta + \sigma_\epsilon^2/2 = 0$ and $c = 1$, $\tilde{k}, k > 0$, $\lim_{a\to 0}\tilde{k}' = \lim_{a\to 0} k' = \infty$, $k'' < \tilde{k}'' < 0$, and there exists an action $\bar{a}$ above the (objectively) optimal one for which $k(\bar{a}) = \tilde{k}(\bar{a})$.[5] In our example, hence, if the agent knew or could learn the true fundamental, then despite her misspecification she would choose the optimal action. To ensure that our example satisfies Assumptions 1 and 2, we furthermore suppose that $\tilde{k}'(a)/|\tilde{k}''(a)|$ is bounded from above, and that $(k'(a)/k(a)) - (\tilde{k}'(a)/\tilde{k}(a))$ is bounded from above.

*Confirmatory Bias and Other Belief-Dependent Information-Processing Mistakes.* Although it

[5] To see why the optimal action satisfies $k'(a) = 1$, note that the mean of a log-normal distribution with parameters $\mu$ and $\sigma^2$ is $\exp(\mu + \sigma^2/2)$, so that the expectation of $\exp(b_t)$ is $\exp(\Theta + \sigma_\epsilon^2/2) = 1$.

is superficially a different problem, our framework can also be used to model the agent's updating when she does not take (endogenous) actions, but how she perceives information depends on her current beliefs. An important example of such a bias is confirmatory bias, which Rabin and Schrag (1999) define as a "cognitive bias that leads individuals to misinterpret new information as supporting previously held hypotheses." A considerable body of evidence from psychology is consistent with confirmatory bias; see, e.g., Bruner and Potter (1964), Lord, Ross and Lepper (1979), Benjamin (2019). We interpret confirmatory bias as updating that is biased toward the general direction, and not the precise level, of current beliefs. To take Lord, Ross, and Lepper's main experimental setting, for instance, those broadly in support of capital punishment are too prone to interpret evidence as favoring capital punishment.

To model confirmatory bias, suppose that $\tilde{\theta}_0 = 0$, $\tilde{Q}(a,b) = b - L(a-b)$ and $Q(a,b) = (b + m(a)) - L(a - (b + m(a)))$, where $L$ is a symmetric loss function satisfying $|L'(x)| < k < 1$ for all $x$, and $m$ is a twice differentiable function. Given that the agent's posterior is always symmetric and that she has no experimentation motive, she chooses the action corresponding to her mean beliefs: $a_t = \tilde{\theta}_{t-1}$. Hence, we get that $\tilde{Q}(a_t, b_t) = b_t - L(\tilde{\theta}_{t-1} - b_t)$, and $Q(a_t, b_t) = (b_t + m(\tilde{\theta}_{t-1})) - L(\tilde{\theta}_{t-1} - (b_t + m(\tilde{\theta}_{t-1})))$. By setting up a problem in which the action is a simple function of beliefs, therefore, we have transformed an updating problem that depends on actions to one that depends on beliefs. The term $m(\tilde{\theta}_{t-1})$ can then be thought of as a bias in the agent's perception of information that she does not account for when she updates. While the bias can in principle depend on the agent's current beliefs in an arbitrary way, confirmatory bias most naturally corresponds to the case in which $m(a)$ is increasing in $a$ with $m(a) > 0$ for $a > 0$ and $m(a) < 0$ for $a < 0$. It is easy to verify that if $m(a)$ and $m'(a)$ are bounded from above and below, this example satisfies Assumptions 1 and 2.

A similar formalism can be used to capture conservatism and base-rate neglect, whereby a person's updating underweights the signal or prior, respectively (Benjamin, 2019). Suppose that $\tilde{Q}(a_t, b_t) = b_t - L(a_t - b_t)$ as above, but in a variant to the above, $Q(a_t, b_t) = (b_t + m(b_t - a_t)) - L(a_t - (b_t + m(b_t - a_t)))$. Then, $m(b_t - a_t) = m(b_t - \tilde{\theta}_{t-1})$ is the agent's bias in perceiving new information, so that the bias depends on how the signal relates to her newest belief. Conservatism bias can be modeled by assuming that $-x < m(x) < 0$ for $x > 0$ and $-x > m(x) > 0$ for $x < 0$; i.e., the agent understands the direction in which she should update her beliefs, but perceives the signal as weaker than it is. Conversely, base-rate neglect corresponds to $m(x) > 0$ for $x > 0$ and $m(x) < 0$ for $x < 0$; i.e., the agent perceives the signal as stronger than it is. Note that in this specification, the agent underweights the prior, but does not completely ignore it.

