Unrealistic Expectations and Misguided Learning*

Paul Heidhues
DICE, University of Düsseldorf

Botond Kőszegi
Central European University

Philipp Strack
UC Berkeley

April 21, 2017

Abstract

We explore the learning process and behavior of an individual with unrealistically high expectations ("overconfidence") when outcomes also depend on an external fundamental that affects the optimal action. Moving beyond existing results in the literature, we show that the agent's beliefs regarding the fundamental converge under weak conditions. Furthermore, we identify a broad class of situations in which "learning" about the fundamental is self-defeating: it leads the individual systematically away from the correct belief and toward lower performance. Due to his overconfidence, the agent—even if initially correct—becomes too pessimistic about the fundamental. As he adjusts his behavior in response, he lowers outcomes and hence becomes even more pessimistic about the fundamental, perpetuating the misdirected learning. The greater is the loss from choosing a suboptimal action, the *further* the agent's action ends up from optimal. We partially characterize environments in which self-defeating learning occurs, and show that the decisionmaker learns to take the optimal action if and only if a specific *non*-identifiability condition is satisfied. In contrast to an overconfident agent, an underconfident agent's misdirected learning is self-limiting and therefore not very harmful. We argue that the decision situations in question are common in economic settings, including delegation, organizational, effort, and public-policy choices.

*We are grateful to Itzhak Gilboa, three anonymous referees, Roland Bénabou, Daniel Benjamin, Volker Nocke, Demian Pouzo, Francesco Squintani, as well as seminar and conference audiences for insightful comments and feedback. Heidhues and Kőszegi thank the European Research Council for financial support under Starting Grant #313341. Part of this research was carried out while Heidhues and Kőszegi visited the WZB, which they thank for its hospitality. Mats Köster and Kinga Marczell provided excellent research assistance.

1 Introduction

Large literatures in psychology and economics suggest that in many situations, individuals have unrealistically positive beliefs about their traits or prospects, and researchers have began to investigate the nature of this "overconfidence" and study its implications for economic interactions.[1] One important question concerning individuals with overconfident or otherwise biased beliefs is how they update these beliefs when information comes in. Indeed, classical results identify conditions under which learning leads to correct beliefs (e.g., Savage, 1954, Chapter 3), and more recent research explores ways in which a biased learning process can lead to overconfident beliefs (e.g., Gervais and Odean, 2001, Chiang, Hirshleifer, Qian and Sherman, 2011, Jehiel, 2016).

In this paper, we investigate how overconfident individuals update not their overconfident beliefs, but their beliefs about other decision-relevant variables. Moving beyond existing results in the literature, we show that beliefs converge under weak conditions. Furthermore, we identify a broad and economically important class of situations in which an overconfident person's inferences are self-defeating: they lead him systematically away from the correct belief and toward lower performance. For example, if a team member is overly full of himself and is hence disappointed by his team's performance, he concludes that his teammates are less talented or lazier than he thought. He responds by increasing his control of the team, lowering team performance. He misinterprets this low performance as reflecting even more negatively on his teammates, perpetuating the misdirected learning further. Perversely, the greater is the loss from choosing a suboptimal action, the *further* the agent's action ends up from optimal. We partially characterize environments in which self-defeating learning occurs, and show that the decisionmaker's long-run behavior is optimal if and only if a specific *non*-identifiability condition is satisfied. In contrast to an overconfident agent, an underconfident agent's misdirected learning is self-limiting and therefore not very harmful.

We present our framework in Section 2.1. In each period $t \in \{1, 2, 3, \dots\}$, the agent produces observable output $q_t = Q(e_t, a, \phi) + \epsilon_t$, which depends on his action e_t, his ability or other output-relevant parameter a, an unknown fundamental ϕ, and a noise term ϵ_t. We assume that Q is increasing in a and ϕ, and that the optimal action is increasing in ϕ.[2] The noise terms ϵ_t are inde-

[1] We use the term "overconfidence" to mean broadly any unrealistic beliefs, especially about ability or other important personal characteristics, that lead a person to expect good outcomes. The same expression is often used more specifically to denote overly narrow confidence intervals (Moore and Healy, 2008). We review evidence for and theoretical work on overconfidence in Section 7.

[2] So long as these effects are monotonic, the above directional assumptions are just normalizations.

pendently and identically distributed mean-zero random variables with a log-concave distribution that has full support on $\mathbb{R}$. The agent uses Bayes' rule to update beliefs, and chooses the myopically optimal action in every period. Crucially, he is overconfident: while his true ability is A, he believes with certainty that it is $\tilde{a} > A$. Finally, for most of the paper we assume that the optimal action depends in weakly opposite ways on ability and the fundamental—that is, it is weakly decreasing in ability. This assumption is sufficient for generating self-defeating learning.

In Section 2.2, we argue that beyond team production, this reduced-form model captures a number of economically important situations in which individuals may have unrealistic expectations. A principal may not know how intrinsically motivated his employees are, and hence what level of control or explicit incentives maximizes performance. A person may not know how nice his partner or friend is, and hence how deferentially or assertively he should act to elicit the best outcomes from the relationship. A student or employee may not know the return to effort, and hence how much effort is optimal. And a policymaker may not know the scale of underlying problems in the economy, and hence what policy leads to the best outcomes.

Adapting Esponda and Pouzo's (2016a) concept of Berk-Nash equilibrium to our setting, in Section 2.3 we define a stable belief according to the intuitive consistency property that—when taking the action that he perceives as optimal given his false belief—the average output the agent expects coincides with the average output he produces. Because at a stable belief the agent has no reason to question his beliefs, it is where his beliefs can be expected to converge. Motivated by this observation, in Sections 3, 4, and 6 we study the properties of stable beliefs, assuming that beliefs converge there; and in Section 5, we take up convergence. We assume that there is a unique stable belief, and identify sufficient conditions on the primitives for this to be the case (Proposition 1). This assumption simplifies our statements regarding the properties of limiting beliefs, and is crucial for our convergence proof.

In Section 3, we study the key properties of the agent's learning process and limiting beliefs, summarize a variety of casual evidence for our central mechanism, and discuss economic implications. We establish that—as in the case of the overconfident team member above—the agent's learning process is self-defeating: if his initial action is not too far from optimal, then the opportunity to change his action in response to what he learns leads to more incorrect beliefs, and more suboptimal behavior, than if he could not change his action (Proposition 2). Furthermore, limiting beliefs satisfy a surprising and perverse comparative static: the more important it is for the agent to take the right action—that is, the greater is the loss from a suboptimal action—the *further* his beliefs end up from the truth, and the further his behavior ends up from optimal (Proposition 3). Intuitively, when choosing a suboptimal action is more harmful, the agent hurts himself more through his misguided learning. To develop a consistent theory of his observations, therefore, he must become more pessimistic about the fundamental.

We also consider what happens if—similarly to a dissatisfied team member deciding whether to replace his teammate—the agent can choose between the above task and an outside option. Consistent with the literature on overconfidence, the agent might be too prone to enter into and initially persist in the task. In contrast to received wisdom, however, our model predicts that the agent's growing pessimism about the fundamental may induce him to *exit* the task too easily, and by implication to jump too much between tasks. This prediction is consistent with the observation that many documented effects of overconfidence in economic settings, such as the pursuit of mergers and innovations by overconfident CEOs or the creation of new businesses by overconfident entrepreneurs, pertain to *new* directions.

In Section 4, we ask what happens when our sufficient condition for self-defeating learning—that the optimal action depend in opposite ways on ability and the fundamental—is not satisfied, and argue that self-defeating learning still occurs if and only if the optimal action depends sufficiently less on ability than on the fundamental. As a conceptually interesting case, we show that long-run behavior is always optimal if and only if Q has the form $V(e, S(a, \phi))$—that is, ability and the fundamental are *not* separately identifiable (Proposition 4). This conclusion contrasts with the lesson from classical learning settings that non-identifiability hinders efficient learning. Intuitively, because ability and the fundamental do not have independent effects on output, the agent's misinference about the fundamental can fully compensate his overconfidence, and hence in the long run overconfidence does not adversely affect him. If his team's output depends on his effort and the team's total ability, for instance, the agent correctly infers total ability, and chooses effort optimally.

In Section 5, we show that if Q satisfies some regularity conditions, then the agent's beliefs converge to the stable belief with probability one (Theorem 1). Furthermore, under the additional assumption that Q is linear in ϕ, beliefs converge to the stable belief even when actions are dynamically optimal (Theorem 2). To prove convergence, 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. Worse, beliefs constitute a function-valued process, whose transitions are, due to the endogeneity of the agent's action, driven by shocks that are neither independently nor identically distributed. Little is known in general about the asymptotic behavior of the posterior distribution in such infinite-dimensional models when the observations are not i.i.d., even when the model is correctly specified (e.g., Ghosal and van der Vaart, 2007, pages 192-3). To find a handle on the problem, we employ a method that to our knowledge has not been used in the literature on learning with misspecified models. We focus on the agent's extremal beliefs: what levels of the fundamental does he in the limit conclusively rule out? Given the structure of our problem, this puts bounds on his long-run actions, which restrict his extremal beliefs further. Using this contraction logic, we show that the agent rules out everything but a single point.

Knowing that beliefs converge with probability one allows us to make another economically relevant point: given that in the long run the agent correctly predicts average output, his beliefs pass a specification test based on comparing the empirical distribution of noise terms he estimates to the true distribution (Proposition 5). In this sense, the agent's overconfidence is—even with infinite data—stable.

In Section 6, we analyze variants of our framework using a simple output function of the form $Q(e, a, \phi) = a + \phi - L(e - \phi)$ for a symmetric loss function L. We consider a setting in which the agent is initially uncertain about a, and observes—in addition to output—a noisy signal of his relative contribution to output, $a - (\phi - L(e_t - \phi))$. It being highly subjective, however, the agent observes his own contribution with a positive bias. As a result, he develops overconfidence about a, and his limiting belief about ϕ is identical to that in our basic model (Proposition 6).

We also study underconfident agents (for whom $\tilde{a} < A$), and identify an interesting asymmetry: while an overconfident agent's limiting utility loss from misguided learning can be an arbitrarily large multiple of his overconfidence $\tilde{a} - A$, an underconfident agent's limiting utility loss is bounded by his underconfidence $A - \tilde{a}$ (Proposition 7). To understand the intuition, consider an underconfident agent who starts off with the correct mean belief about the fundamental. Upon observing higher levels of output than he expected, the agent concludes that the fundamental is better than he thought, and revises his action. The resulting utility loss, however, leads him to reassess the optimistic revision of his belief, bringing his beliefs back toward the true fundamental. Hence, his misinference regarding the fundamental—which with overconfidence was self-reinforcing—is now self-correcting.

In Section 7, we relate our paper to the two big literatures it connects, that on the implications of overconfidence and that on learning with misspecified models. From a methodological perspective, our model is a special case of Esponda and Pouzo's (2016a) framework for games when players have misspecified models. Because we have an individual-decisionmaking problem with a specific structure, we can derive novel and economically important results that are not possible in the general framework. In particular, to our knowledge our paper is the first to study the implications of overconfidence for inferences about other decision-relevant variables, and how these inferences interact with behavior. We are also unaware of other papers characterizing when self-defeating learning does versus does not occur in an individual-decisionmaking context.

In Section 8, we conclude by discussing some potential applications of our framework for multi-person situations.

2 Learning Environment

In this section, we introduce our basic framework, outline possible economic applications of it, and perform a few preliminary steps of analysis.

2.1 Setup

In each period $t \in \{1, 2, 3, \dots\}$, the agent produces observable output $q_t \in \mathbb{R}$ according to $Q(e_t, a, \Phi) + \epsilon_t$, where $e_t \in (\underline{e}, \overline{e})$ is his action, $a \in \mathbb{R}$ is his unchanging ability, $\Phi \in (\underline{\phi}, \overline{\phi})$ is an unobservable unchanging fundamental, and ϵ_t is random noise. Throughout the paper, we make the following basic assumptions. First, the state Φ is drawn from a continuous prior distribution $\pi_0 : (\underline{\phi}, \overline{\phi}) \to \mathbb{R}_{>0}$ with bounded positive density everywhere on $(\underline{\phi}, \overline{\phi})$, and we suppose that the agent has a correct prior belief π_0. Second, the ϵ_t are i.i.d. continuously distributed mean-zero random variables. Denoting their cumulative distribution function by F and their strictly positive density by $f : \mathbb{R} \to \mathbb{R}_{>0}$, we impose a version of log concavity: the second derivative of $\log f$ is strictly negative and bounded from below.[3] Third, output satisfies some regularity properties and normal-

[3] We think of strict bounded log concavity as an economically weak restriction: it guarantees that an increase in output increases the agent's belief about the fundamental in the sense of the monotone likelihood ratio property, and that any signal is non-trivially informative. Examples of distributions that have full support on $\mathbb{R}$ and satisfy the

izations: Q is twice continuously differentiable, with (i) $Q_{ee} < 0$ and $Q_e(\underline{e}, a, \phi) > 0 > Q_e(\overline{e}, a, \phi)$ for all a, ϕ; (ii) $Q_a, Q_\phi > 0$, and (iii) $Q_{e\phi} > 0$. Part (i) guarantees that there is always a unique myopically optimal action. Part (ii) implies that output is increasing in ability and the fundamental, and Part (iii) implies that the optimal action is increasing in the fundamental. So long as the effects implied by Parts (ii) and (iii) are monotonic, our directional assumptions on them are just normalizations, and do not affect the logic and message of our results. Indeed, if any of these derivatives was negative, we could change variables to reverse the orientation of a, ϕ, or e, and obtain a model in which the same derivative is positive.

We assume for most of the paper that the agent chooses his action myopically in each period, aiming to maximize that period's expected output. This assumption is irrelevant for our analysis of the properties of limiting (point) beliefs in Sections 3, 4, and 6. We do use the assumption for our main convergence result, although under stronger conditions we establish convergence also when the agent optimizes dynamically.

Crucially, we posit that the agent is overoptimistic about his ability: while his true ability is A, he believes with certainty that it is $\tilde{a} > A$. Let $\Delta = |A - \tilde{a}|$ denote the degree of the agent's overconfidence. Given his inflated self-assessment, the agent updates his beliefs about the fundamental in a Bayesian way. To guarantee that the agent can always find a fundamental that is consistent with the average output he produces, we assume that for all e, there is a $\tilde{\phi} \in (\underline{\phi}, \overline{\phi})$ such that $Q(e, A, \Phi) = Q(e, \tilde{a}, \tilde{\phi})$. Note that because $Q_\phi > 0$, this $\tilde{\phi}$ is unique.

We specify the agent's belief about ability as degenerate for two main reasons. Technically, the assumption generates a simple and tractable model of persistent overconfidence, allowing us to focus on our research question of learning about *other* variables. More importantly, we view an assumption of overconfident beliefs that are not updated downwards as broadly realistic. Quite directly, such an assumption is consistent with the view of many psychologists that individuals are extremely reluctant to revise self-views downwards (e.g., Baumeister, Smart and Boden, 1996). More generally, such an assumption can be thought of as a stand-in for forces explored in the psychology and economics literatures (but not explicitly modeled here) that lead individuals to maintain unrealistically positive beliefs. To show that our model is consistent with this perspective, in Section 6.1 we allow for the agent to be uncertain about a, and show that a biased learning process

assumption include the normal and logistic distributions.

leads to a setting essentially equivalent to ours.[4]

To complete the description of our model, we state a sufficient condition for self-defeating misguided learning to occur:

Assumption 1 (Sufficient Condition for Self-Defeating Learning)**.** $\text{sgn}(Q_{ea}) \neq \text{sgn}(Q_{e\phi})$.

Note that with the normalization $Q_{e\phi} > 0$, Assumption 1 is equivalent to $Q_{ea} \leq 0$. This assumption plays a central role in our paper: in Section 3, we explore the implications of misguided learning under Assumption 1, and in Section 4, we study what happens when Assumption 1 is not satisfied. Economically, we think of Assumption 1 as allowing for two possibilities:

(i) The optimal action is almost insensitive to ability, at least relative to how sensitive it is to the fundamental ($Q_{ea} \approx 0 \ll Q_{e\phi}$).[5] This is likely to be the case in many applications in which the agent is looking to fine-tune a decision—such as the design of a public policy or organizational incentive system—to circumstances he views as largely external to his overconfident beliefs.

(ii) In small-scale economic settings, however, the optimal action may be non-trivially sensitive to ability. Then, Assumption 1 requires that the optimal action depend in opposite ways on ability and the fundamental. This assumption naturally holds for delegation, as the optimal extent of delegation depends in opposite ways on the decisionmaker's ability and his teammate's ability. In contrast, the assumption does not hold if output always depends on ability and the fundamental in similar ways, such as when it depends on $a + \phi$.

The following two parametric examples are useful to keep in mind when developing our results.

Example 1: Loss-Function Specification. The output function has the form

$$Q(e, a, \phi) = a + \phi - L(e - \phi), \tag{1}$$

where $\underline{e} = \underline{\phi} = -\infty$, $\overline{e} = \overline{\phi} = \infty$, $a \in \mathbb{R}$, and L is a symmetric loss function with $L(0) = 0$ and $|L'(x)| < k < 1$ for all x. Economically, this specification captures a situation in which the agent wants to adjust his action to some underlying state of the world. Researchers have used similar, loss-function-based, specifications in Crawford and Sobel (1982) and the large literature on cheap

[4] The model in Section 6.1 replaces a misspecified point belief about a with misspecified learning about a, and hence is also not a fully rational model. But we view the misspecified nature of the agent's learning as a feature, not a bug: any model in which the agent is correctly specified and keeps learning until he has the correct belief about his ability contradicts observed widespread overconfidence among individuals who have had plenty of opportunity to learn about themselves.

[5] As we argue in more detail in Section 4, if Q_{ea} is positive but small relative to $Q_{e\phi}$, our qualitative results regarding misdirected learning survive unchanged.

talk, expert advice, and organizations (e.g., Alonso, Dessein and Matouschek, 2008) following it, as well as in Morris and Shin (2002) and the subsequent literature on coordination.

Example 2: Effort Specification. The output function has the form

$$Q(e, a, \phi) = (a + e)\phi - c(e), \tag{2}$$

where $\underline{e} = \underline{\phi} = 0$, $\overline{e} = \overline{\phi} = \infty$, $a > 0$, and c is a strictly convex cost function with $c(0) = c'(0) = 0$, and $\lim_{e\to\infty} c'(e) = \infty$. In this example, the action e is effort, and the fundamental ϕ determines the return to effort.

Consistent with our motivating phenomenon of overconfidence, in describing and discussing our model and results we interpret a as ability. But more broadly, a could stand for any variable that leads the agent to be steadfastly and unrealistically optimistic about his prospects. For example, he may have overly positive views about his country or organization, or—as in the case of policymakers below—he may have overly optimistic beliefs about the technology generating output.

2.2 Applications

In this section, we argue that several economically important settings fit our model of Section 2.1. In each of these settings, other individuals are also involved, and for a full account it is necessary to model their behavior as well. Nevertheless, to focus on a single agent's inferences and behavior, we abstract from the decisions of others, and model their effect only in reduced form.

Application 1: Delegation. The decisionmaker is working in a team with another agent, and must decide how much of the work to delegate. The expected output of the team is $Q(e, a, \phi) = a\, u(e) + \phi\, v(e)$, where $e \in (0, 1)$ is the proportion of the job the decisionmaker delegates, ϕ is the teammate's ability, and $u(e), v(e) > 0, u'(e), u''(e), v''(e) < 0, v'(e) > 0$. Then, the higher is the decisionmaker's ability and the lower is the teammate's ability, the lower is the optimal extent of delegation.

Application 2: Control in Organizations. A principal is deciding on the incentive system to use for an agent who chooses two kinds of effort, overt effort x^o (e.g., writing reports) and discretionary effort x^d (e.g., helping others in the organization). The principal can incentivize overt effort, for instance through monitoring (e.g., requiring and reading reports) or explicit incentives written on an objective signal of overt effort. For simplicity, we assume that the principal chooses x^o directly. Consistent with the literature on multitasking starting with Holmström and Milgrom (1991), we also assume that discretionary effort is a decreasing function of overt effort. In addition, we suppose that discretionary effort is an increasing function of the agent's intrinsic motivation ϕ. Writing discretionary effort as $x^d = x^d(x^o, \phi)$, the principal's profit is $R(a, x^o, x^d(x^o, \phi))$, where a is the quality of the organization, the principal's ability, or other factor affecting overall productivity. Supposing that the optimal overt effort is decreasing in intrinsic motivation,[6] this model reduces to our setting with $Q(e, a, \phi) = R(a, -e, x^d(-e, \phi))$.[7]

Application 3: Assertiveness versus Deference. The decisionmaker is in a personal relationship—such as partnership or friendship—with another individual. The output q is how "nicely" the other person is acting, including how willing she is to comply with the agent's requests, how much of the unpleasant joint work she does, etc. The fundamental ϕ is how nice the other person is, and e is how deferentially the agent acts toward her. This action could range from being extremely deferential (high e) to being extremely aggressive or even violent (low e). Finally, a stands for the agent's talent or attractiveness. Output is determined according to the loss-function specification of Example 1: the partner tends to act more nicely if she is a nicer person, if the agent is more talented or attractive, and if the agent's behavior is more attuned to the partner's niceness.

Of course, the deference or aggressiveness of an agent, or the niceness of his partner's behavior, are not typical outcomes studied by economists. But these are clearly observable—with noise—to individuals, and manifestations—at least of extreme choices—might be observable to researchers as well.[8]

Application 4: Work and Return to Effort. The agent is an employee or student who must decide how hard to work at his job or school. He periodically observes the output of his efforts, such as promotions, grades, or other rewards, but he does not know the return to effort. Output is given by the effort specification in Example 2.

Application 5: Public policy. A policymaker aims to maximize the performance of some aspect of the economy, q, which depends on his policy choice e, a fundamental ϕ, his ability or his party's or country's potential a, and noise according to the loss-function specification in Example 1. In particular, output q could be the well-being of the population in relation to drug-related crime,

[6] One simple functional form capturing our discussion is $R(a, x^o, x^d(x^o, \phi)) = a + x^o + x^d(x^o, \phi)$, where the discretionary effort function satisfies $\frac{\partial x^d(x^o,\phi))}{\partial x^o} < 0$, $\frac{\partial x^d(x^o,\phi))}{\partial \phi} > 0$, and $\frac{\partial^2 x^d(x^o,\phi))}{\partial x^o \partial \phi} < 0$.

[7] An alternative interpretation of the same framework is that x^o is the agent's "mechanical" input into the organization, x^d is his "creative" input, and ϕ is his ability. In this interpretation, creative input depends on creative effort—which is a substitute to overt effort—and ability.

[8] For instance, Card and Dahl (2011) study the role of emotions in family violence using police records.

ϕ the underlying condition of the population with respect to drug use, and e the degree of drug liberalization. The extent of restrictions must be optimally aligned with underlying conditions to minimize drug-related crime. In a completely different example, q could represent the overall performance of the economy, e the country's openness toward other countries in trade, exchange of ideas, and movement of people, and ϕ the optimal degree of integration with the world economy.[9]

One may argue that since politicians are heavily scrutinized and receive a lot of feedback, they should learn about themselves and not be overconfident. Yet by extensively documenting and studying overconfidence in another high-flying and heavily scrutinized group, top CEO's, Malmendier and Tate (2005) and the literature following it have shown that feedback does not necessarily eliminate overconfidence. Hence, while we are unaware of direct compelling evidence that politicians are overconfident, it is plausible that many are. In addition, a policymaker may have unrealistic expectations not only because of his trust in himself, but also because of his unrealistic beliefs about policy tools. He may, for instance, have false beliefs about how much of an increase in economic growth or reduction in crime can be achieved with the right policy.

2.3 Preliminaries

Let $e^*(\phi)$ denote the optimal action when the fundamental is ϕ. We define the *surprise function* as

$$\Gamma(\phi) = Q(e^*(\phi), A, \Phi) - Q(e^*(\phi), \tilde{a}, \phi). \tag{3}$$

If the agent believes the fundamental to be ϕ, then he takes the action $e^*(\phi)$, and therefore expects average output to be $Q(e^*(\phi), \tilde{a}, \phi)$. In reality, he obtains an average output of $Q(e^*(\phi), A, \Phi)$, so that $\Gamma(\phi)$ represents the average surprise he experiences. Applying Esponda and Pouzo's (2016a) Berk-Nash equilibrium—their solution concept for games when players may have misspecified models—to our setting, we define:[10]

Definition 1. A stable belief is a Dirac measure on a state ϕ_∞ at which the agent is not surprised by average output, i.e. $\Gamma(\phi_\infty) = 0$.

Intuitively, a stable belief is the only type of point belief that the agent finds no reason to abandon. If the agent holds a stable belief about the fundamental, then he takes an action such that he

[9] This example was suggested to us by Francesco Squintani.

[10] Lemma 4 in the appendix shows formally that a stable belief and the corresponding optimal action constitute a Berk-Nash equilibrium, and that every pure-strategy Berk-Nash equilibrium is of that form.

produces on average exactly as much as he expects, confirming his belief. In contrast, if the agent holds any other point belief about the fundamental, then he takes an action such that he obtains a non-zero average surprise, accumulating evidence that the fundamental is something else. As a result, one would expect the agent's beliefs to converge to a stable belief.

Motivated by these observations, in the rest of the paper we perform two distinct types of analysis. In Sections 3, 4, and 6, we study properties of the agent's stable beliefs, assuming that his beliefs converge to a stable belief. And in Section 5, we confirm in the central case of our model that—under somewhat stronger technical assumptions—beliefs indeed converge to a stable belief.

For our analysis, we assume that there is a unique stable belief. Although our model's mechanism and insights regarding the properties of stable beliefs hold more generally, this assumption simplifies many of our statements. In addition, the assumption is crucial for our convergence proof. We identify sufficient conditions for a unique stable belief:

Proposition 1. *There is a unique stable belief if:*

(i) Q_a is bounded, Q_ϕ is bounded away from zero, and overconfidence ($\tilde{a} - A$) is sufficiently small; or

(ii) Q takes the form in Example 1; or

(iii) Q takes the form in Example 2 and $c'''(e) \geq 0$.

3 Main Mechanism and Economic Implications

In this section, we lay out the main forces in our model, and discuss economic implications. For this purpose, we suppose that Assumption 1 holds. Although we state our results for general Q, throughout we use the loss-function specification of Example 1 for illustration. For these illustrations, we normalize $A = \Phi = 0$, and suppose that the prior on ϕ is symmetric and has mean equal to the true fundamental.

3.1 Self-Defeating Learning

Example 1—Fixed Action. As a benchmark case, we suppose for a moment that e_t is exogenously given and constant over time at level $e = e^*(\Phi) = 0$. Then, average output converges to $Q(e, A, \Phi) = 0$. The agent, in contrast, believes that if the state is ϕ, then average output is $Q(e, \tilde{a}, \phi) =$

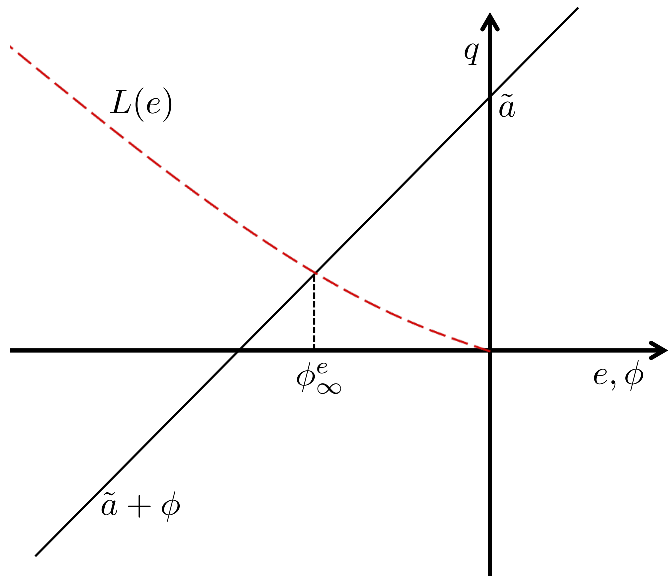

Figure 1: Limiting Beliefs with Fixed Action

$\tilde{a} + \phi - L(e - \phi) = \tilde{a} + \phi - L(\phi)$. To explain an average output of zero, therefore, he comes to believe that the fundamental is ϕ^e_∞ satisfying $\tilde{a} + \phi^e_\infty - L(\phi^e_\infty) = 0$.

Figure 1 illustrates. The agent's belief about the fundamental (ϕ) as well as his action (e) are on the horizontal axis, and output is on the vertical axis. The agent's limiting belief is given by the intersection of the loss function through the origin and the line $\tilde{a} + \phi$: at this point, the loss from taking the (in his mind) suboptimal action exactly explains the average output of zero.

Clearly, $\phi^e_\infty < \Phi$. Intuitively, the agent is surprised by the low average output he observes, and concludes that the fundamental is worse than he thought. This tendency to attribute failures to external factors provides a formalization of part of the self-serving attributional bias documented by Miller and Ross (1975) and the literature following it. Indeed, one account of the bias is that individuals have high expectations for outcomes, and update when outcomes fall short (Tetlock and Levi, 1982, Campbell and Sedikides, 1999). But while the agent's inference about the fundamental is misguided—it takes him away from his correct prior mean—in the current setup with a fixed action it is harmless or potentially even beneficial. For instance, because the agent now correctly predicts average output, he makes the correct choice when deciding whether to choose this task over an outside option with a given level of utility.[11]

[11] As a simple implication, if the agent's high belief about ability is due to ego utility as in Kőszegi (2006), then misdirected learning allows him to have his cake and eat it too: he can maintain the pleasure from believing that his

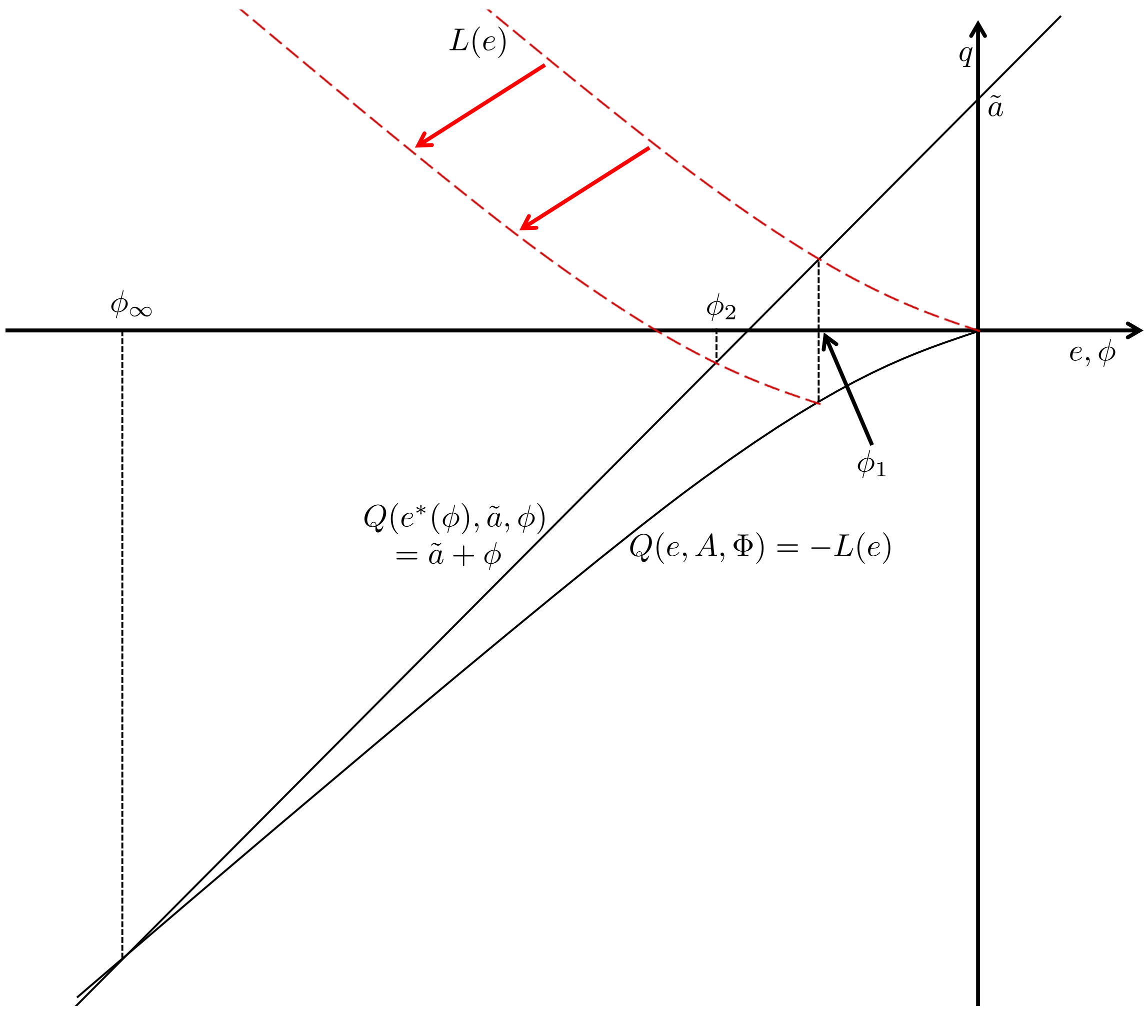

Figure 2: Learning with Endogenous Action

Example 1—Endogenous Action. Now suppose that the agent chooses his action optimally given his beliefs. We illustrate the learning dynamics in Figure 2, again putting the agent's belief (ϕ) as well as his action (e) on the horizontal axis and output on the vertical axis. The "average output possibilities curve" $Q(e, A, \Phi) = -L(e)$ represents true average output as a function of the agent's action, and the "perceived average achievable output line" $Q(e^*(\phi), \tilde{a}, \phi) = \tilde{a} + \phi$ represents the average output the agent believes is reachable as a function of the fundamental ϕ. Since $e^*(\phi) = \phi$, the agent's stable belief assigns probability one to the intersection of the two curves. It is apparent from the figure that the limiting belief is further than in the case of exogenously fixed actions above.

ability is high, while not suffering any losses associated with incorrect beliefs.

Hence, learning is self-defeating in a basic sense: the fact that the agent can optimize given his learning makes his learning even more inaccurate, with the corresponding bad effect on his actions. In particular, the agent's limiting per-period utility loss relative to the optimal action can be an arbitrarily large multiple of his overconfidence Δ.