*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 $\tilde{Q}(a_t, b_t)$.

As a specific case, we consider a simplified version of the main example in Nyarko (1991) and the non-convergence result in Fudenberg et al. (2017). Suppose that true demand is $b_t - 3a_t$ 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 $Q(a_t, b_t) = a_t(b_t - 3a_t)$ 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 $\tilde{\theta}/2$. Suppose for a moment 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 non-trivial amount, so her beliefs also cannot converge to non-degenerate belief.[6] 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. Theorem 1 below implies that these non-convergence examples rely on the discreteness of the state or action space. We allow for both to be continuous, and under these circumstances beliefs and actions converge.

[6] 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 the probability density function of a normal distribution with mean zero and variance $\sigma_\epsilon^2$ by $\phi(\cdot)$. Upon deducing $\tilde{b}(b, a)$, Bayes' Rule yields the new belief

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

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 non-vanishing probability, contradicting that beliefs converge.

*Misestimating Equilibrium Feedback.* Our framework can also be used to analyze representative-agent models in which the agent interprets macroeconomic observations incorrectly. As a simple example, suppose that $\theta$ is a measure of the state of the economy, $a_t$ is an action — such as one's consumption or housing choice — that individuals look to align with the state of the economy, and the true fundamental $\Theta$ is normalized to zero. We use the same formal model as for confirmatory bias. Agents choose actions to maximize the expectation of $\tilde{Q}(a_t, b_t) = b_t - L(a_t - b_t)$, and they observe a macroeconomic outcome — such as growth — given by $Q(a_t, b_t) = b_t + m(a_t) - L(a_t - (b_t + m(a_t)))$. While in reality the aggregate outcome is increasing in the level of individuals' choices, the representative agent does not understand this. In an alternative interpretation motivated by Dal Bó, Dal Bó and Eyster (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.

*Overconfidence.* A large literature in psychology documents, and recent economic research explores 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 $\tilde{Q}(a, b) > Q(a, b)$ — for any action and any state, the agent expects higher output than is realistic. 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.

# 5 Convergence of the Agent's Belief

## 5.1 Main Argument

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 observes and 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 belief. Within our more specific domain, our convergence result also substantially supersedes that of Esponda and Pouzo (2016a, Theorem 3), which applies only for priors close to the limiting belief and for actions that are only close to optimal. Finally, our earlier paper on misguided learning by overconfident individuals (Heidhues et al., 2018) establishes convergence using a completely different technique, which relies on more restrictive assumptions about the agent's misspecification, including that there is only one possible long-run belief and that the overconfidence distorts the subjectively optimal action away from the optimal one in the same direction as an underestimation of the state.

On the other hand, 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 at 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 belief converges.

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

$$\tilde{\theta}_{t+1} = \tilde{\theta}_t + \gamma_t \left[\tilde{b}_{t+1} - \tilde{\theta}_t\right], \text{ where } \gamma_t = \frac{h_\epsilon}{(t+1)h_\epsilon + h_0}. \tag{4}$$

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

$$g(t, \tilde{\theta}_t) = \mathbb{E}_{b_{t+1}} \left[\tilde{b}(b_{t+1}, a^*(t+1, \tilde{\theta}_t))|\Theta\right] - \tilde{\theta}_t. \tag{5}$$

The above expectation is the one of an outside observer who knows the true fundamental $\Theta$ 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 belief is updated in the limit, we define

$$g(\tilde{\theta}) = \lim_{t\to\infty} g(t, \tilde{\theta}) = \mathbb{E}_{b_{t+1}} \left[\tilde{b}(b_{t+1}, a^*(\tilde{\theta}))|\Theta\right] - \tilde{\theta}.$$

We denote by $C$ the set of points where $g$ intersects zero, i.e., the agent's mean surprise equals zero:

$$C = \{\tilde{\theta} \colon g(\tilde{\theta}) = 0\}. \tag{6}$$

Intuitively, the agent cannot have a long-run belief $\tilde{\theta}$ outside $C$ 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 upwards; and if $g(\tilde{\theta}_t) < 0$, then the agent's mean beliefs drift downwards. Hence, any limiting belief must be in $C$. Furthermore, not every belief in $C$ is stable: whenever $g$ is negative on the left of $\tilde{\theta}$ and positive on the right of $\tilde{\theta}$, the agent's perceived signals would still in expectation push her away from $\tilde{\theta}$. We thus have the following definition:

**Definition 1** (Stability)**.** A point $\tilde{\theta} \in C$ is stable if there exists a $\delta > 0$ such that $g(\theta) < 0$ for $\theta \in (\tilde{\theta}, \tilde{\theta} + \delta)$ and $g(\theta) > 0$ for $\theta \in (\tilde{\theta} - \delta, \tilde{\theta})$.