How does the agent drive himself into such a highly suboptimal situation? We give a heuristic argument, assuming that the agent updates his action increasingly rarely, and pretending that his average output and beliefs reach their limit in finite time after each update. Suppose that for a long time, the agent chooses the optimal action given his initial beliefs, $e = 0$, and ends up with an average output of zero. Then, the belief ϕ_1 he develops about the fundamental is the same as in the case of exogenously fixed actions: it is given by the intersection of the loss function starting at $(0, 0)$ (the higher red dashed curve) and the perceived average achievable output line.

Now suppose that the agent updates his action, and for an even longer time keeps taking the optimal action given his new belief, $e = \phi_1$. Then, he ends up with an average output of $-L(\phi_1)$. To identify the ϕ_2 consistent with these observations, we draw a copy of the loss function starting from his current action-average output location, and find the intersection of this curve with the perceived average achievable output line. This is his new belief ϕ_2. Continuing with this logic gives a sense of the agent's learning dynamics, and illustrates why he ends up at ϕ_∞.

We briefly discuss how our prediction of self-defeating learning can manifest itself in the applications outlined in Section 2.2. In the context of manager-subordinate relationships (an example of Application 1), Manzoni and Barsoux (1998, 2002) describe a common problem—the "set-up-to-fail syndrome"—that is due to a dynamic similar to our mechanism. When the subordinate performs below expectations, the manager reacts by increasing control and assigning less inspiring tasks. This undermines the employee's motivation, lowering performance and eliciting further negative reactions from the manager.

In the context of organizations (Application 2), the notion of dysfunctional vicious circles has been described in sociology (March and Simon, 1958, Crozier, 1964, Masuch, 1985). Management is dissatisfied with the performance of the organization, and attempts to improve things with increased monitoring, control, or other bureaucratic measures. These measures backfire, for instance because they alienate employees. Management reacts by tightening bureaucratic measures further.[12]

[12] Masuch (1985) also argues that one reason for such vicious cycles is that people have unrealistically high expectations regarding how well things will work. The cycle ends when management runs out of control tools, or, as

In the context of personal relationships (Application 3), our model provides one possible mechanism linking unrealistically high self-views to aggressive behavior. Although it had been hypothesized that aggression is associated with low self-esteem, Baumeister et al. (1996) and the subsequent literature reviewed by Lambe, Hamilton-Giachritsis, Garner and Walker (forthcoming) argue that aggressive individuals tend to have high self-esteem that is not based on compelling facts. Our model says that in this situation, the agent comes to feel that his partner does not treat him with the respect he deserves from a nice partner, and reacts by becoming more assertive. As a result of this misguided reaction, he receives even less respect from his partner, leading him to react even more strongly. The relationship deteriorates, and potentially ends up abusive and/or violent.[13]

Although we have not found discussions that are reminiscent of our mechanism in our other applications, self-defeating interpretations of worse-than-expected outcomes also seem plausible in these settings. In the context of work or study effort (Application 4), when a person with unrealistic expectations does not obtain the rewards he expects, he concludes that the return to talent and effort is lower than he thought, and lowers his effort in response. But because effort *is* important, this lowers his output more than he expects, leading him to become even more pessimistic about the return to effort. As a case in point, our mechanism may contribute to the widespread underestimation of the returns to college education documented by Bleemer and Zafar (2015), but more research is necessary to determine the extent and nature of its role.

In an interesting contrast to our model, existing work on overconfidence, such as Bénabou and Tirole (2002), Gervais, Heaton and Odean (2011) and de la Rosa (2011), has emphasized that if ability and effort are complements, then overconfidence can benefit a person by leading him to exert higher effort. Our model says that if ability and effort have separable effects, or even if they are complements, but the complementarity is low, then exactly the opposite is the case.

In the context of the war on drugs (an example of Application 5), a policymaker may interpret drug problems as indicating that he should crack down on drugs, only to make the problem worse and the reaction harsher. In the context of nationalism, a policymaker (or citizens) may react to disappointing economic outcomes by concluding that globalization does not hold as much promise

described for instance by Argyris (1982), when management settles into a suboptimal set of practices.

[13] The mechanism that Baumeister and coauthors describe verbally is somewhat different from ours. They hypothesize that violence serves as a way for a person to protect his unrealistically high self-views when these are threatened by an outsider. While this is consistent with our mechanism if we interpret a "threat" as treating the agent with less respect than he believes he deserves, the relationship of the two mechanisms warrants further research.

as they have hoped. This leads the country to adopt more nationalistic and protectionist policies, exacerbating the problem and hardening the conclusion that globalization is a failure.

General Q. We now state the self-defeating nature of learning formally and for general Q:

Proposition 2. *Suppose Assumption 1 holds. If the agent's action is exogenously fixed at $e \geq e^*(\Phi)$, then his beliefs converge to a Dirac measure on ϕ^e_∞, where $\phi_\infty < \phi^e_\infty < \Phi$.*

Proposition 2 implies that if the agent starts off with an action at or above the optimal action, then the opportunity to change his action in response to his inferences leads to more incorrect long-run beliefs than if he could not change his action. By continuity, the same is the case if his initial action is below but close to the optimal action. Of course, the self-defeating nature of learning also implies that if the initial action is not too far from the optimal action, then the opportunity to update his action lowers the agent's long-run average utility.

We conclude this section by illustrating in the loss-function specification that the directional assumptions we have made on the effect of the fundamental on output and the optimal action ($Q_\phi, Q_{e\phi} > 0$) are indeed irrelevant for the logic of our results. First, suppose that the production function is of the form $Q(e, a, \phi) = a - \phi - L(e - \phi)$, so that an increase in the fundamental lowers expected output. As Figure 3 makes clear, this leaves the logic—including that the agent's learning is self-defeating and can leave his beliefs arbitrarily far from the truth relative to Δ—completely unchanged: the average output possibilities curve is still $-L(e)$, and the perceived average achievable output line is now $\tilde{a} - \phi$, so the mirror image of our previous analysis applies. Consider also the possibility that the optimal action depends negatively on the state ($Q_{e\phi} < 0$): $Q(e, a, \phi) = a + \phi - L(e + \phi)$. With the trivial change of variables $\phi' = -\phi$, this is equivalent to the case above, so once again the same analysis and the same insights result.

3.2 The Importance of Being Right

Example 1. Because it provides a surprising and potentially testable prediction of our model, we consider the comparative statics of limiting beliefs with respect to the loss function. An increase in the loss function increases the agent's incentive to act in accordance with the fundamental. His action responds in an unfortunate way: by ending up *further* from the fundamental. If the loss function shifts up, then the curve $-L(e_t)$ in Figure 2 shifts down, and the limiting belief ϕ_∞ therefore moves to the left. Intuitively, a steeper loss function means that the agent hurts himself

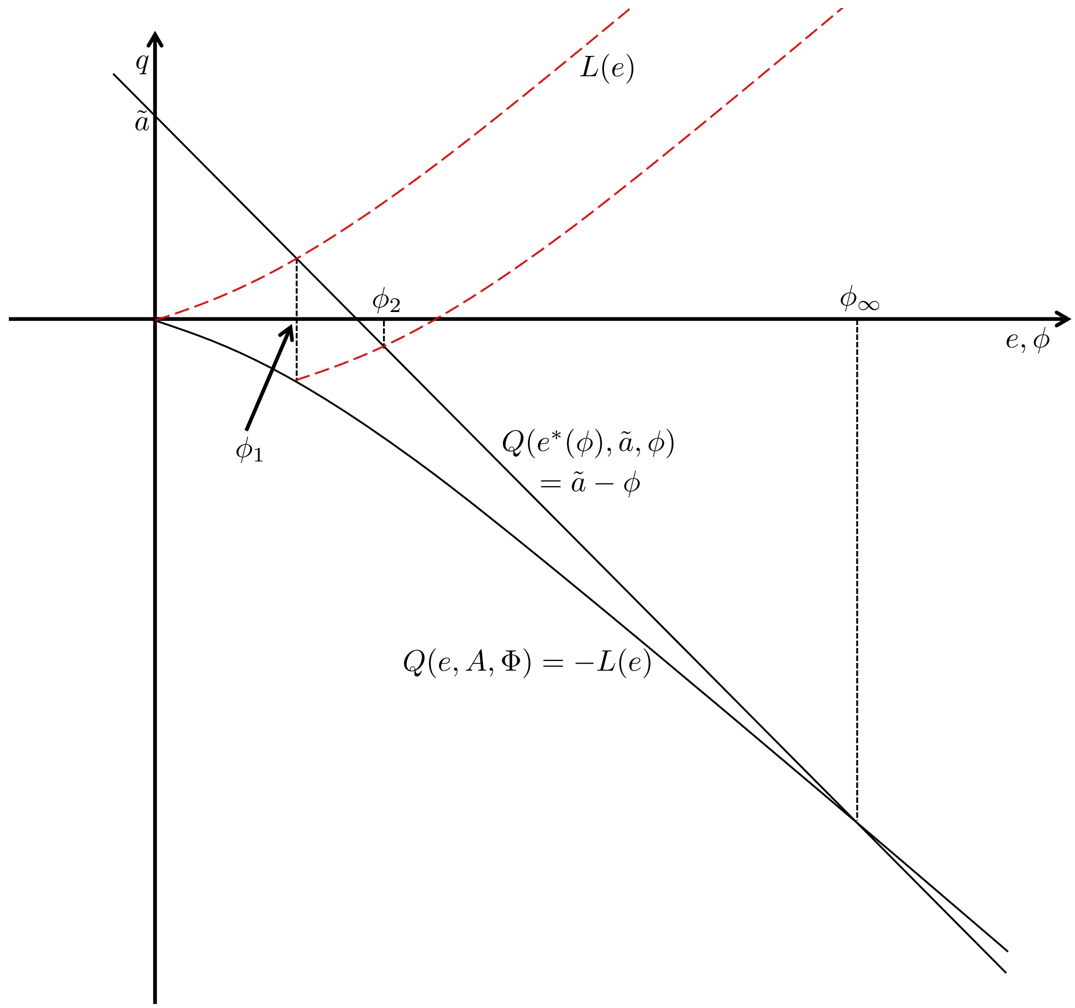

Figure 3: Learning when the Fundamental Lowers Output

more through his misinferences. To develop a consistent theory of his observations, therefore, he must become more pessimistic about the world.

General Q. To generalize the above result to arbitrary production functions, we define

$$\begin{aligned} R(a,\phi) &= Q(e^*(a,\phi),a,\phi) \\ L(e,a,\phi) &= Q(e^*(a,\phi),a,\phi) - Q(e,a,\phi), \end{aligned} \tag{4}$$

and observe that $Q(e,a,\phi) = R(a,\phi) - L(e,a,\phi)$. Intuitively, $R(a,\phi)$ is the average achievable output given a and ϕ, and $L(e,a,\phi)$ is the loss relative to the achievable output due to choosing a suboptimal action. We compare two technologies Q_1, Q_2 with corresponding $e_1^*, e_2^*, R_1, R_2, L_1, L_2$.

Proposition 3. *Suppose Assumption 1 holds, $e_1^*(a,\phi) = e_2^*(a,\phi)$ and $R_1(a,\phi) = R_2(a,\phi)$ for all a,ϕ, and $L_1(e,a,\phi) < L_2(e,a,\phi)$ for all $a,\phi,e \neq e^*(a,\phi)$. Then, the agent's stable belief is further below the true fundamental Φ under technology Q_2 than under technology Q_1.*

The key step in the proof of Proposition 3 is that an increase in the loss function decreases the surprise function. An increase in the loss function means that the agent hurts himself more through his misinference-induced suboptimal behavior, and this must mean that he is more negatively surprised by his average output. And a downward shift in the surprise function lowers the agent's stable belief.

3.3 Outside Options

In the above model, we have assumed that the agent participates in the task in every period regardless of his beliefs. It is natural to consider an environment in which he has an outside option, such as another task he could perform. In the manager-employee setting, for instance, a manager can keep working with the current employee, or he can fire the employee and reopen or discontinue the position. Let the utility of the outside option be $\underline{u}$. We suppose that the agent correctly understands $\underline{u}$, and discuss the implications of relaxing this assumption below. For simplicity, we consider the loss-function specification of Example 1. Interesting new issues arise if the prior on ϕ is not too dispersed and $\underline{u}$ is in the intermediate range, so that the agent starts off with the task but eventually quits.

There are then two cases. If $\underline{u} > A$, then the agent should not be in this task in the first place, so both his entry into it and his initial persistence in it are suboptimal. The prediction that overconfident individuals—overestimating their ability to perform well—are more likely to enter into and persist in ability-sensitive tasks is consistent with common intuition and evidence from both psychology and economics. Overconfidence is often invoked as one explanation for why many new businesses fail in the first few years, and indeed Landier and Thesmar (2009) find that entrepreneurs of small startups have overconfident expectations about future growth. Similarly, several studies suggest that CEO overconfidence is associated with a higher likelihood of pursuing risky actions, such as making acquisitions (Malmendier and Tate, 2008) and undertaking innovation (Galasso and Simcoe, 2011, Hirshleifer, Low and Teoh, 2012). And persisting in tasks (often for too long) is regarded as one of the main characteristics of overconfidence (McFarlin, Baumeister and Blascovich, 1984, for example).

In contrast, if $\underline{u} < A$, then the agent should be in the task, so overconfidence generates suboptimal exit: ironically, the agent stops performing the task *because* he overestimates his ability to do well in it. Intuitively, he is more prone to exit than a realistic agent because his negative inferences about the fundamental eventually negate his overconfidence.

Importantly, an overconfident agent is likely to overestimate not only output in the current task, but—in as much as it depends in part on ability—also the outside option $\underline{u}$. Such overestimation exacerbates the tendency to exit the task, and can generate interesting dynamics when there are multiple types of alternative tasks for the agent to choose. The logic of our model suggests that the agent first seeks out another ability-sensitive task in which he believes a different fundamental determines outcomes, and then successively jumps from one such task to the next. And once he runs out of these tasks, he chooses a less ability-sensitive task and sticks with it.[14]

The prediction that overconfidence leads individuals to eventually quit superior tasks, to jump too much between tasks, and to eventually prefer less ability-sensitive tasks, contrasts with the typical view on the implications of overconfidence. While we have not found direct evidence for this prediction, it is consistent with the observation that many documented effects of overconfidence in economic settings, such as the tendency of overconfident CEO's to undertake mergers or innovations, pertain to the pursuit of new directions rather than to persistence in old directions.

Our result that an overconfident agent jumps between tasks suggests one possible reason for the persistence—though not for the emergence—of overconfidence. Since the agent often abandons tasks and prefers tasks for which previous learning does not apply, he can keep updating mostly about external circumstances, slowing down learning about his ability.

4 When Learning Is Not Detrimental

In Section 3, we have shown that self-defeating learning—whereby the agent's response to his misguided inferences leads to even more misguided inferences and even more suboptimal behavior—always occurs under Assumption 1. We now discuss what happens when Assumption 1 is not

[14] A related point to ours is made in a bargaining context by Bénabou and Tirole (2009), who show that an individual may leave a productive partnership in part because he overestimates his outside option. Our model predicts that the agent may well exit even if he correctly understands the outside option. As a result, he may exit even if no ability-sensitive alternative tasks are available.

satisfied, allowing us to partially characterize the conditions that facilitate self-defeating learning.

Suppose, therefore, that $Q_{ea} > 0$, so that the optimal action is strictly increasing in ability. To understand the key issue, suppose also that the agent starts off with a belief that is concentrated around the true fundamental. Then, because he is overconfident and $Q_{ea} > 0$, his initial action is too high. As in the rest of our analysis, the surprisingly low average output he observes leads him to revise his belief about the fundamental downwards—and, as a consequence, to choose lower actions. Because his initial action was too high, this adjustment is in the right direction. We then have two cases. It is possible that in the limit misdirected learning increases output, so that self-defeating learning does not occur. It is, however, also possible that misdirected learning lowers the action below optimal, at which point the logic by which further learning occurs is analogous to that in Section 3. Then, self-defeating learning may occur: any further negative inference about ϕ leads the decisionmaker to choose lower actions, lowering output and reinforcing his negative inference.

As a conceptually interesting question, as well as to say more about when each of the above two cases obtains, we ask when a misspecified agent's long-run behavior is optimal. Call the action that is optimal given the unique stable belief the *stable action.*

Proposition 4. *The following are equivalent:*

I. For any A, $\tilde{a}$, and Φ, the agent's stable action is optimal (i.e., maximizes true expected output).

II. The agent's stable action is identical to that with an output function $Q(e, a, \phi) = V(e, S(a, \phi))$, where $V_S, S_a, S_\phi > 0$, and for any $a, \phi, \tilde{a}$, there is a unique ϕ' such that $S(a, \phi) = S(\tilde{a}, \phi')$.

Proposition 4 says that agents with wrong beliefs about ability behave optimally in the long run if and only if long-run behavior can be described by an output function that depends only on a summary statistic S of ability and the fundamental, and not independently on the two variables. If this is the case, then the agent can in the limit correctly deduce S from average output, so he correctly predicts how changes in his action affect output. As a result, he chooses the optimal action. Our proof establishes that this kind of production function is not only sufficient, but also necessary for learning to be optimal in the limit.

An interesting aspect of Proposition 4 is that the agent is able to find the optimal action exactly when the problem is *not* identifiable—when his observations of output do not allow him to separately learn a and ϕ. This beneficial role of non-identifiability is in direct contrast with what one might expect based on the statistical learning literature, where non-identifiability is *defined* as a property of the environment that hinders learning. Yet it is exactly the non-identifiability of the problem that allows the overconfident agent to choose his long-run action well: because ability and the fundamental do not have independent effects on output, the agent's misinference about the fundamental can fully compensate his overconfidence regarding ability, and hence overconfidence does not adversely affect him.

In the team production setting, for instance, suppose that the agent—instead of making delegation decisions—chooses his effort level, and output depends on effort and the total ability of the team (i.e., output takes the form $V(e, a + \phi)$). Then, although the agent still underestimates his teammates, he is able to deduce the team's total ability $a + \phi$ from output. As a result, he chooses the optimal action.

Notice that statement II of Proposition 4 implies that controlling for their effect on output, the optimal action is equally sensitive to ability and the fundamental ($e^*_a(a, \phi)/Q_a(e^*(a, \phi), a, \phi) = e^*_\phi(a, \phi)/Q_\phi(e^*(a, \phi), a, \phi)$). This insight indicates that if $Q_{ea} > 0$, then changes in the agent's beliefs about the fundamental eventually induce actions that are significantly lower than optimal—and hence self-defeating learning occurs—if the optimal action is sufficiently more sensitive to the fundamental than to ability. And adding our insights from Section 3, we conclude that self-defeating learning occurs if the optimal action either (i) depends sufficiently less on ability than on the fundamental; or (ii) depends in opposite ways on ability and the fundamental.

5 Convergence

In this section, we establish conditions under which the agent's beliefs converge to the stable belief. We also argue that when the agent's beliefs converge, his overconfidence is stable—that is, he will not realize he is wrong—in a specific sense.

5.1 Convergence with Myopic Actions

To establish convergence, we maintain the same assumptions as in Section 2.1, but impose stronger conditions on some of the derivatives of Q:[15]

[15] Example 1 satisfies Assumption 2 so long as L'' is bounded. Example 2 does not satisfy Assumption 2 as stated, but it does so under economically minor modifications: that $\overline{e} < \infty$, $\overline{\phi} < \infty$, and a is positive, bounded, and bounded away from zero.

Assumption 2. (i) $|Q_e| < \kappa_e$; (ii) $Q_a \leq \overline{\kappa}_a$, $0 < \underline{\kappa}_\phi \leq Q_\phi \leq \overline{\kappa}_\phi$, and (iii) $|Q_{\phi\phi}| \leq \overline{\kappa}_{\phi\phi}$.

The various bounds in Assumption 2 guarantee that the agent's inferences from output and his reactions to these inferences are always of comparable size.[16]

Our main result in this section is:

Theorem 1. *Suppose Assumptions 1 and 2 hold. Then, the agent's beliefs almost surely converge in distribution to the unique stable belief.*

The interdependence between actions and beliefs—that the agent's action depends on his belief, and the bias in his inferences in turn depends on his action—creates several difficulties in our convergence proof. To start, 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. Even one central necessary component of the convergence of beliefs, the concentration of beliefs, requires some properties of the path of actions. And indeed, convergence does not hold in some other models with endogenous actions, such as those of Nyarko (1991) and Fudenberg, Romanyuk and Strack (forthcoming), and the convergence result of Esponda and Pouzo (2016a, Theorem 3) applies only for priors close to the stable belief and for actions that are only close to optimal. To make matters especially difficult, beliefs constitute a function-valued process whose transitions are driven by shocks that are neither independently nor identically distributed. Little is known in general about the asymptotic behavior of the posterior distribution in such infinite-dimensional models when the observations are not i.i.d., even when the model is correctly specified (e.g., Ghosal and van der Vaart, 2007, pages 192-3).

To find a handle on the problem, we employ a method that to our knowledge has not been used in the literature on learning with misspecified models. We focus on extremal beliefs: what levels of the fundamental does the agent in the limit conclusively rule out? Given the structure of our problem, this puts bounds on his long-run actions, which restrict his extremal beliefs further. Using this contraction argument, we show that the agent rules out everything but the root of Γ.

We explain the detailed logic of our proof in six steps.

[16] Specifically, the lower bound on Q_ϕ ensures that the agent always makes a non-trivial inference from an increase in output. The upper bound on Q_ϕ bounds how much the agent learns from the output of a single period. Similarly, the bounds on Q_e and Q_a ensure that changing the action or ability has a limited effect on output. The condition on $Q_{\phi\phi}$ helps us to bound the second derivative of the subjective posterior log-likelihood.

Step 1. We show that the change in beliefs can in the long-run be approximated well by the expected change. To argue this, we use that the log-likelihood function is the average of the log-likelihoods of the realized outputs. This is the technically most demanding step in the proof, as—due to the endogeneity of actions—the log-likelihood is an average of non-identical and non-independent random variables. We adapt existing versions of the law of large numbers to the types of non-i.i.d. random variables our framework generates.

Step 2. We establish that if the agent is on average positively surprised by output for belief ϕ, then by Step 1 the derivative of his subjective log-likelihood goes to infinity almost surely at ϕ. An analogous statement holds for negative surprises.

Step 3. If the agent has realistic beliefs about ability ($\tilde{a} = A$), then—no matter his action—his average surprise is positive for $\phi < \Phi$, and negative for $\phi > \Phi$. By Step 2, his beliefs converge to Φ almost surely.

Step 4. Now we turn to the overconfident agent. We define $\underline{\phi}_\infty$ as the supremum of fundamentals such that in the long run, the agent almost surely convinces himself that the true fundamental is above $\underline{\phi}_\infty$. We define $\overline{\phi}_\infty$ analogously. We show that the beliefs of an overconfident agent can be bounded—in the sense of the monotone likelihood ratio property—relative to the beliefs of a realistic agent. Using that the beliefs of a realistic agent converge to Φ, this means that $\underline{\phi}_\infty$ and $\overline{\phi}_\infty$ exist. And because the agent is overconfident, his beliefs about the fundamental must in the long run be below the truth: $\underline{\phi}_\infty \leq \overline{\phi}_\infty \leq \Phi$. The goal for the rest of the proof is to show that $\underline{\phi}_\infty = \phi_\infty = \overline{\phi}_\infty$.

Step 5. We show that the above bounds on long-run beliefs also bound long-run actions. In particular, we know that the optimal action is increasing in the fundamental. Therefore, in the long run the agent's action must be on (or arbitrarily close to) the interval $[e^*(\underline{\phi}_\infty), e^*(\overline{\phi}_\infty)]$.

Step 6. Now we show by contradiction that $\underline{\phi}_\infty \geq \phi_\infty$. Supposing that $\underline{\phi}_\infty < \phi_\infty$, we establish that the agent's average surprise at $\underline{\phi}_\infty$ is positive in the long run. To see this, note that because $\underline{\phi}_\infty < \phi_\infty$, the average surprise is positive for the action $e^*(\underline{\phi}_\infty)$. Furthermore, since the agent overestimates ability and underestimates the fundamental (and $Q_{ea} \leq 0, Q_{e\phi} > 0$), he underestimates Q_e. This implies that the agent's average surprise is positive for any action near or above $e^*(\underline{\phi}_\infty)$, that is, for any long-run action. By Step 2, in the long run the agent convinces himself that the fundamental is above, and bounded away from, $\underline{\phi}_\infty$, contradicting the definition of $\underline{\phi}_\infty$.

An analogous argument establishes that $\overline{\phi}_\infty \leq \phi_\infty$.

5.2 Convergence with Dynamically Optimal Actions when Q is Linear in ϕ

In Section 5.1, we have established convergence by assuming that the agent takes the myopically optimal action. In the current section, we consider dynamically optimal actions. We assume that the agent has discount factor $\delta \in (0,1)$, and in each period chooses a current action and a (history-contingent) strategy for future actions to maximize discounted expected output. Because e_t affects how much the agent learns about the fundamental, the myopically optimal action is in general not dynamically optimal, making it difficult to bound actions based on beliefs. For example, the agent will take an action associated with a low payoff this period if that action reveals a lot of information about optimal future actions. Nevertheless, we establish convergence in a special case of our model.

Assumption 3. Suppose (i) $Q(e,a,\phi) = \phi H(e,a) + G(e,a)$; (ii) the second derivative of $\log f$ is bounded away from zero; (iii) the prior distribution π_0 of Φ satisfies $-\infty < \underline{\kappa}_\pi \leq (\partial^2/\partial\phi^2)\log\pi_0(\phi) \leq \overline{\kappa}_\pi < \infty$ as well as $\underline{\phi} \geq 0$ and $\pi_0(\underline{\phi}) = 0$; and (iii) e_t is chosen from a bounded interval.

The substantive new assumption is (i), which says that average output is linear in ϕ. The other assumptions are regularity conditions that add to Assumption 2.[17]

Theorem 2. *Suppose Assumptions 1, 2, and 3 hold and the agent chooses dynamically optimal actions. Then, the agent's beliefs almost surely converge in distribution to the unique stable belief.*

To understand the key new idea, notice that for any e_t, the agent believes that

$$\phi + \frac{\epsilon_t}{H(e_t,\tilde{a})} = \frac{q_t - G(e_t,\tilde{a})}{H(e_t,\tilde{a})}.$$

Hence, the agent believes that the observations $(q_t - G(e_t,\tilde{a}))/H(e_t,\tilde{a})$ are independently distributed signals with mean ϕ, albeit with a variance that depends on e_t. By the logic underlying the law of large numbers, the agent's subjective beliefs therefore concentrate in the long run for any sequence of actions. This implies that the subjective expected benefit from learning about the fundamental Φ vanishes, so the agent must eventually choose actions that are close to myopically optimal. As a result, the logic of Theorem 1 applies.

[17] The output function of Example 2 is linear in the fundamental and thus the example satisfies Assumption 3 so long as e is bounded. Example 1 does not satisfy Assumption 3.

5.3 Stability of Overconfidence

A central assumption of our paper is that the agent has an unrealistically high point belief about his ability. In this section, we ask whether his observations might lead him to conclude that something about his beliefs is awry, which might lead him to reconsider his belief about ability.[18] We show that if beliefs and actions converge, then there is a strong sense in which the agent's overconfidence can be considered stable—even after he has observed infinite data.

More precisely, we construct a specification test for the agent's beliefs in the following way. Suppose that the agent's beliefs converge to the stable belief. Given these limiting beliefs, the agent looks back and extracts the noise realizations he thinks have generated his observations: $\tilde{\epsilon}_t = q_t - Q(e_t, \tilde{a}, \phi_\infty)$. Now the agent takes the infinite subsample $\{\tilde{\epsilon}_{t_1}, \tilde{\epsilon}_{t_2}, \dots\}$ of the $\tilde{\epsilon}_t$, where $t_1, t_2, \dots$ are ex-ante specified. Let $\hat{F}_i(x) = |\{i' \leq i | \tilde{\epsilon}_{t_{i'}} \leq x\}|/i$ be the empirical frequency of observations below x in the first i elements of his subsample. The agent expects this empirical cumulative distribution to match the true cumulative distribution of ϵ_t in the long run. It does:

Proposition 5. *Suppose that beliefs converge in distribution to the unique stable belief. Then, for any sequence t_i and any x, $\lim_{i \to \infty} \hat{F}_i(x) = F(x)$.*

Intuitively, in the long run the agent settles on beliefs that lead him to predict average output accurately, so that he also extracts the noise terms accurately. Hence, he cannot be surprised about the long-run distribution of noise terms.

It is worth noting the role of the convergence of actions—as distinct from the convergence of beliefs—in our result. Suppose, for example, that the agent is forced to take each of the two actions e_1 and $e_2 > e_1$ infinitely many times at pre-specified dates, but at vanishing rates, so that his beliefs still converge to ϕ_∞. Since $\tilde{a} > A$ and $\phi_\infty < \Phi$, $Q_{e\phi} > 0$ and $Q_{ea} \leq 0$ imply that $Q_e(e, A, \Phi) > Q_e(e, \tilde{a}, \phi_\infty)$. The agent therefore underestimates the difference in average output with e_2 versus e_1, and hence Proposition 5 fails in a strong way: the empirical mean of $\tilde{\epsilon}_t$ must be non-zero either at e_1 or at e_2. This example illustrates that the informational environment makes it possible for the agent to discover his overconfidence—but he takes actions under which he does not. In particular, his action does not vary enough for him to reject his belief about ability.

Of course, a person may use a more sophisticated specification test that hinges on the speed

[18] See Gagnon-Bartsch, Rabin and Schwartzstein (2017) for a more detailed exploration of when a person with a misspecified model may discover that he is wrong.

of convergence of his beliefs. Unfortunately, we are unable to determine whether the agent passes such a test, because our convergence proofs cannot say anything about the speed of convergence, not even in the case of a realistic agent. We are skeptical, however, that a typical economic agent would perform specification tests of such sophistication.[19]

6 Extensions and Modifications

In this section, we discuss some economically relevant variants of our model. For simplicity, we restrict attention to the loss-function specification of Example 1 throughout the section. In addition, as in Sections 3 and 4, we assume that beliefs converge to the Dirac measure on the root of the surprise function, and study only properties of limiting beliefs.

6.1 Biased Learning about Ability

Throughout the paper, we have assumed that the agent holds point beliefs about ability that are too high, thinking of this assumption as a stand-in for forces that generate overconfidence. We now show through a simple example that the mechanisms we have identified are consistent with a setting in which overconfidence arises endogenously through biased learning. Although the precise bias we assume is different, one can think of this exercise as integrating Gervais and Odean's (2001) model of "learning to be overconfident" with our model of misdirected learning.

Suppose that the agent has continuously distributed prior beliefs about (a, ϕ) with full support on the plane. In each period, he observes output $q_t = a+\phi-L(e_t-\phi)+\epsilon_t$ as previously. In addition, he observes a noisy measure of his relative contribution to output, $r_t = a-(\phi-L(e_t-\phi)+\epsilon_t)+\epsilon'_t+\Delta'$, where the ϵ'_t have mean zero. However, the agent perceives his relative contribution with a bias: while he believes that $\Delta' = 0$, in reality $\Delta' = 2\Delta$.

The assumption that individuals perceive their own contribution to performance in a biased way is supported by a variety of evidence in psychology. For instance, Ross and Sicoly (1979) find that when married couples are asked about the portion of various household chores they perform, their answers typically add up to more than 100 percent. Babcock and Loewenstein (1997) provide

[19] Similarly, if noise is not additively separable from Q, then the agent's belief may fail a variant of the distribution-based test in Proposition 5. Even this conclusion, however, hinges on the agent being certain about the distribution of noise. If he is not certain, then the long-run distribution of output may lead him to make inferences about the noise distribution rather than to conclude that his belief about ability is incorrect.

evidence that parties in a negotiation interpret the same information differently and in a self-serving way. Summarizing the literature, Bénabou and Tirole (2016) explain that selective attention to and recall of information, asymmetric interpretation of information, and asymmetric updating can all contribute to individuals' biased self-evaluations. Our model captures these possibilities in one simple reduced-form way.

Logically, it is possible that the agent evaluates not only r_t, but also q_t in a biased way. But in most situations motivating our analysis, a person's contribution to output is more subjective than output itself, and is hence more vulnerable to distortion. Bénabou and Tirole (2009) make a similar distinction.

Because the agent now observes both q_t and r_t, we redefine his surprise function for this two-dimensional case. If the agent believes that his ability is a and the fundamental is ϕ, then he takes the action ϕ. We therefore write

$$\Gamma_q(\phi, a) = [A + \Phi - L(\phi - \Phi)] - [a + \phi]$$
$$\Gamma_r(\phi, a) = [A - (\Phi - L(\phi - \Phi)) + \Delta'] - [a - \phi]$$

The first line is the agent's surprise function for output, which follows the same logic as in our basic model. The second line is the agent's surprise function for his relative contribution. Due to his bias, the average relative contribution he perceives is $A - (\Phi - L(\phi - \Phi)) + \Delta'$. But because he is not aware of his bias, he expects his average relative contribution to be $a - \phi$. Setting both surprises to zero, adding, and rearranging gives that the stable belief about ability must be $\tilde{a}_\infty = A + \Delta'/2 = A + \Delta$—exactly the point belief that we imposed exogenously in our basic model. Plugging this into the unchanged surprise function for output, we get that the stable belief about the fundamental must also be the same as in our basic model:

Proposition 6. *There is a unique stable belief $\tilde{a}_\infty, \phi_\infty$, where $\tilde{a}_\infty = A + \Delta = \tilde{a}$, and ϕ_∞ is the unique root of Γ defined in Equation* (3)*.*

Intuitively, the agent believes that the average of total output and his relative contribution, $\sum_t (q_t + r_t)/2$, provides an unbiased estimate of his ability. This estimator, however, converges to $A + \Delta$, so that he develops overconfident beliefs over time. Given that he becomes overconfident about ability, he is misled about the fundamental in an identical way as in our basic model.