We denote by $H$ the set of stable points in $C$. The main result of the paper shows that $H$ completely describes the set of possible long-run beliefs.

**Theorem 1.** *Suppose Assumptions 1 and 2 hold, and that $C$ is finite. Then, almost surely $\tilde{\theta}_\infty = \lim_{t\to\infty} \tilde{\theta}_t$ exists and lies in $H$, which is non-empty.*

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 only depends on time $t$, we study dynamics of the agent's mean beliefs $\tilde{\theta}_t$. Although — as we have 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 $a^*(\tilde{\theta}_{s-1})$ — that is, the action the agent would choose if she was fully confident of the fundamental. We show that the mistake we make in approximating $a^*(s, \tilde{\theta}_{s-1})$ in this way is of order $1/s$, 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 $1/s$. Since when updating 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 $1/s^2$, and hence the total error in approximating beliefs from period $t$ onwards is of order $\sum_{s=t}^{\infty} 1/s^2$, i.e., finite. Furthermore, as $t \to \infty$, the approximation error from period $t$ onwards 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 $g(\tilde{\theta}_t)$ defined above. Note, however, that the expected change 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 $\tau_t$ 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

$$z'(\tau) = g(z(\tau)), \tag{7}$$

which is a deterministic equation. Intuitively, since $z(\tau)$ is a sped-up version of $\tilde{\theta}_t$, as $\tau$ 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 Equation (7) converges to a stable point, the agent's beliefs converge to a stable point.

It is worth noting some ways in which our convergence argument can be generalized. Consistent with much of the literature, we have assumed that while the agent uses a misspecified model for interpreting observations, she applies Bayes' Rule perfectly when updating her beliefs. Such a high degree of rationality may be implausible, especially for a misspecified agent. But the above logic makes clear that the convergence of beliefs, and even the same limiting beliefs, obtain under different plausible assumptions on how the agent updates. For convergence as well as the determination of limiting beliefs, what matters is that — as in Equation (4) — the agent's belief moves in the direction of her perceived signal $\tilde{b}_{t+1}$, and it does so by an amount on the order of $1/t$. Besides Bayes' Rule with a normal-normal structure, a number of other updating processes generate such a property. For instance, in a version of adaptive expectations, the agent (perhaps heuristically) behaves as if she had point beliefs equal to the average of her observed signals. Then, any finite variance distribution for $\epsilon_t$ yields the same set of limiting beliefs. Relatedly, if we maintain the assumption of Bayesian updating, we can posit any distribution with mean zero and finite variance for the *true* noise, so long as the agent's subjective model is normal-normal.

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 $\hat{Q}(a_t, b_t)$, she observes $Q(a_t, b_t)$, and she believes that she observes $\tilde{Q}(a_t, b_t)$. For notational simplicity, our model imposes $\hat{Q} = \tilde{Q}$, but as long as $\hat{Q}$ satisfies Assumption 1 and $\tilde{Q}$ and $Q$ satisfy Assumption 2, our convergence proof remains valid for $\hat{Q} \neq \tilde{Q}$. Obviously, assuming a different objective function does in general change the optimal action, thus $g$, and consequently the set of stable limiting beliefs.

## 5.2 Using the Framework: Examples

We illustrate the use of our main result in our applications, excepting overconfidence. An application to overconfidence — albeit in a somewhat different setting — is in Heidhues et al. (2018).

*Confirmatory Bias.* While Theorem 1 implies that the agent's beliefs almost always converge, our example of confirmatory bias illustrates that the limit may not be unique. Suppose that $\Theta = \tilde{\theta}_0 = 0$, $m(0) = 0$, $m'(0) > 1$, $m'(a) > 0$ and $m''(a) < 0$ for $a > 0$, $\lim_{a\to\infty} m'(a) < 1$, and $m(-a) = -m(a)$. It is easy to see that $\tilde{b}(b_t, a_t) = b_t + m(a_t)$, so $g(\tilde{\theta}) = m(\tilde{\theta}) - \tilde{\theta}$. This implies that there is a unique $\tilde{\theta}_\infty > 0$ such that $g(\tilde{\theta}_\infty) = 0$, and therefore $C = \{-\tilde{\theta}_\infty, 0, \tilde{\theta}_\infty\}$. But 0 is not a stable fixed point of $g$, so $H = \{-\tilde{\theta}_\infty, \tilde{\theta}_\infty\}$. By Theorem 1, the agent's beliefs converge with probability one 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 of $\tilde{\theta}_\infty$ and $-\tilde{\theta}_\infty$ with probability one-half. Which one the agent's beliefs converge to depends on her early observations. If she draws sufficiently many positive observations early on, for instance, her confirmatory bias leads her to develop a positive bias.

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.

The model also implies that if $m'(0) < 1$ (with the other assumptions remaining unchanged), then despite her confirmatory bias the agent develops correct beliefs in the long run. If the confirmatory bias is sufficiently weak, therefore, it does not threaten correct long-run beliefs.

*Fertilizer.* Using the example's functional form in Equation (1), we have $\tilde{k}(a^*(\tilde{\theta}))\exp(\tilde{b}(b, a^*(\tilde{\theta}))) = k(a^*(\tilde{\theta}))\exp(b)$, so that $\tilde{b}(b, a^*(\tilde{\theta})) - b = \log(k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta})))$.[7] Using this fact in Equations (5)

[7] Based on the above, we can verify that Assumptions 1 and 2 hold. Because $\tilde{b}_a(b, a) = (k'(a)/k(a)) - (\tilde{k}'(a)/\tilde{k}(a))$, it is bounded, so Part (ii) of Assumption 2 is satisfied. Since for all $a \geq \bar{a}$, $\log(k(a)/\tilde{k}(a)) \leq 0$, setting

and (6), we obtain $g(\tilde{\theta}) = \mathbb{E}_b\left[b+\log(k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta})))|\Theta\right]-\tilde{\theta}_t = \Theta-\tilde{\theta}+\log(k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta})))$. Furthermore, the subjectively optimal action $a^*(\tilde{\theta})$ satisfies the first-order condition $\tilde{k}'(a^*(\tilde{\theta}))\exp(\tilde{\theta} + \sigma_\epsilon^2/2) = 1$. This can be rewritten as $\tilde{k}'(a^*(\tilde{\theta}))\exp(\Theta + \sigma_\epsilon^2/2 + \tilde{\theta} - \Theta) = 1$, so using $\Theta + \sigma_\epsilon^2/2 = 0$ yields $\Theta - \tilde{\theta} = \log(\tilde{k}'(a^*(\tilde{\theta})))$. Hence, we get $g(\tilde{\theta}) = \log\left[\tilde{k}'(a^*(\tilde{\theta}))k(a^*(\tilde{\theta}))/\tilde{k}(a^*(\tilde{\theta}))\right]$. 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 $\tilde{\theta} \to \infty$.

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 $\Theta$. There could also be a stable zero below $\Theta$.[8] 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, then she uses it heavily. This leads to high productivity, which she attributes in a large part to the technology rather than her usage. 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.

*Misspecified Demand Function.* Here, $\tilde{b}(b_t, a_t)$ solves the equation $b_t - 3a_t = \tilde{b}_t - a_t$, so $\tilde{b}(b_t, a_t) = b_t - 2a_t$. Using $a^*(\tilde{\theta}) = \tilde{\theta}/2$ yields $g(\tilde{\theta}) = E_{b_t}[b_t - 2a_t] - \tilde{\theta} = \Theta - 2\tilde{\theta}$, which (since $\Theta = 6$) implies that $H = \{3\}$. Hence, the agent comes to believe that the fundamental is 3, which

$\Delta = \max_{a\in[0,\bar{a}]}\log(k(a)/\tilde{k}(a))$ verifies that Part (i) of Assumption 2 holds. Clearly, our assumption on $\tilde{k}$ ensure that Assumption 1 Part (i) holds. Because $|\tilde{Q}_{aa}| = -\tilde{k}''(a)\exp(b)$ and $|\tilde{Q}_{ab}| = \tilde{k}'(a)\exp(b)$, these are integrable with respect to a normal distribution over $b$, and so Part (ii) also holds. Part (iii) holds since $(|\tilde{Q}_{ab}|/|\tilde{Q}_{aa}|) = (\tilde{k}'(a)/|\tilde{k}''(a)|)$, and hence bounded by assumption.