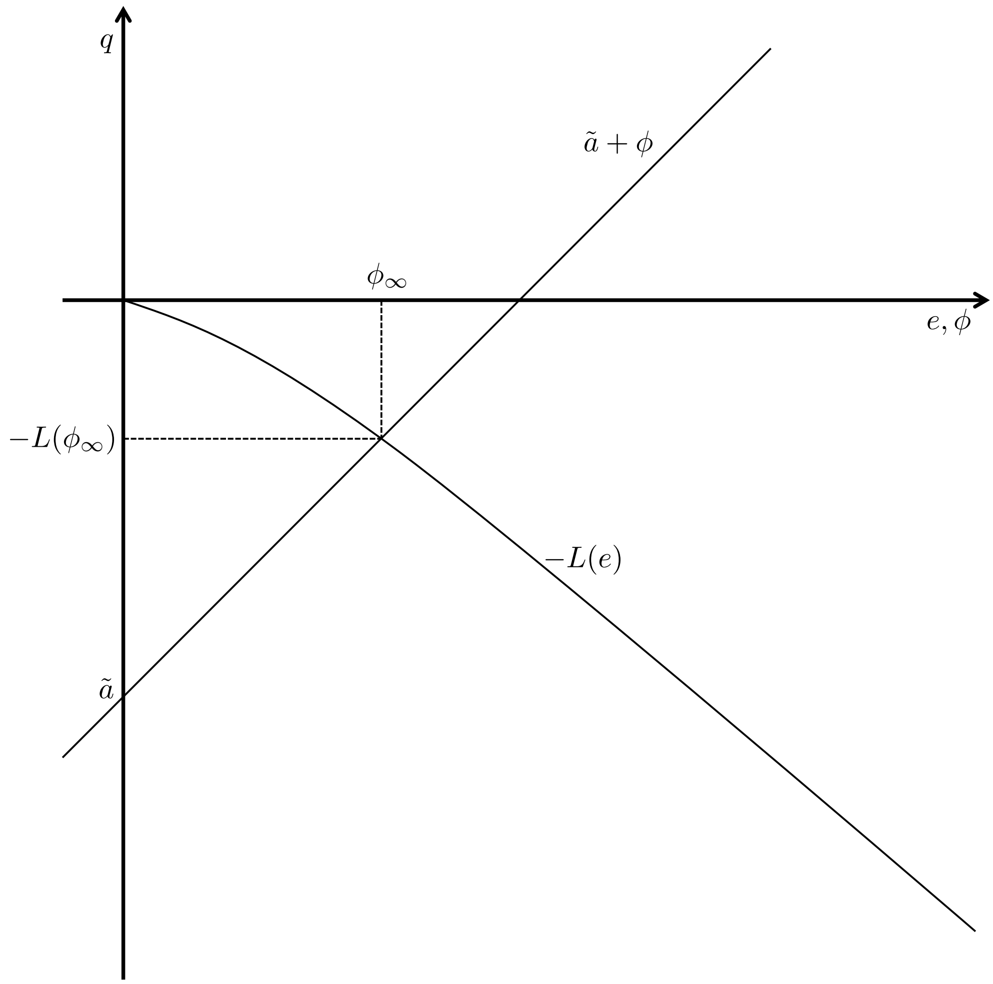

Figure 4: Limiting Belief and Loss with Underconfidence

6.2 Underconfidence

While many or most people are overconfident, some are "underconfident"—they have unrealistically *low* beliefs about themselves. We briefly discuss the implications of underconfidence for misguided learning. In Figure 4, we redraw the relevant parts of Figure 2 for underconfident beliefs ($\tilde{a} < A$), again normalizing A and Φ to 0. As for overconfident agents, any possible stable belief is given by the intersection of the average output possibilities curve $-L(e)$ and the perceived average achievable output line $\tilde{a} + \phi$. If $\tilde{a} < A$, the two curves intersect to the right of the true fundamental: since the agent is pessimistic about his own ability, he becomes overly optimistic about the fundamental.

Furthermore, it is apparent from the figure that in the limit the agent's loss from underconfidence is bounded by Δ. Formally:

Proposition 7. *Suppose Q takes the loss-function form in Example 1, and $\tilde{a} < A$. Then, there is a unique stable belief ϕ_∞, which satisfies $0 < \phi_\infty - \Phi < \Delta$ and $L(\phi_\infty - \Phi) < \Delta$.*

These results contrast sharply with those in the overconfident case, where the limiting belief is always more than Δ away from the true fundamental ($\Phi - \phi_\infty > \Delta$), and the associated loss can be an arbitrarily large multiple of Δ. To understand the intuition, consider again an agent who has a symmetric prior with mean equal to the true fundamental. Due to his underconfidence, he is then likely to observe some better performances than he expects. As a result, he concludes that the fundamental is better than he thought, and revises his action. The resulting utility loss, however, leads him to reassess the optimistic revision of his belief, bringing his beliefs back toward the true fundamental. In this sense, the agent's misinference regarding the fundamental is self-correcting—in contrast to the logic in the case of overconfidence, where the misinference is self-reinforcing. Moreover, because a utility loss of Δ or more cannot be explained by a combination of underconfidence in the amount of Δ and an unrealistically positive belief about the fundamental (which increases expected output), any consistent belief must generate a utility loss of less than Δ.

7 Related Literature

Our theory connects two literatures, that on overconfidence and that on learning with misspecified models. While we discuss other more specific differences below, to our knowledge our paper is the first one to study the implications of overconfidence for inferences about other, decision-relevant exogenous variables. More recently, Le Yaouanq and Hestermann (2016) study the same question, focusing on the issue of persistence, which we cover only briefly in Section 3.3. We are also unaware of previous research characterizing when self-defeating learning does versus does not occur in an individual-decisionmaking context.

7.1 Overconfidence

Our paper studies the implications of unrealistic expectations regarding a variable for learning about another variable. In many applications, the most plausible source of such unrealistic expectations is overconfidence—the topic of an extensive literature in economics and psychology.

A plethora of classical evidence in psychology as well as economics suggests that on average people have unrealistically positive views of their traits and prospects (e.g., Weinstein, 1980, Svenson, 1981, Camerer and Lovallo, 1999). Recently, Benoît and Dubra (2011) have argued that much of this evidence is also consistent with Bayesian updating and correct priors, and thus does not conclusively demonstrate overconfidence. In response, a series of careful experimental tests have documented overconfidence in the laboratory in a way that is immune to the Benoît-Dubra critique (Burks, Carpenter, Goette and Rustichini, 2013, Charness, Rustichini and van de Ven, 2014, Benoît, Dubra and Moore, 2015). In addition, there is empirical evidence that consumers are overoptimistic regarding future self-control (Shui and Ausubel, 2004, DellaVigna and Malmendier, 2006, for instance), that truck drivers persistently overestimate future productivity (Hoffman and Burks, 2013), that genetically predisposed individuals underestimate their likelihood of having Huntington's disease (Oster, Shoulson and Dorsey, 2013), that unemployed individuals overestimate their likelihood of finding a job (Spinnewijn, 2012), and that some CEOs are overoptimistic regarding the future performance of their firms (Malmendier and Tate, 2005). Moreover, in all of these domains the expressed or measured overconfidence predicts individual choice behavior. For example, CEOs' overconfidence predicts the likelihood of acquiring other firms (Malmendier and Tate, 2008), of using internal rather than external financing (Malmendier and Tate, 2005), of using short-term debt (Graham, Harvey and Puri, 2013), of engaging in financial misreporting (Schrand and Zechman, 2012), and of engaging in innovative activity (Hirshleifer et al., 2012). While all of these papers look at the relationship between overconfidence and behavior, they do not theoretically investigate the implications of overconfidence for (misdirected) learning about other variables.

A number of theoretical papers explain why agents become (or seem to become) overconfident. In one class of papers, the agent's learning process is tilted in favor of moving toward or stopping at confident beliefs (Gervais and Odean, 2001, Zábojnik, 2004, Bénabou and Tirole, 2002, Kőszegi, 2006). In other papers, non-common priors or criteria lead agents to take actions that lead the average agent to expect better outcomes than others (Van den Steen, 2004, Santos-Pinto and Sobel, 2005). Finally, some papers assume that an agent simply chooses unrealistically positive beliefs because he derives direct utility from such beliefs (Brunnermeier and Parker, 2005, Oster et al., 2013). While these papers provide foundations for overconfident beliefs and some feature learning, they do not analyze how overconfidence affects learning about other variables.

Many researchers take the view that overconfidence can be individually and socially beneficial even beyond providing direct utility.[20] Our theory is not in contradiction with this view, but it does predict circumstances under which overconfidence can be extremely harmful.

7.2 Learning with Misspecified Models

On a basic level, an overconfident agent has an incorrect view of the world, and hence our paper is related to the literature on learning with misspecified models, that is models in which the support of the prior does not include the true state of the world. Closely related to our theory, Esponda and Pouzo (2016a) develop a general framework for studying repeated games in which players have misspecified models.[21] Methodologically, our model is a special case of theirs in which there is one player.[22] Building on Berk (1966), Esponda and Pouzo 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 stable beliefs have a similar property. Because of our specific setting, we derive stronger results on convergence of beliefs and establish many other properties of the learning process and limiting beliefs.

Our convergence results with endogenous actions contrast with Nyarko (1991), who provides an example in which a misspecified myopic agent's beliefs do not converge. Focusing on the case in which the agent's subjective state space is binary, Fudenberg et al. (forthcoming) fully characterize asymptotic actions and beliefs for any level of patience. Even for a myopic agent beliefs often do not converge. Furthermore, they provide a simple example in which the beliefs of a myopic agent converge, but those of a more patient agent do not. In our model the subjective state space is continuous, and we provide conditions for a continuous state space under which beliefs converge, which allows us to study the implications of overconfidence.

Taking the interpretation that at most one prior can be correct, multi-agent models with non-common priors can also be viewed as analyzing learning with misspecified models. In this literature, papers ask how different agents' beliefs change relative to each other, but do not study the interac-

[20] See, e.g., Taylor and Brown (1988) for a review of the relevant psychology literature, and Bénabou and Tirole (2002) and de la Rosa (2011) for economic examples.

[21] Esponda and Pouzo (2016b) extend the analysis to single-agent Markov decision problems.

[22] Technically, we differ in considering a continuous state space, which allows us to characterize stable beliefs through our no-surprise condition.

tion with behavior. Dixit and Weibull (2007) construct examples in which individuals with different priors interpret signals differently, so that the same signal can push their beliefs further from each other. Similarly, Acemoglu, Chernozhukov and Yildiz (forthcoming) consider Bayesian agents with different prior beliefs regarding the conditional distribution of signals given (what we call) the fundamental, and show that the agents' beliefs regarding the fundamental do not necessarily converge.[23]

Misdirected learning also occurs in some other settings in which individuals have misspecified models of the world. In the social-learning model of Eyster and Rabin (2010) and in many cases also in that of Bohren (2013), agents do not sufficiently account for redundant information in previous actions. With more and more redundant actions accumulating, this mistake is amplified, preventing learning even in the long run. Esponda (2008) studies an adverse-selection environment in which—similarly to the notion of cursed equilibrium by Eyster and Rabin (2005)—a naive buyer underestimates the effect of his price offer on the quality of supplied products. As the buyer learns from experience that the quality is low, he adjusts his price offer downwards, leading to an even worse selection of products and perpetuating the misguided learning. We explore the implications of a very different mistake than these papers.

The interaction between incorrect inferences and behavior we study is somewhat reminiscent of Ellison's (2002) model of academic publishing, in which researchers who are biased about the quality of their own work overestimate publishing standards, making them tougher referees and thereby indeed toughening standards. In contrast to our work, updating is ad-hoc, and the model relies on "feedback" from others on the evolution of the publishing standard. In a similar but more distant vein, Blume and Easley (1982) ask when traders in an exchange economy learn the information held by others through the observation of equilibrium prices, allowing traders to entertain incorrect models. Traders consider a finite set of models—including the true model—and use a "boundedly rational" learning rule that ignores how learning by traders affects equilibrium prices. They show that the true model is locally stable but also that there could be cycles or an incorrect model can

[23] Andreoni and Mylovanov (2012) and Kondor (2012) develop closely related models within the common-prior paradigm. In their models, there are two sources of information about a one-dimensional fundamental. Agents receive a series of public signals, and private signals on how to interpret the public signals. Although beliefs regarding the public information converge, beliefs regarding the fundamental do not, as agents keep interpreting the public information differently. Relatedly, Piketty (1995) analyzes a model in which different personal mobility experiences lead dynasties to develop different steady-state beliefs about the importance of birth versus work in success, and hence to prefer different redistribution policies.

be locally stable.

There is also a substantial amount of other research on the learning implications of various mistakes in interpreting information (see for instance Rabin and Schrag, 1999, Rabin, 2002, Madarász, 2009, Rabin and Vayanos, 2010, Benjamin, Rabin and Raymond, 2012, Spiegler, 2016). Overconfidence is a different type of mistake—in particular, it is not directly an inferential mistake—so our results have no close parallels in this literature.

Methodologically, our theory confirms Fudenberg's (2006) point that it is often insufficient to do behavioral economics by modifying one assumption of a classical model, as one modeling change often justifies other modeling changes as well. In our setting, the agent's false belief about his ability leads him to draw incorrect inferences regarding the fundamental, so assuming that an overconfident agent is otherwise classical may be misleading.

8 Conclusion

While our paper focuses exclusively on individual decisionmaking, the possibility of self-defeating learning likely has important implications for multi-agent situations. For example, it has been recognized in the literature that managerial overconfidence can benefit a firm both because it leads the manager to overvalue bonus contracts and because it can lead him to exert greater effort (de la Rosa, 2011, for example). Yet for tasks with the properties we have identified, misguided learning can also induce a manager to make highly suboptimal decisions. Hence, our analysis may have implications for the optimal allocation of decisionmaking authority for a manager.

Beyond how an overconfident agent operates in a standard economic environment, it seems relevant to understand how multiple overconfident individuals interact with each other. As a simple example, consider again our application to assertiveness in personal relationships (Application 3), where an overconfident agent misinterprets his partner's actions and ends up treating her more and more assertively. This dynamic is likely to be reinforced if the partner is also overconfident and hence similarly misinterprets the agent's actions, creating a downward spiral on both sides. Peace has a chance only if the partners "indulge" each other's overconfidence by holding overly positive views of each other.

A completely different issue is that different individuals may have different interpretations as to what explains unexpectedly low performance. For instance, a Democrat may interpret poor health outcomes as indicating problems with the private market, whereas a Republican may think that the culprit is government intervention. In the formalism of our model, one side believes that output is increasing in ϕ, while the other side believes that output is decreasing in ϕ. Similarly to Dixit and Weibull's (2007) model of political polarization, decisionmakers with such opposing theories may prefer to adjust policies in different directions. Our model highlights that unrealistically high expectations regarding outcomes can play an important role in political polarization. Furthermore, our model makes predictions on how the two sides interpret each other's actions. It may be the case, for instance, that if a Republican has the power to make decisions, he engages in self-defeating learning as our model predicts, with a Democrat looking on in dismay and thinking that—if only he had the power—a small adjustment in the opposite direction would have been sufficient. If the Democrat gets power, he adjusts a little in the opposite direction, initially improving performance, but then he engages in self-defeating learning of his own. Frequent changes in power therefore help keep self-defeating learning in check. Meanwhile, an independent who also has unrealistic expectations but is unsure about which side's theory of the world is correct always tends to move his theory toward the theory of the party in opposition, as that party's theory is better at explaining current observations.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz, "Fragility of Asymptotic Agreement under Bayesian Learning," *Theoretical Economics*, forthcoming.

Alonso, Ricardo, Wouter Dessein, and Niko Matouschek, "When Does Coordination Require Centralization?," *American Economic Review*, 2008, *98* (1), 145–179.

Andreoni, James and Tymofiy Mylovanov, "Diverging Opinions," *American Economic Journal: Microconomics*, 2012, *4* (1), 209–232.

Andrews, Donald WK, "Generic Uniform Convergence," *Econometric Theory*, 1992, *8* (2), 241–257.

Argyris, Chris, *Reasoning, Learning and Action*, San Francisco, CA: Jossey-Bass, 1982.

Babcock, Linda and George Loewenstein, "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives*, 1997, *11* (1), 109–126.

Baumeister, Roy F., Laura Smart, and Joseph M. Boden, "Relation of Threatened Egotism to Violence and Aggression: The Dark Side of High Self-Esteem," *Psychological Review*, 1996, *103* (1), 5–33.

Bénabou, Roland and Jean Tirole, “Self-Confidence and Personal Motivation,” *Quarterly Journal of Economics*, 2002, *117* (3), 871–915.

_ **and** _ , “Over My Dead Body: Bargaining and the Price of Dignity,” *American Economic Review Papers and Proceedings*, 2009, *99* (2), 459–465.

_ **and** _ , “Mindful Economics: The Production, Consumption, and Value of Beliefs,” *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

Benjamin, Daniel J., Matthew Rabin, and Collin Raymond, “A Model of Non-Belief in the Law of Large Numbers,” 2012. Working Paper, Cornell University.

Benoît, Jean-Pierre and Juan Dubra, “Apparent Overconfidence,” *Econometrica*, 2011, *79* (5), 1591–1625.

_ **,** _ **, and Don A. Moore**, “Does the Better-Than-Average Effect Show That People Are Overconfident? Two Experiments.,” *Journal of the European Economic Association*, 2015, *13* (2), 293–329.

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

Bleemer, Zachary and Basit Zafar, “Intended College Attendance: Evidence from an Experiment on College Returns and Costs,” 2015. Working Paper.

Blume, Lawrence and David Easley, “Learning to Be Rational,” *Journal of Economic Theory*, 1982, *26*, 340–351.

Bohren, Aislinn, “Informational herding with Model Mispecification,” June 2013. Working Paper.

Brunnermeier, Markus and Jonathan Parker, “Optimal Expectations,” *American Economic Review*, 2005, *95* (4), 1092–1118.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini, “Overconfidence and Social Signalling,” *The Review of Economic Studies*, 2013, *80* (3), 949–983.

Camerer, Colin and Dan Lovallo, “Overconfidence and Excess Entry: An Experimental Approach,” *American Economic Review*, 1999, *89* (1), 306–318.

Campbell, W. Keith and Constantine Sedikides, “Self-Threat Magnifies the Self-Serving Bias: A Meta-Analytic Integration,” *Review of General Psychology*, 1999, *3* (1), 23–43.

Card, David and Gordon B. Dahl, “Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior,” *Quarterly Journal of Economics*, March 2011, *126* (1), 103–143.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven, “Self-Confidence and Strategic Behavior,” 2014. Working Paper.

Chiang, Yao-Min, David Hirshleifer, Yiming Qian, and Ann E. Sherman, “Do Investors Learn from Experience? Evidence from Frequent IPO Investors,” *Review of Financial Studies*, 2011, *24* (5), 1560–1589.

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

Crozier, Michel, *The Bureaucratic Phenomenon*, Chicago: University of Chicago Press, 1964.

de la Rosa, Leonidas Enrique, "Overconfidence and Moral Hazard," *Games and Economic Behavior*, 2011, *73* (2), 429–451.

DellaVigna, Stefano and Ulrike Malmendier, "Paying Not to Go to the Gym," *American Economic Review*, 2006, *96* (3), 694–719.

Dixit, Avinash K. and Jörgen Weibull, "Political Polarization," *Proceedings of the National Academy of Sciences*, 2007, *104* (18), 7351–7356.

Ellison, Glenn, "Evolving Standards for Academic Publishing: A q-r Theory," *Journal of Political Economy*, October 2002, *110* (5), 994–1034.

Esponda, Ignacio, "Behavioral Equilibrium in Economies with Adverse Selection," *American Economic Review*, 2008, *98* (4), 1269–1291.

_ 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. available at `http://arxiv.org/abs/1502.06901`.

Eyster, Erik and Matthew Rabin, "Cursed Equilibrium," *Econometrica*, 2005, *73* (5), 1623–1672.

_ and _ , "Naïve Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2010, *2* (4), 221–243.

Fudenberg, Drew, "Advancing Beyond Advances in Behavioral Economics," *Journal of Economic Literature*, 2006, *44* (3), 694–711.

_ , Gleb Romanyuk, and Philipp Strack, "Active Learning with a Misspecified Prior," *Theoretical Economics*, forthcoming.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein, "Channeled Attention and Stable Errors," 2017. Working Paper.

Galasso, Alberto and Timothy S. Simcoe, "CEO Overconfidence and Innovation," *Management Science*, 2011, *57* (8), 1469–1484.

Gervais, Simon and Terrance Odean, "Learning To Be Overconfident," *Review of Financial Studies*, 2001, *14* (1), 1–27.

_ , J. B. Heaton, and Terrance Odean, "Overconfidence, Compensation Contracts, and Capital Budgeting," *Journal of Finance*, 2011, *66* (5), 1735–1777.

Ghosal, Subhashis and Aad van der Vaart, "Convergence Rates of Posterior Distributions for Noniid Observations," *Annals of Statistics*, 2007, *35* (1), 192–223.

Graham, John R., Campbell R. Harvey, and Manju Puri, "Managerial Attitudes and Corporate Actions," *Journal of Financial Economics*, 2013, *109* (1), 103 – 121.

Hirshleifer, David, Angie Low, and Siew Hong Teoh, "Are Overconfident CEOs Better Innovators?," *Journal of Finance*, 08 2012, *67* (4), 1457–1498.

Hoffman, Mitch and Stephen Burks, "Training Contracts, Worker Overconfidence, and the Provision of Firm-Sponsored General Training," 2013. Working Paper.

Holmström, Bengt and Paul Milgrom, "Multi-Task Principal-Agent Analyzes: Incentive Contracts, Asset Ownership,andJob Design," *Journal of Law, Economics and Organization*, 1991, *7*, 24–52. special issue.

Jehiel, Philippe, "Investment Strategy and Selection Bias: An Equilibrium Perspective on Overoptimism," 2016. Working Paper.

Kőszegi, Botond, "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, 2006, *4* (4), 673–707.

Kondor, Péter, "The More We Know about the Fundamental, the Less We Agree on the Price," *Review of Economic Studies*, 2012, *79* (3), 1175–1207.

Lambe, Sinead, Catherine Hamilton-Giachritsis, Emily Garner, and Julian Walker, "The Role of Narcissism in Aggression and Violence: A Systematic Review," *Trauma, Violence, & Abuse*, forthcoming.

Landier, Augustin and David Thesmar, "Financial Contracting with Optimistic Entrepreneurs," *Review of Financial Studies*, 2009, *22* (1), 117–150.

Madarász, Kristóf, "Information Projection: Model and Applications," 2009. SSRN Paper #1407190.

Malmendier, Ulrike and Geoffrey Tate, "CEO Overconfidence and Corporate Investment," *Journal of Finance*, 2005, *60* (6), 2661–2700.

_ **and** _ , "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction," *Journal of Financial Economics*, 2008, *89* (1), 20–43.

Manzoni, Jean-François and Jean-Louis Barsoux, "The Set-Up-to-Fail Syndrome," *Harvard Business Review*, March-April 1998, pp. 101–113.

_ **and** _ , *The Set-Up-to-Fail Syndrome*, Harvard Business School Press, 2002.

March, James G. and Herbert A. Simon, *Organizations*, New York: Wiley, 1958.

Masuch, Michael, "Vicious circles in organizations," *Administrative Science Quarterly*, 1985, pp. 14–33.

McFarlin, Dean B., Roy F. Baumeister, and Jim Blascovich, "On Knowing When to Quit: Task Failure, Self-Esteem, Advice, and Nonproductive Persistence," *Journal of Personality*, 1984, *52* (2), 138–155.

Miller, Dale T. and Michael Ross, "Self-serving Biases in the Attribution of Causality: Fact or Fiction?," *Psychological Bulletin*, 1975, *82* (2), 213–225.

Moore, Don A. and Paul J. Healy, "The Trouble with Overconfidence," *Psychological Review*, 2008, *115* (2), 502–517.

Morris, Stephen and Hyun Song Shin, "Social Value of Public Information," *American Economic Review*, 2002, *92* (5), 1521–1534.

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

Oster, Emily, Ira Shoulson, and E. Ray Dorsey, "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease," *American Economic Review*, 2013, *103* (2), 804–30.

Piketty, Thomas, "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, August 1995, *110*, 551–584.

Rabin, Matthew, "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, 2002, *117* (3), 775–816.

_ **and Dimitri Vayanos**, "The Gambler's and Hot-Hand Fallacies: Theory and Applications," *Review of Economic Studies*, 2010, *77* (2), 730–778.

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

Ross, Michael and Fiore Sicoly, "Egocentric Biases in Availability and Attribution," *Journal of Personality and Social Psychology*, 1979, *37* (3), 322–336.

Santos-Pinto, Luis and Joel Sobel, "A Model of Positive Self-Image in Subjective Assessments," *American Economic Review*, 2005, *95* (5), 1386–1402.

Saumard, Adrien and Jon A. Wellner, "Log-Concavity and Strong Log-Concavity: A Review," *Statistics Surveys*, 2014, *8*, 45–114.

Savage, Leonard J., *The Foundations of Statistics*, New York, Wiley, 1954.

Schrand, Catherine M. and Sarah L.C. Zechman, "Executive overconfidence and the slippery slope to financial misreporting," *Journal of Accounting and Economics*, 2012, *53* (1–2), 311 – 329.

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

Shui, Haiyan and Lawrence M. Ausubel, "Time Inconsistency in the Credit Card Market," 2004. Available at http://ssrn.com/abstract=586622.

Spiegler, Ran, "Bayesian Networks and Boundedly Rational Expectations," *Quarterly Journal of Economics*, 2016, *131* (3), 1243–1290.

Spinnewijn, Johannes, "Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs," 2012. Working Paper.

Svenson, Ola, "Are We Less Risky and More Skillful Than Our Fellow Drivers?," *Acta Psychologica*, 1981, *18*, 473–474.

Taylor, Shelley E. and Jonathan D. Brown, "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin*, 1988, *103* (2), 193–210.

Tetlock, Philip E. and Ariel Levi, "Attribution Bias: On the Inconclusiveness of the Cognition-Motivation Debate," *Journal of Experimental Social Psychology*, 1982, *18* (1), 68–88.

Van den Steen, Eric, "Rational Overoptimism (and Other Biases)," *American Economic Review*, 2004, *94* (4), 1141–1151.

Weinstein, Neil D., "Unrealistic Optimism About Future Life Events," *Journal of Personality and Social Psychology*, 1980, *39*, 806–820.

Williams, David, *Probability with martingales*, Cambridge university press, 1991.

Yaouanq, Yves Le and Nina Hestermann, "It's Not My Fault! Ego, Excuses and Perseverance," 2016. Working Paper.

Zábojnik, Ján, "A Model of Rational Bias in Self-Assessments," *Economic Theory*, 2004, *23* (2), 259–282.

A Proofs

A.1 Auxiliary Results

This preliminary section introduces useful notation for the main proofs, shows that output has an intuitive influence on subjective beliefs, and apply a well-known result from Berk (1966) on the convergence of subjective beliefs with fixed actions to our setting.

Let $\mathcal{L}(x) = \mathbb{E}[\log f(x + \epsilon_t)]$, where the expectation is taken with respect to ϵ_t. Note that as $\epsilon_1, \epsilon_2, ...$ are i.i.d., this expectation is independent of t. Our first result shows that the log-concavity of f implies that $\mathcal{L}$ is single-peaked, with its peak at zero. Let $g(x) = \frac{\partial}{\partial x}(\log f(x))$, and let $|\underline{\kappa}_f|$ denote the bound on the absolute value of g', i.e $|\underline{\kappa}_f| > |g'|$.

Lemma 1. *$\mathcal{L}(x)$ is well-defined for all $x \in \mathbb{R}$. Furthermore,*

i) $\mathcal{L}(\cdot)$ is strictly concave; and *ii)* $\mathcal{L}'(x) > 0$ *for* $x < 0$ *and* $\mathcal{L}'(x) < 0$ *for* $x > 0$.

Proof of Lemma 1. As the distribution of ϵ is log-concave it has finite variance $\sigma^2 = \mathbb{E}[|\epsilon_t|^2]$ (see for example Proposition 5.2 in Saumard and Wellner (2014)). Our bounded log-concavity assumption on f implies that

$$\begin{aligned}|\log(f(x+\epsilon))| &= \left|\log(f(x)) + \int_x^{x+\epsilon} g(z)\mathrm{d}z\right| = \left|\log(f(x)) + \epsilon g(x) + \int_x^{x+\epsilon}\int_x^z g'(y)\mathrm{d}y\mathrm{d}z\right| \\ &\le |\log(f(x))| + |\epsilon|\,|g(x)| + |\epsilon|^2\frac{|\underline{\kappa}_f|}{2}.\end{aligned}$$

It thus follows that

$$\begin{aligned}|\mathcal{L}(x)| = \left|\mathbb{E}[\log f(x+\epsilon_t)]\right| &\le \mathbb{E}[|\log f(x+\epsilon_t)|] \le |\log(f(x))| + \mathbb{E}[|\epsilon_t|]\,|g(x)| + \mathbb{E}[|\epsilon_t|^2]\frac{|\underline{\kappa}_f|}{2} \\ &\le |\log(f(x))| + \sigma\,|g(x)| + \sigma^2\frac{|\underline{\kappa}_f|}{2},\end{aligned}$$

where we used in the last step that by Jensen's inequality $\mathbb{E}[|\epsilon|] \le \sqrt{\mathbb{E}[|\epsilon|^2]} = \sigma$. Thus, $\mathcal{L}(x)$ is well defined for every x.

As our log-concavity assumption implies that $|g|$ and $|g'|$ are bounded by an integrable function, majorized convergence yields that the derivatives of $\mathcal{L}(\cdot)$ are given by

$$\begin{aligned}\mathcal{L}'(x) &= \int_{\mathbb{R}} g(x+\epsilon)f(\epsilon)\mathrm{d}\epsilon, \\ \mathcal{L}''(x) &= \int_{\mathbb{R}} g'(x+\epsilon)f(\epsilon)\mathrm{d}\epsilon \in [\underline{\kappa}_f, 0).\end{aligned}$$

Since $g' < 0$, $\mathcal{L}$ is strictly concave. To see that the peak is at zero, note that

$$\mathcal{L}(x) - \mathcal{L}(0) = \mathbb{E}[\log f(x+\epsilon_t) - \log f(\epsilon_t)] = -\mathbb{E}\Big[\log\Big(\frac{f(\epsilon_t)}{f(x+\epsilon_t)}\Big)\Big].$$

The right-hand side equals minus the Kullback-Leibler Divergence. By Gibb's inequality the Kullback-Leibler Divergence is larger or equal to zero, holding with equality if and only if both densities coincide a.e.. Hence, the right-hand side is minimized at $x = 0$, so that $\mathcal{L}(\cdot)$ is maximized at $x = 0$. Finally, as $\mathcal{L}$ is strictly concave and maximized at 0, it follows that $\mathcal{L}'(0) = 0$, and $\mathcal{L}'(x)$ is thus positive for $x < 0$ and negative for $x > 0$. $\square$

Denote by $\ell_0 : \mathbb{R} \to \mathbb{R}$ the subjective prior log-likelihood of the agent. By Bayes' rule the agent's subjective log-likelihood $\ell_t : \mathbb{R} \to \mathbb{R}$ assigned to the state ϕ in period t is given by

$$\ell_t(\phi) = \sum_{s=1}^t \log f(q_s - Q(e_s, \tilde{a}, \phi)) + \ell_0(\phi). \tag{5}$$

The density function $\pi_t : \mathbb{R} \to \mathbb{R}_+$ of the agent's subjective belief in period t equals

$$\pi_t(\phi) = \frac{e^{\ell_t(\phi)}}{\int_{-\infty}^{\infty} e^{\ell_t(z)} \mathrm{d}z} .$$

Denote by $\tilde{\mathbb{P}}_t[\cdot] = \tilde{\mathbb{P}}[\cdot \mid q_1, \ldots, q_t]$ the agents subjective probability measure over states conditional on the outputs $q_1, \ldots, q_t$ and by $\Pi_t : \mathbb{R} \to [0, 1]$ the cdf of the agent's subjective belief

$$\Pi_t(z) = \tilde{\mathbb{P}}_t[\Phi \leq z] = \int_{-\infty}^{z} \pi_t(\phi)\mathrm{d}\phi .$$

We sometimes write $\ell_t(\phi; q)$ and $\pi_t(\phi; q)$ if we want to highlight the dependence of the agent's belief on the outputs $q_1, \ldots, q_t$ observed in previous periods.

Definition 2 (Monotone Likelihood Ratio Property)**.** A distribution with density $\pi : \mathbb{R} \to \mathbb{R}_+$ is greater than a distribution with density $\pi' : \mathbb{R} \to \mathbb{R}_+$ in the sense of monotone likelihood ratio $\pi' \leq_{MLR} \pi$ if and only if for all states $\phi' \leq \phi$,

$$\frac{\pi'(\phi)}{\pi(\phi)} \leq \frac{\pi'(\phi')}{\pi(\phi')} . \tag{MLRP}$$

The next Lemma shows that higher observed output q_t in any period t leads to a higher posterior belief in the sense of monotone likelihood ratios (MLR).

Lemma 2 (Beliefs are Monotone in Output)**.** *If $q'_s \leq q_s$ for all $s \leq t$ then*

$$\pi_t(\cdot; q') \leq_{MLR} \pi_t(\cdot; q).$$

Proof of Lemma 2. Condition (MLRP) is equivalent to

$$0 \leq \log\left[\frac{\pi_t(\phi; q)}{\pi_t(\phi'; q)}\right] - \log\left[\frac{\pi_t(\phi; q')}{\pi_t(\phi'; q')}\right] = [\ell_t(\phi; q) - \ell_t(\phi'; q)] - [\ell_t(\phi; q') - \ell_t(\phi'; q')] .$$

Hence, it suffices to show that for all $s \leq t$

$$\frac{\partial^2 \ell_t(\phi; q)}{\partial q_s \partial \phi} \geq 0 .$$

Taking the derivative of the log-likelihood given in Eq. 5 yields that for $s \leq t$

$$\frac{\partial^2 \ell_t(\phi; q)}{\partial q_s \partial \phi} = -g'(q_s - Q(e_s, \tilde{a}, \phi))Q_\phi(e_s, \tilde{a}, \phi) .$$

As $g' < 0$ and $Q_\phi > 0$, the result follows. $\square$

Our next result shows that beliefs converge when the agent takes the same action in every period. As in this case the outputs are independent, the result follows immediately from the characterization of long-run beliefs with i.i.d. signals in Berk (1966).

Lemma 3 (Berk (1966)). *Suppose that the agent takes a fixed action e in all periods and there exists a state $\phi^e_\infty \in (\underline{\phi}, \overline{\phi})$ that satisfies*

$$Q(e, A, \Phi) = Q(e, \tilde{a}, \phi^e_\infty) \,. \tag{6}$$

Then the agent's belief a.s. converges and concentrates on the unique state ϕ^e_∞.