[8] 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 non-trivially 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 $g$.

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.

*Conservatism and Base-Rate Neglect.* We establish that under the symmetry assumption that $m(x) = -m(-x)$, conservatism and base-rate neglect lead to correct long-run beliefs for any $\Theta$. In this case, $\tilde{b}(b_t, a_t) = b_t + m(b_t - a_t)$, so $g(\tilde{\theta}) = E[b + m(b - \tilde{\theta})] - \tilde{\theta} = E[b - \tilde{\theta} + m(b - \tilde{\theta})]$. Notice that for both conservatism and base-rate neglect, the function $x + m(x)$ satisfies the same symmetry assumption as $m$, and that $b - \tilde{\theta}$ is distributed normally with mean $\Theta - \tilde{\theta}$. This implies that $g(\tilde{\theta}) > 0$ for $\tilde{\theta} < \Theta$, $g(\Theta) = 0$, and $g(\tilde{\theta}) < 0$ for $\tilde{\theta} > \Theta$. Hence, $H = \{\Theta\}$, and thus the agent's beliefs converge with probability 1 to $\Theta$. Intuitively, under both biases the agent updates in the right direction, albeit not by the right amount. So long as she does so, she must eventually end up near the correct beliefs.

*Equilibrium Feedback.* Making the same assumptions on $m$ as in the application to confirmatory bias above, we get the same formal model with a different economic interpretation. If $m$ is sufficiently steep at zero, then the economy converges either to an overly high or an overly low level of activity relative to what is justified by the fundamental. When beliefs happen to be randomly high, individuals choose high actions, and as they misinterpret the resulting observations as indicating strong fundamentals, they can develop optimistic, “irrationally exuberant” beliefs about the economy. The converse happens when beliefs are randomly low.

# 6 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.

## References


**Benjamin, Daniel J.**, "Errors in Probabilistic Reasoning and Judgmental Biases," in Douglas B. Bernheim, Stefano DellaVigna, and David I. Laibson, eds., *Handbook of Behavioral Economics*, Vol. 2, Elsevier, 2019.

**Berk, Robert H.**, "Limiting Behavior of Posterior Distributions when the Model Is Incorrect," *Annals of Mathematical Statistics*, 1966, *37* (1), 51–58.

**Bó, Ernesto Dal, Pedro Dal Bó, and Erik Eyster**, "The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects," *The Review of Economic Studies*, 2018, *85* (2), 964–998.

**Bruner, Jerome S. and Mary C. Potter**, "Inference in Visual Recognition," *Science*, 1964, *144* (3617), 424–425.

**Duflo, Esther, Michael Kremer, and Jonathan Robinson**, "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya," *American Economic Review*, 2008, *98* (2), 482–488.

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

**_ and _** , "Equilibrium in Misspecified Markov Decision Process," 2016. Working Paper, available at `http://arxiv.org/abs/1502.06901`.

**Fudenberg, Drew, Gleb Romanyuk, and Philipp Strack**, "Active Learning with a Misspecified Prior," *Theoretical Economics*, 2017, *12* (3), 1155–1189.

**Heidhues, Paul, Botond Kőszegi, and Philipp Strack**, "Unrealistic Expectations and Misguided Learning," *Econometrica*, 2018, *86* (4), 1159–1214.

**Kushner, Harold J. and G. George Yin**, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, 2003.

**Lord, Charles G., Lee Ross, and Mark R. Lepper**, "Biased Assimilation and Attitude Polarization: The Effect of Prior Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology*, 1979, *37* (11), 2098–2109.

**Nyarko, Yaw**, "Learning in Mis-Specified Models and the Possibility of Cycles," *Journal of Economic Theory*, 1991, *55* (2), 416–427.

**Rabin, Matthew and Joel Schrag**, "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 1999, *114* (1), 37–82.

**Shalizi, Cosma Rohilla**, "Dynamics of Bayesian Updating with Dependent Data and Misspecified Models," *Electronic Journal of Statistics*, 2009, *3*, 1039–1074.