Proof of Lemma 3. For fixed actions, outputs $q_1, q_2, \ldots$ are i.i.d. random variables. The expecion of an outside observer of the log-likelihood that the agent assigns to the state ϕ after observing a single signal q_t is given by

$$\mathbb{E}[\log f(q_t - Q(e, \tilde{a}, \phi))] = \mathbb{E}[\log f(Q(e, A, \Phi) - Q(e, \tilde{a}, \phi) + \epsilon_t)] = \mathcal{L}(Q(e, A, \Phi) - Q(e, \tilde{a}, \phi)) \,. \tag{7}$$

By the main theorem in Berk (1966, page 54), the agent's subjective belief concentrates on the set of maximizers of (7). By Lemma 1, $\mathcal{L}(\cdot)$ is uniquely maximized at zero and, hence, (7) is maximized whenever (6) is satisfied. As Q_ϕ is positive and bounded away from zero, there exists a unique such point, and the agent's subjective belief converges to a Dirac measure on that point. □

Let us denote a Dirac measure on the state ϕ by δ_ϕ.

Lemma 4. *(e, π) is a pure-strategy Berk-Nash equilibrium if and only if π is a stable belief δ_ϕ and $e = e^*(\phi)$.*

Proof. For a given state ϕ' and action e, the Kullback-Leibler-divergence between the true distribution of signals and the distribution that the agent expects is given by

$$\mathbb{E}\left[\log \frac{f(\epsilon_t)}{f(Q(e, A, \Phi) - Q(e, \tilde{a}, \phi') + \epsilon_t)}\right] = \mathcal{L}(0) - \mathcal{L}(Q(e, A, \Phi) - Q(e, \tilde{a}, \phi')) \,.$$

By the definition of Berk-Nash equilibrium, the agent assigns positive probability only to states that minimize the Kullback-Leibler-divergence. Since, by Lemma 1, $\mathcal{L}(\cdot)$ is maximized at 0, it follows that the agent's belief is a Dirac measure on the state ϕ that satisfies $Q(e, A, \Phi) - Q(e, \tilde{a}, \phi) = 0$. Because the equilibrium action must be optimal given this belief, it follows that $(e^*(\phi), \delta_\phi)$ is a pure-strategy Berk-Nash equilibrium if and only if ϕ is a stable belief. □

A.2 Main Results

A.2.1 Proving Properties of Limiting Beliefs

We first show that all stable beliefs are in an interval around the true state Φ:

Lemma 5. *Let $\overline{\kappa}_a \geq Q_a$ be the upper bound on Q_a and $0 < \underline{\kappa}_\phi \leq Q_\phi$ the lower bound on Q_ϕ. Any root of Γ lies in the interval $I_{\tilde{a}} = [\Phi - \frac{\overline{\kappa}_a}{\underline{\kappa}_\phi}(\tilde{a} - A), \Phi]$. Furthermore, $\Gamma(\Phi) < 0$ and $\Gamma(\Phi - \frac{\overline{\kappa}_a}{\underline{\kappa}_\phi}(\tilde{a} - A)) \geq 0$, so if $I_{\tilde{a}} \subset (\underline{\phi}, \overline{\phi})$ then Γ has at least one root in $I_{\tilde{a}}$.*

Proof of Lemma 5. Note that for $\tilde{a} = A$, the surprise function

$$\Gamma(\phi) = Q(e^*(\phi), A, \Phi) - Q(e^*(\phi), \tilde{a}, \phi),$$

has a unique root at $\phi = \Phi$ since $Q_\phi > 0$. Furthermore, since $Q_a > 0$ and $Q_\phi > 0$, when $\tilde{a} > A$ any root of Γ must be less than Φ. Now for any $\phi < \Phi$,

$$\begin{aligned}
\Gamma(\phi) &\geq \min_e \{Q(e, A, \Phi) - Q(e, \tilde{a}, \phi)\} \\
&= \min_e \{Q(e, A, \Phi) - Q(e, A, \phi) + Q(e, A, \phi) - Q(e, \tilde{a}, \phi)\} \\
&\geq \underline{\kappa}_\phi(\Phi - \phi) - \overline{\kappa}_a(\tilde{a} - A).
\end{aligned}$$

Therefore, any root of Γ must line in the interval $I_{\tilde{a}} = [\Phi - \frac{\overline{\kappa}_a}{\underline{\kappa}_\phi}(\tilde{a} - A), \Phi]$. Furthermore, since $A < \tilde{a}$ we have that $\Gamma(\Phi) < 0$, and by the above inequality $\Gamma(\Phi - \frac{\overline{\kappa}_a}{\underline{\kappa}_\phi}(\tilde{a} - A)) \geq 0$. It thus follows from the continuity of Γ that Γ has at least one root in $I_{\tilde{a}}$. $\square$

Proof of Proposition 1.

Proof of Part (i): We first show that for $\tilde{a} - A$ sufficiently small Γ can have at most one root. Since $e^*(\phi)$ is implicitly defined through $Q_e(e^*(\phi), \tilde{a}, \phi) = 0$ and Q is twice continuously differentiable with $Q_{ee} < 0$, $e^*(\phi)$ is a continuous function of ϕ and $\tilde{a}$. By the implicit function theorem, $e^{*'}(\phi) = -Q_{e\phi}/Q_{ee}$ and hence is also a continuous function of ϕ and $\tilde{a}$. Thus,

$$\Gamma'(\phi) = Q_e(e^*(\phi), A, \Phi)\ (e^*)'(\phi) - Q_\phi(e^*(\phi), \tilde{a}, \phi)$$

is a continuous function of ϕ and $\tilde{a}$. Since for $\tilde{a} = A$ and $\phi = \Phi$,

$$\Gamma'(\phi)|_{\phi=\Phi, \tilde{a}=A} = -Q_\phi(e^*(\phi), A, \Phi) < 0,$$

continuity of $\Gamma'(\phi)$ implies that there exists a pair $\eta_A, \eta_\Phi > 0$ such that for all $\tilde{a} \in [A, A + \eta_A)$ and $\phi \in (\Phi - \eta_\Phi, \Phi]$ one has $\Gamma'(\phi) < 0$. Thus, for any $\tilde{a} \geq A$ that satisfies $\tilde{a} < A + \min\{\eta_A, \eta_\Phi + \frac{\underline{\kappa}_\phi}{\overline{\kappa}_a}\}$ one has that $\Gamma'(\phi) < 0$ over the relevant interval $I_{\tilde{a}}$. By Lemma 5, all roots of Γ lie in $I_{\tilde{a}}$ and thus for all such $\tilde{a}$, Γ has a unique root. Furthermore, for $\tilde{a} - A$ small enough $I_{\tilde{a}} \subset (\underline{\phi}, \overline{\phi})$ and hence by

Lemma 5 Γ crosses zero in $(\underline{\phi}, \overline{\phi})$. We conclude that Γ has a unique root if overconfidence $(\tilde{a} - A)$ is sufficiently small.

Proof of Part (ii): When Q takes the form in Example 1, then $e^*(\phi) = \phi$ and

$$\Gamma(\phi) = -(\tilde{a} - A) + (\Phi - \phi) - L(\Phi - \phi).$$

Since $L'(x) < 1$, $\Gamma'(\phi) < 0$ and hence Γ has at most one root. Finally, as $(\underline{\phi}, \overline{\phi}) = \mathbb{R}$, we have that $I_{\tilde{a}} \subset (\underline{\phi}, \overline{\phi})$ and hence there exists a stable belief by Lemma 5.

Proof of Part (iii): Since $Q_a > 0$ and $Q_\phi > 0$, when $\tilde{a} > A$ any root of Γ must be less than Φ. ϕ is a root of Γ if and only if $(A + e^*(\phi))\Phi = (\tilde{a} + e^*(\phi))\phi$, or

$$e^*(\phi) = -A + (\tilde{a} - A)\phi/(\Phi - \phi) \tag{8}$$

The right-hand side of this equation is increasing and convex over the interval $(0, \Phi)$, negative at $\phi = 0$, and approaches ∞ as ϕ approaches Φ.

Furthermore, for $e^*(\phi)$ to be optimal given ϕ, we must have $c'(e^*(\phi)) = \phi$. This implies that $e^{*\prime}(\phi) = 1/c''(e^*(\phi))$, so that if $c'''(e) \geq 0$, then $e^*(\phi)$ is concave. Furthermore $e^*(0) = 0$ and $e^*(\Phi)$ is finite, so that it equals the right-hand-side of (8) at exactly one point. □

Proof of Proposition 2. By Lemma 3, if for any fixed action e the unique solution ϕ^e_∞ to

$$0 = Q(e, A, \Phi) - Q(e, \tilde{a}, \phi^e_\infty)$$

satisfies $\phi^e_\infty \in (\overline{\phi}, \underline{\phi})$, beliefs converge to a Dirac measure on ϕ^e_∞. Since $\tilde{a} > A$ and $Q_a, Q_\phi > 0$, we have $\phi^e_\infty < \Phi$. Let ϕ_∞ be the state corresponding to the unique stable belief. Since $\phi_\infty \in (\overline{\phi}, \underline{\phi})$ and $\phi^e_\infty < \Phi < \overline{\phi}$, we have that $\phi_\infty \leq \phi^e_\infty$ implies $\phi^e_\infty \in (\overline{\phi}, \underline{\phi})$.

By definition of ϕ_∞,

$$\Gamma(\phi_\infty) = Q(e(\phi_\infty), A, \Phi) - Q(e(\phi_\infty), \tilde{a}, \phi_\infty) = 0.$$

Since $Q_a > 0$ and $Q_\phi > 0$, when $\tilde{a} > A$ one has $\phi_\infty < \Phi$. Furthermore, when the agent's belief are a Dirac measure on ϕ, he chooses the myopically optimal action that satisfies

$$Q_e(e(\phi), \tilde{a}, \phi) = 0.$$

By the implicit function theorem, $e^{*'}(\phi) = -Q_{e\phi}/Q_{ee} > 0$. Hence, $e(\phi_\infty) < e^*(\Phi)$.

Since $Q_{e\phi} > 0$ and, by Assumption 1, $Q_{ae} \leq 0$, the derivative

$$\frac{\partial}{\partial e}\left[Q(e, A, \Phi) - Q(e, \tilde{a}, \phi_\infty^e)\right] = -\int_A^{\tilde{a}} Q_{ae}(e, a, \Phi)da + \int_{\phi_\infty^e}^{\Phi} Q_{\phi e}(e, \tilde{a}, \phi)d\phi > 0.$$

Because $Q(e, A, \Phi) - Q(e, \tilde{a}, \phi_\infty^e) = 0$ and $e(\phi_\infty) < e^*(\Phi) \leq e$, therefore,

$$Q(e(\phi_\infty), A, \Phi) - Q(e(\phi_\infty), \tilde{a}, \phi_\infty^e) < 0 = Q(e(\phi_\infty), A, \Phi) - Q(e(\phi_\infty), \tilde{a}, \phi_\infty).$$

Since $Q_\phi > 0$, it follows that $\phi_\infty < \phi_\infty^e$. $\square$

Proof of Proposition 3. Since $Q_a > 0$, $Q_\phi > 0$ and $\tilde{a} > A$, the surprise function Γ_i is negative for all $\phi \geq \Phi$ and hence has a negative slope at its unique root $\phi_{\infty,i}$. Furthermore, when the agent's belief are a Dirac measure on ϕ, he chooses the myopically optimal action that satisfies

$$Q_e(e(\phi), \tilde{a}, \phi) = 0.$$

By the implicit function theorem, $e^{*'}(\phi) = -Q_{e\phi}/Q_{ee} > 0$ and $e^{*'}(\tilde{a}) = -Q_{ea}/Q_{ee} \leq 0$. Since $\phi_{\infty,i} < \Phi$ and $\tilde{a} > A$, the agent chooses a suboptimally low stable action. Because $\Gamma_i(\phi) = R_i(A, \Phi) - R_i(\tilde{a}, \phi) - L_i(e^*(\tilde{a}, \phi), A, \Phi)$, $\Gamma_1 > \Gamma_2$ pointwise at all but the optimal action, the fact that the agent chooses some suboptimal action implies that $\phi_{\infty,1} > \phi_{\infty,2}$. $\square$

Proof of Proposition 4. $I \Rightarrow II$. Denote by $\tilde{\phi}(A, \tilde{a}, \Phi)$ the state corresponding to the unique stable belief when perceived ability equals $\tilde{a}$, true ability equals A, and the true state equals Φ. Denote by $e^*(a, \phi)$ the optimal action when the ability equals a and the state equals ϕ.

Since the action is objectively optimal for the state Φ as well as subjectively optimal when the agent holds beliefs $\tilde{\phi}(A, \tilde{a}, \Phi)$, we have

$$e^*(A, \Phi) = \arg\max_e Q(e, A, \Phi) = \arg\max_e Q(e, \tilde{a}, \tilde{\phi}(A, \tilde{a}, \Phi)) = e^*(\tilde{a}, \tilde{\phi}(A, \tilde{a}, \Phi)). \tag{9}$$

Furthermore, by the definition of a stable belief, the agent gets no surprise:

$$Q(e^*(A, \Phi), A, \Phi) = Q(e^*(A, \Phi), \tilde{a}, \tilde{\phi}(A, \tilde{a}, \Phi)). \tag{10}$$

We establish some properties of $e^*(a, \phi)$. First, we know that $e^*_\phi(a, \phi) > 0$. Second, we show that $e^*_a(a, \phi) > 0$ (recall that we do not assume $Q_{ea} \geq 0$ in Proposition 4). Totally differentiating

Equation (10) with respect to $\tilde{a}$ and using that $Q_a > 0$, we get that $\tilde{\phi}_{\tilde{a}}(a, \tilde{a}, \phi) < 0$. Furthermore, by Equation (9) we have $e^*(a, \phi) = e^*(\tilde{a}, \tilde{\phi}(a, \tilde{a}, \phi))$. Totally differentiating this equality with respect to $\tilde{a}$ gives that $e^*_{\tilde{a}}(\tilde{a}, \tilde{\phi}_{\tilde{a}}(a, \tilde{a}, \phi)) = -e^*_{\phi}(\tilde{a}, \tilde{\phi}_{\tilde{a}}(a, \tilde{a}, \phi))\tilde{\phi}_{\tilde{a}}(a, \tilde{a}, \phi) > 0$. Setting $\tilde{a} = a$ and using that $\tilde{\phi}(a, a, \phi) = \phi$ establishes our claim that $e^*_a(a, \phi) > 0$.

Given these properties, we can define $S(a, \phi) = e^*(a, \phi)$, and this function satisfies $S_a, S_\phi > 0$. Furthermore, using again that $e^*(a, \phi) = e^*(\tilde{a}, \tilde{\phi}(a, \tilde{a}, \phi))$, note that for any $a, \tilde{a}, \phi$, the unique ϕ' satisfying $S(a, \phi) = S(\tilde{a}, \phi')$ is $\phi' = \tilde{\phi}(a, \tilde{a}, \phi)$. Now we define $V(e, S) = S - L(|e - S|)$, where $L(\cdot)$ is any strictly increasing function satisfying $L'(x) < 1$ everywhere. By construction, $V_S > 0$. And because $S(\tilde{a}, \tilde{\phi}(A, \tilde{a}, \Phi)) = S(A, \Phi)$, $\Gamma(\tilde{\phi}(A, \tilde{a}, \Phi)) = 0$. Thus, the Dirac measure on $\tilde{\phi}(A, \tilde{a}, \Phi)$ is the stable belief and $S(\tilde{a}, \tilde{\phi}(A, \tilde{a}, \Phi)) = e^*(\tilde{a}, \tilde{\phi}(A, \tilde{a}, \Phi)) = e^*(A, \Phi)$ is the stable action.

II $\Rightarrow$ *I*. Because $V_S, S_a, S_\phi > 0$, for any action e there is a unique $\tilde{\phi}$ such that $V(e, S(A, \Phi)) = V(e, S(\tilde{a}, \tilde{\phi}))$. At this $\tilde{\phi}$, $\Gamma(\tilde{\phi}) = 0$, and hence $\delta_{\tilde{\phi}} = \delta_{\phi_\infty}$ is a stable belief. Hence the stable action $e^*(\phi_\infty)$ of the agent satisfies $V_e(e^*(\phi_\infty), S(\tilde{a}, \phi_\infty)) = 0$, and since $V_e(e, S(\tilde{a}, \phi_\infty)) = V_e(e, S(A, \Phi))$, the stable action is optimal if the output function takes the form $V(e, S(a, \phi))$. $\square$

A.2.2 Belief Concentration in the Myopic Case

We now prove Theorem 1.

Step 1: In the Long Run The Expected Change Approximates the Actual Change in Beliefs.

To characterize the long-run behavior of the agent's beliefs, we show that the change in his beliefs in the long-run can be approximated well by the expected change. To establish this, we use the fact that the log-likelihood function is the average of the log-likelihoods of the realized outputs. As the log-likelihood is an average of non-identical and non-independent random variables, however, we need to generalize existing versions of the law of large numbers to non-i.i.d. random variables. To do so, we use the fact that a square-integrable martingale divided by its quadratic variation, converges to zero whenever the quadratic variation goes to infinity.

The following proposition states a law of large numbers like result for square integrable martingales with bounded quadratic variation. Recall the definition of the quadratic variation of a martingale $(y_t)_t$ as

$$[y]_t = \sum_{s=1}^{t-1} \mathbb{E}[(y_{s+1} - y_s)^2 | \mathcal{F}_s] \,, \tag{11}$$

where $\mathcal{F}_s$ denotes filtration of an outside observer who knows the state at time s, i.e. the expectation is taken with respect to all information available at time s. For brevity, we refer to the martingale $(y_t)_t$ as martingale y.