# A Proofs

**Proof of Lemma 1.** We established in the text that the agent believes at the end of period $t-1$ that $b_t$ is normally distributed with mean $\tilde{\theta}_{t-1}$ and variance $h_\epsilon^{-1} + [h_0 + (t-1)h_\epsilon]^{-1}$. Now because $\tilde{Q}$ is strictly concave in $a_t$, so is the expectation $\tilde{\mathbb{E}}_{t-1,\tilde{\theta}_{t-1}}\left[\tilde{Q}(a, \tilde{b})\right]$ and thus if an optimal action exists it is unique and characterized by the first order condition

$$0 = \tilde{\mathbb{E}}_{t-1,\tilde{\theta}_{t-1}}\left[\tilde{Q}_a(a^*(t, \tilde{\theta}_{t-1}), \tilde{b})\right] . \tag{8}$$

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. As $|\tilde{Q}_{ab}|$ and $|\tilde{Q}_{aa}|$ are integrable with respect to a normal distribution over the external state by Assumption 1, we can apply the dominated convergence theorem to show that the right-hand-side of Equation 8 is differentiable in $a$ and $\theta$. We can thus apply the implicit function theorem (using the fact that $\tilde{b}_t$ is distributed according to $\tilde{\theta}_{t-1} + \sigma_t \eta$, where $\eta$ denotes the standard normal distribution $\mathcal{N}(0,1)$ and $\sigma_t^2 = \frac{1}{h_\epsilon} + \frac{1}{h_0 + (t-1)h_\epsilon}$) to establish that

$$\frac{\partial a^*(t, \tilde{\theta}_{t-1})}{\partial \tilde{\theta}_{t-1}} = -\frac{\tilde{\mathbb{E}}_{t,\tilde{\theta}_{t-1}}\left[\tilde{Q}_{ab}(a^*(t, \tilde{\theta}_{t-1}), \tilde{b})\right]}{\tilde{\mathbb{E}}_{t,\tilde{\theta}_{t-1}}\left[\tilde{Q}_{aa}(a^*(t, \tilde{\theta}_{t-1}), \tilde{b})\right]} ,$$

exists. Furthermore, by Assumption 1

$$\left| \frac{\partial a^*(t, \tilde{\theta}_{t-1})}{\partial \tilde{\theta}_{t-1}} \right| \le \frac{\tilde{\mathbb{E}}_{t,\tilde{\theta}_{t-1}}\left[\kappa \left|\tilde{Q}_{aa}(a^*(t, \tilde{\theta}_{t-1}), \tilde{b})\right|\right]}{\tilde{\mathbb{E}}_{t,\tilde{\theta}_{t-1}}\left[\left|\tilde{Q}_{aa}(a^*(t, \tilde{\theta}_{t-1}), \tilde{b})\right|\right]} = \kappa ,$$

and hence the derivative is bounded from above and below. □

In the proof of Theorem 1, we make use of the following fact.

**Lemma 2** (Convergence to the Limiting Action.)**.** *Suppose Assumption 1 holds. There exists a constant $d > 0$ such that for any $t$ and any $\tilde{\theta}_{t-1}$, we have $|a^*(t, \tilde{\theta}_{t-1}) - a^*(\tilde{\theta})| \le \frac{1}{t}\, d$.*

**Proof of Lemma 2.** Consider the optimal action as a function of the posterior variance $\sigma_t^2 = \frac{1}{h_\epsilon} + \frac{1}{h_0 + (t-1)h_\epsilon}$. As the optimal action $a^*$ is interior and the output $\tilde{Q}$ is differentiable with respect to the action, we have that

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

As $|\tilde{Q}_{ab}|$ and $|\tilde{Q}_{aa}|$ are integrable with respect to a normal distribution over the external state by Assumption 1, we can apply the dominated convergence theorem to show that the right-hand-side of Equation 8 is differentiable in $a$ and $\eta$. We can thus apply the implicit function theorem to get

$$\frac{\partial a^*(\sigma, \tilde{\theta})}{\partial \sigma} = -\frac{\mathbb{E}\left[\tilde{Q}_{ab}(a^*(\sigma, \tilde{\theta}), \tilde{\theta} + \sigma\eta)\eta \mid \eta \sim \mathcal{N}(0,1)\right]}{\mathbb{E}\left[\tilde{Q}_{aa}(a^*(\sigma, \tilde{\theta}), \tilde{\theta} + \sigma\eta) \mid \eta \sim \mathcal{N}(0,1)\right]}.$$

Using that $|\tilde{Q}_{ab}(a_t, a, b_t)|/|\tilde{Q}_{aa}(a_t, a, b_t)| \leq \kappa$ for all $a_t, b_t$, thus,

$$\begin{aligned}\left|\frac{\partial a^*(\sigma, \tilde{\theta})}{\partial \sigma}\right| &\leq \frac{\mathbb{E}\left[|\tilde{Q}_{ab}(a^*(\sigma, \tilde{\theta}), \tilde{\theta} + \sigma\eta)||\eta| \mid \eta \sim \mathcal{N}(0,1)\right]}{\mathbb{E}\left[|\tilde{Q}_{aa}(a^*(\sigma, \tilde{\theta}), \tilde{\theta} + \sigma\eta)| \mid \eta \sim \mathcal{N}(0,1)\right]} \\ &\leq \frac{\mathbb{E}\left[\kappa|\tilde{Q}_{aa}(a^*(\sigma, \tilde{\theta}), \tilde{\theta} + \sigma\eta)||\eta| \mid \eta \sim \mathcal{N}(0,1)\right]}{\mathbb{E}\left[|\tilde{Q}_{aa}(a^*(\sigma, \tilde{\theta}), \tilde{\theta} + \sigma\eta)| \mid \eta \sim \mathcal{N}(0,1)\right]} = \kappa\mathbb{E}\left[|\eta| \mid \eta \sim \mathcal{N}(0,1)\right] =: d\,.\end{aligned}$$

We thus have that

$$\begin{aligned}|a^*(\sigma_t, \theta) - a^*(\sigma_\infty, \theta)| &= \left|\int_{\sigma_\infty}^{\sigma_t} \frac{\partial a^*(z, \tilde{\theta})}{\partial z} \mathrm{d}z\right| \leq \int_{\sigma_\infty}^{\sigma_t} \left|\frac{\partial a^*(z, \tilde{\theta})}{\partial z}\right| \mathrm{d}z \\ &\leq d(\sigma_t - \sigma_\infty) = d\frac{1}{h_0 + (t-1)h_\epsilon}\,.\end{aligned}$$

□

**Proof of Theorem 1.** We begin by showing that $H$ is non-empty. It follows from Assumption 2 that for the $\Delta$ in the assumption

$$\begin{aligned}g(\Theta + 2\Delta) &= \mathbb{E}_{b_{t+1}}\ [\tilde{b}(b_{t+1}, a^*(\tilde{\theta}))|\Theta] - \Theta - 2\Delta \leq \mathbb{E}_{b_{t+1}}\ [b_{t+1} + \Delta|\Theta] - \Theta - 2\Delta \\ &= \Theta + \Delta - \Theta - 2\Delta \leq -\Delta\,.\end{aligned}$$

By the same argument $g(\Theta - 2\Delta) \geq \Delta$. As $g$ is strictly positive at $\Theta - 2\Delta$ and strictly negative at $\Theta - 2\Delta$ and continuous it crosses zero at least once from above. Since $C$ is finite, this point is a stable point according to Definition 1.

The convergence result is Theorem 2.1 in Kushner and Yin (2003), page 127 applied to dynamics given in Equation (4)

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

which (when constrained to a compact set) is a special case of the dynamics of Equation (1.2) on page 120 in Kushner and Yin (2003). We will first establish that the sequence of mean beliefs is bounded with probability one, and then apply the theorem for the case in which the dynamics are bounded with probability one and there is no constraint set. Second, we verify that the conditions for the theorem apply.

First, by Assumption 2 , $|b_t - \tilde{b}_t| \in [0, \Delta]$. Since $\tilde{\theta}_0 = \theta_0$ and $\theta_t - \tilde{\theta}_t = (1-\gamma_t)(\theta_{t-1} - \tilde{\theta}_{t-1}) + \gamma_t(b_t - \tilde{b}_t)$ for $t \geq 1$, $|\theta_t - \tilde{\theta}_t| \in [0, \Delta]$ is bounded. Hence, because $(\theta_t)_t$ converges to $\Theta$ almost surely, $\limsup_{t\to\infty} \tilde{\theta}_t \leq \Theta + \Delta$ and $\liminf_{t\to\infty} \tilde{\theta}_t \geq \Theta - \Delta$ are almost surely bounded.

Second, we now check the conditions that are necessary for the Theorem to be applicable. Below we follow the enumeration of assumptions used in Kushner and Yin:

(1.1) The theorem requires $\sum_{t=1}^{\infty} \gamma_t = \infty$ and $\lim_{t\to\infty} \gamma_t = 0$. Immediate from the definition of $\gamma_t = \frac{h_\epsilon}{(t+1)h_\epsilon + h_0}$.

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

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

By the norm inequality for the $L_2$ norm we have that

$$\begin{aligned}\sqrt{\mathbb{E}\left[|\tilde{b}_{t+1} - \tilde{\theta}_t|^2 \mid \Theta\right]} &\leq \sqrt{\mathbb{E}\left[|\tilde{b}_{t+1} - b_{t+1}|^2 \mid \Theta\right]} + \sqrt{\mathbb{E}\left[|b_{t+1} - \theta_t|^2 \mid \Theta\right]} + \sqrt{\mathbb{E}\left[|\theta_t - \tilde{\theta}_t|^2 \mid \Theta\right]} \\ &\leq 2\Delta + \sqrt{\mathbb{E}\left[|b_{t+1} - \Theta|^2 \mid \Theta\right]} + \sqrt{\mathbb{E}\left[|\theta_t - \Theta|^2 \mid \Theta\right]} \\ &= 2\Delta + h_\epsilon^{-1/2} + \sqrt{\mathbb{E}\left[|\theta_t - \Theta|^2 \mid \Theta\right]}\,.\end{aligned}$$

Where the second to last line follows as $|\theta_t - \tilde{\theta}_t| \leq \Delta$ and $|b_t - \tilde{b}_t| \leq \Delta$. The result follows as the expected $L_2$ distance of the correct belief of an outsider $\theta_t$ from the state $\Theta$ is monotone decreasing.

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

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

where $g$ is defined in Equation 5.

(A2.3) The theorem requires $g(\tilde{\theta})$ to be continuous in $\tilde{\theta}$. By Lemma 1, $a^*(\tilde{\theta})$ is differentiable and thus continuous in $\tilde{\theta}$. By Assumption 2, $\tilde{b}$ is differentiable in $a$ and thus continuous $a$. Hence, $g(\tilde{\theta}) = \mathbb{E}\left[\tilde{b}(a^*(\tilde{\theta}), b) - \tilde{\theta} \mid \Theta\right]$ is continuous in $\tilde{\theta}$.

(A2.4) The theorem requires $\sum_{t=1}^{\infty}(\gamma_t)^2 < \infty$. Immediate from the definition of $\gamma_t = \frac{h_\epsilon}{(t+1)h_\epsilon + h_0}$.

(A2.5) The theorem requires $\sum_{t=1}^{\infty} \gamma_t |\beta_t| < \infty$ *w.p.*1. Observe that

$$\beta_t = \mathbb{E}\left[\tilde{b}(a^*(t,\tilde{\theta}), b) - \tilde{b}(a^*(\tilde{\theta}), b) \mid \Theta\right] .$$

To bound $\beta_t$, we use that by Assumption 2 the derivative of the perceived signal with respect to the action $\frac{\partial \tilde{b}}{\partial a}$ exists and is bounded from above and below. Let $d_1 > 0$ be a bound such that $\left|\frac{\partial \tilde{b}}{\partial a}\right| \leq d_1$. Hence,

$$|\beta_t| \leq \mathbb{E}\left[|\tilde{b}(a^*(t,\tilde{\theta}), b) - \tilde{b}(a^*(\tilde{\theta}), b)| \mid \Theta\right] \leq d_1 \, \mathbb{E}\left[|a^*(t,\tilde{\theta}) - a^*(\tilde{\theta})|\right] .$$

By Lemma 2 $|a^*(t,\tilde{\theta}) - a^*(\tilde{\theta})| < (1/t)d_2$ for some $d_2 > 0$. Consequently,

$$\sum_{t=1}^{\infty} \gamma_t |\beta_t| \leq d_1 d_2 \sum_{t=1}^{\infty} \frac{h_\epsilon}{(t+1)h_\epsilon + h_0} \times \frac{1}{t} < \infty.$$

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

$$f(\tilde{\theta}) = -\int_{x}^{\tilde{\theta}} g(z)\mathrm{d}z \,,$$

for some $x$. As in our case $C$ is finite, $f$ trivially satisfies all conditions of the theorem. □