Proposition 8 (Law of Large Numbers)**.** *Let $(y_t)_t$ be a martingale that satisfies a.s. $[y]_t \leq v\,t$ for some constant $v \geq 0$. We have that a.s.*

$$\lim_{t\to\infty} \frac{y_t}{t} = 0\,.$$

Proof of Proposition 8. We first show that y is square integrable. By the law of iterated expectations, we have that

$$\begin{aligned}\mathbb{E}[y_t^2] &= \mathbb{E}[y_1^2 + \sum_{s=1}^{t-1} y_{s+1}^2 - y_s^2] = y_1^2 + \mathbb{E}\left[\sum_{s=1}^{t-1} \mathbb{E}[y_{s+1}^2 - y_s^2 \mid \mathcal{F}_s]\right] \\ &= y_1^2 + \mathbb{E}\left[\sum_{s=1}^{t-1} \mathbb{E}\left[(y_{s+1} - y_s)^2 \mid \mathcal{F}_s\right]\right] = y_1^2 + \mathbb{E}\left[\,[y]_t\right] \leq y_1^2 + vt\,.\end{aligned}$$

Consequently, the martingale y is square integrable.

In the next step we show that $\lim_{t\to\infty} \frac{y_t}{t} = 0$ almost surely. If the limit of the quadratic variation $[y]_\infty = \lim_{t\to\infty}[y]_t$ exists, the martingale y converges almost surely (Theorem 12.13 in Williams, 1991), and hence $\lim_{t\to\infty} {}^{y_t}/_t = 0$. Hence, from now on we consider the remaining histories for which the quadratic variation does not converge. We can rewrite the average value of the martingale y as

$$\frac{y_t}{t} = \frac{y_t}{[y]_t} \cdot \frac{[y]_t}{t}\,.$$

Because the square quadratic variation is monotone increasing by definition, it must go to infinity after those histories. Furthermore, (as argued for example in Section 12.14 in Williams, 1991), for any square integrable martingale and any history for which $\lim_{t\to\infty}[y]_t = \infty$, one has that almost surely $\lim_{t\to\infty} \frac{y_t}{[y]_t} = 0$. Since $0 \leq [y]_t \leq vt$, it follows that $\limsup_{t\to\infty} \left|\frac{y_t}{t}\right| \leq v \cdot \limsup_{t\to\infty} \left|\frac{y_t}{[y]_t}\right| = 0$ almost surely. Since $\limsup_{t\to\infty} \left|\frac{y_t}{t}\right| = 0$ a.s., $\lim_{t\to\infty} \frac{y_t}{t} = 0$ a.s.. □

To apply the result of Proposition 8 we will, for every fixed belief ϕ, consider the Doob decomposition (Section 12.11 in Williams, 1991) of the derivative of the log-likelihood process into a martingale $(y_t(\phi))_t$ and a previsible process $(z_t(\phi))_t$, henceforth $y(\phi)$ and $z(\phi)$.

Let $m_t(\phi) := Q(e_t, A, \Phi) - Q(e_t, \tilde{a}, \phi)$. We define $x_t(\phi)$ as

$$\begin{aligned} x_t(\phi) &= -\left(g(m_t(\phi) + \epsilon_t) - \mathbb{E}[g(m_t(\phi) + \epsilon_t) \mid \mathcal{F}_t]\right) Q_\phi(e_t, \tilde{a}, \phi) \\ &= -\left(g(m_t(\phi) + \epsilon_t) - \mathbb{E}[g(m_t(\phi) + \epsilon_t) \mid e_t]\right) Q_\phi(e_t, \tilde{a}, \phi) \\ &= -\big(g(m_t(\phi) + \epsilon_t) - \mathcal{L}'(m_t(\phi))\big) Q_\phi(e_t, \tilde{a}, \phi)\,, \end{aligned} \tag{12}$$

where the second equality above uses the fact that an outside observe who knows the true state needs only to condition on the current action e_t to calculate the expected surprise in the next period. Furthermore, we define $y_t(\phi) = \sum_{s=1}^{t} x_s(\phi)$ and $z_t(\phi)$ as

$$z_t(\phi) = -\sum_{s \le t} \mathcal{L}'(m_s(\phi)) Q_\phi(e_s, \tilde{a}, \phi)\,.$$

Lemma 6. *For every ϕ, the processes $(y(\phi), z(\phi))$ have the following properties:*

i) $\ell'_t(\phi) = \ell'_0(\phi) + y_t(\phi) + z_t(\phi)$;

ii) $|x_t(\phi)| \le \overline{\kappa}_\phi |\underline{\kappa}_f|(|\epsilon_t| + \sigma)$; *and*

iii) $y_t(\phi)$ *is a martingale with* $[y(\phi)]_t \le t\,3(\overline{\kappa}_\phi |\underline{\kappa}_f| \sigma)^2$.

Proof of Lemma 6. *i)* follows immediately from the definition. Furthermore, by construction $\mathbb{E}[x_t(\phi)|\mathcal{F}_t] = \mathbb{E}[x_t(\phi) \mid e_t] = 0$ and hence y is a martingale. Using the bound on the absolute value of the derivative of g, $|\underline{\kappa}_f|$, *ii)* follows since:

$$\begin{aligned} |x_t(\phi)| &= |g(m_t(\phi) + \epsilon_t) - \mathbb{E}[g(m_t(\phi) + \epsilon_t) \mid e_t]|\, Q_\phi(e_t, \tilde{a}, \phi) \\ &= |[g(m_t(\phi) + \epsilon_t) - g(m_t(\phi))] - \mathbb{E}[g(m_t(\phi) + \epsilon_t) - g(m_t(\phi)) \mid e_t]|\, Q_\phi(e_t, \tilde{a}, \phi) \\ &\le \overline{\kappa}_\phi\, |[g(m_t(\phi) + \epsilon_t) - g(m_t(\phi))]| + \overline{\kappa}_\phi\, |\mathbb{E}[g(m_t(\phi) + \epsilon_t) - g(m_t(\phi)) \mid e_t]| \\ &\le \overline{\kappa}_\phi |\underline{\kappa}_f|\, |\epsilon_t| + \overline{\kappa}_\phi\, |\underline{\kappa}_f|\, \mathbb{E}[\,|\epsilon_t|\,]\,, \end{aligned}$$

and by Jensen's inequality $\mathbb{E}[\,|\epsilon_t|\,] \le \sqrt{\mathbb{E}[\,|\epsilon_t|^2\,]} = \sigma$. *iii)* follows from *ii)* and the definition of the quadratic variation (11):

$$\begin{aligned} [y(\phi)]_t &= \sum_{s=2}^{t} \mathbb{E}[x_s^2 \mid \mathcal{F}_{s-1}] \le \overline{\kappa}_\phi^2 |\underline{\kappa}_f|^2 \sum_{s=2}^{t} \mathbb{E}[(|\epsilon_s| + \sigma)^2] \\ &= \overline{\kappa}_\phi^2 |\underline{\kappa}_f|^2 \left\{ \sum_{s=2}^{t} \mathbb{E}[|\epsilon_s|^2] + \sigma^2 + \mathbb{E}[\,|\epsilon_s|]\sigma \right\} \le 3\overline{\kappa}_\phi^2 |\underline{\kappa}_f|^2 \sigma^2 t\,. \end{aligned}$$

$\square$

An immediate corollary from the law of large numbers given in Proposition 8 and Lemma 6 *iii*) is that for every ϕ the average of $y_t(\phi)$ converges to zero almost surely, i.e.

$$\lim_{t\to\infty} \frac{y_t(\phi)}{t} = 0\,.$$

We next want to show that this convergence is uniform in ϕ.

Let $\underline{\phi}_\infty$ be the largest number such that the agent almost always convinces herself eventually that the state Φ is greater $\underline{\phi}_\infty$ and $\overline{\phi}_\infty$ the smallest number such that the agent is eventually convince that the state is below $\overline{\phi}_\infty$, i.e.

$$\underline{\phi}_\infty = \sup\{\phi' \colon \lim_{t\to\infty} \Pi_t(\phi') = 0 \text{ almost surely}\},$$
$$\overline{\phi}_\infty = \inf\{\phi' \colon \lim_{t\to\infty} \Pi_t(\phi') = 1 \text{ almost surely}\}.$$

Definition 3 (Uniform Stochastic Convergence)**.** The sequence $(G_t(\cdot))_t$ converges uniformly over $[\underline{\phi}_\infty, \overline{\phi}_\infty]$ stochastically to zero if and only if

$$\lim_{t\to\infty} \sup_{\phi\in[\underline{\phi}_\infty, \overline{\phi}_\infty]} |G_t(\phi)| = 0 \text{ a.s. .} \tag{U-SCON}$$

For deterministic sequences of real valued function, Ascoli's theorem states that pointwise convergence implies uniform convergence if and only if the functions are equicontinuous. Despite the fact that our sequence of real valued functions is not equicontiniuous for every realization, we use a stochastic analogue—strong stochastic equicontinuity (SSE) as defined on page 245 in Andrews (1992)—to establish uniform convergence below.

Definition 4 (Strong Stochastic Equicontinuity)**.** Let $G^*_t(\phi) = \sup_{s\geq t} |G_s(\phi)|$. A sequence $(G_t(\cdot))_t$ is strongly stochastic equicontinuous if and only if:

i) $G^*_1(\phi) < \infty$ for all $\phi \in [\underline{\phi}_\infty, \overline{\phi}_\infty]$ a.s.; and

ii) for all $\gamma > 0$ there exists $\delta > 0$ such that

$$\lim_{t\to\infty} \mathbb{P}\left[\sup_{\phi\in[\underline{\phi}_\infty, \overline{\phi}_\infty]} \sup_{\phi'\in B_\delta(\phi)} |G^*_t(\phi) - G^*_t(\phi')| > \gamma\right] < \gamma\,.$$

The usefulness of SSE comes from the fact that almost sure convergence in combination with SSE implies U-SCON.

Theorem 3 (Theorem 2 (a) in Andrews (1992))**.** *If G satisfies strong stochastic equicontinuity and G converges pointwise to zero a.s. than G converges uniform stochastically to zero.*

Our next result argues that the sequence ${}^{y_t(\cdot)}/_t$ is strong stochastic equicontinuous and thus converges uniform stochastically to zero.

Lemma 7. $\left(\frac{y_t(\cdot)}{t}\right)_t$ *converges uniform stochastically to zero.*

Proof of Lemma 7. By Lemma 2 in Andrews (1992) strong stochastic equicontinuity of $\left(\frac{y_t(\cdot)}{t}\right)_t$ follows if the absolute value of the derivative of $x_t(\cdot)$ can be uniformly bounded by a random variable B_t for all $\phi \in [\underline{\phi}_\infty, \overline{\phi}_\infty]$ such that $\sup_{t\geq 1} \frac{1}{t}\sum_{s=1}^t \mathbb{E}[B_s] < \infty$ and $\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^t (B_s - \mathbb{E}[B_s]) = 0$.

From the definition of x in Eq. 12 it follows that

$$\begin{aligned}
|x_t'(\phi)| &= \Big| \left(g'(m_t(\phi)+\epsilon_t) - \mathcal{L}''(m_t(\phi))\right)[Q_\phi(e_t,\tilde{a},\phi)]^2 \\
&\qquad\qquad - \left(g(m_t(\phi)+\epsilon_t) - \mathcal{L}'(m_t(\phi))\right) Q_{\phi\phi}(e_t,\tilde{a},\phi)\Big| \\
&\leq \Big|g'(m_t(\phi)+\epsilon_t) - \mathcal{L}''(m_t(\phi))\Big|[Q_\phi(e_t,\tilde{a},\phi)]^2 \\
&\qquad\qquad + \Big| \left(g(m_t(\phi)+\epsilon_t) - \mathcal{L}'(m_t(\phi))\right)\Big| \times \Big|Q_{\phi\phi}(e_t,\tilde{a},\phi)\Big|
\end{aligned}$$

Since $|g'| \leq |\underline{\kappa}_f|$ one has $|\mathcal{L}''(m_t(\phi))| \leq |\underline{\kappa}_f|$, and by Assumption 2 $Q_\phi \leq \overline{\kappa}_\phi$, $|Q_{\phi\phi}| \leq \overline{\kappa}_{\phi\phi}$; using this and substituting $x_t(\phi)$ yields

$$|x_t'(\phi)| \leq 2|\underline{\kappa}_f|\overline{\kappa}_\phi^2 + |x_t(\phi)|\frac{\overline{\kappa}_{\phi\phi}}{\underline{\kappa}_\phi} \leq 2|\underline{\kappa}_f|\overline{\kappa}_\phi^2 + \frac{\overline{\kappa}_{\phi\phi}}{\underline{\kappa}_\phi}\overline{\kappa}_\phi|\underline{\kappa}_f|(|\epsilon_t|+\sigma) =: B_t,$$

where the last inequality follows from Lemma 6 *ii*). As $(B_t)_t$ are i.i.d. it follows that

$$\frac{1}{t}\sum_{s=1}^t \mathbb{E}[B_s] = \mathbb{E}[B_0] < \infty.$$

Furthermore, as B_t also has finite variance the law of large numbers implies $\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^t (B_s - \mathbb{E}[B_s]) = 0$. □

Step 2: No Long-Run Surprises

The next lemma shows that if the agent is always on average surprised by the output for some beliefs—i.e. $m_t(\cdot)$ is bounded away from zero—than the absolute value of the derivative of his subjective log-likelihood goes to infinity almost surely for those beliefs.

Let I denote an interval.

Lemma 8. *(a) If* $\liminf_{t\to\infty} m_t(\phi) \geq \underline{m} > 0$ *for all* $\phi \in I$ *then there exists* $r > 0$ *such that a.s.*

$$\liminf_{t\to\infty} \inf_{\phi\in I} \frac{\ell_t'(\phi)}{t} \geq r\,.$$

(b) If $\limsup_{t\to\infty} m_t(\phi) \leq \overline{m} < 0$ *for all* $\phi \in I$ *then there exists* $r > 0$ *such that a.s.*

$$\limsup_{t\to\infty} \sup_{\phi\in I} \frac{\ell_t'(\phi)}{t} \leq -r\,.$$

Proof of Lemma 8. We show (a) the proof of (b) is analogous. As $\mathcal{L}'$ is decreasing, we have that for all $\phi \in I$

$$\mathcal{L}'(m_t(\phi)) \leq \mathcal{L}'(\underline{m})\,.$$

We use this fact and that $\mathcal{L}'(m) < 0$ to bound $z_t(\phi)$ for all $\phi \in I$,

$$z_t(\phi) = -\sum_{s=1}^{t} \mathcal{L}'(m_t(\phi)) Q_\phi(e_t, \tilde{a}, \phi) \geq \sum_{s=1}^{t} |\mathcal{L}'(\underline{m})| Q_\phi(e_t, \tilde{a}, \phi) \geq t \cdot |\mathcal{L}'(\underline{m})| \underline{\kappa}_\phi\,.$$

Define $r = |\mathcal{L}'(\underline{m})|\underline{\kappa}_\phi > 0$. Using the definition of $y_t(\phi)$ and $z_t(\phi)$, r, and the uniform stochastic convergence of y_t/t to zero that we established in Lemma 7, respectively, we have that

$$\begin{aligned}\liminf_{t\to\infty} \inf_{\phi\in I} \frac{\ell_t'(\phi)}{t} &= \liminf_{t\to\infty} \inf_{\phi\in I} \frac{\ell_0'(\phi) + y_t(\phi) + z_t(\phi)}{t} \geq \left[\liminf_{t\to\infty} \inf_{\phi\in I} \frac{y_t(\phi)}{t}\right] + r \\ &\geq -\left[\limsup_{t\to\infty} \inf_{\phi\in I} \frac{|y_t(\phi)|}{t}\right] + r = r\,.\end{aligned}$$

□

The next lemma argues that if the agent is surprised by the output for an interval of beliefs than he will a.s. assign probability zero to those beliefs in the long-run. Intuitively, as by Lemma 8 the absolute value of the derivative of the agent's subjective likelihood goes to infinity, the absolute value of the derivative of his posterior density goes to infinity. The next lemma shows that this implies that the agent must assign probability zero to those beliefs.

Lemma 9. *i) If* $\liminf_{t\to\infty} m_t(\phi) \geq \underline{m} > 0$ *for all* $\phi \in (l,h) \subset (\underline{\phi}, \overline{\phi})$ *then*

$$\lim_{t\to\infty} \tilde{\mathbb{P}}_t[\Phi \in [l,h)\,] = 0\,.$$

ii) If $\limsup_{t\to\infty} m_t(\phi) \leq \overline{m} < 0$ *for all* $\phi \in (l,h) \subset (\underline{\phi}, \overline{\phi})$ *then*

$$\lim_{t\to\infty} \tilde{\mathbb{P}}_t[\Phi \in (l,h]\,] = 0\,.$$

Proof of Lemma 9. First consider the case where $m_t(\phi) \geq \underline{m} > 0$. Lemma 8 implies that there a.s. exists $r > 0$ such that for sufficiently large t for all $y \in (l,h)$

$$\ell_t'(y) \geq rt\,.$$

Let $\eta = h - l$. We have that the probability the agent assigns to state in $[l+\eta/2, l+\eta]$ satisfies

$$\begin{aligned}
\tilde{\mathbb{P}}_t\left[\Phi \in [l+\eta/2, l+\eta]\right] &= \int_{l+\eta/2}^{l+\eta} \pi_t(z)\mathrm{d}z = \int_{l}^{l+\eta/2} \pi_t(z)\frac{\pi_t(z+\eta/2)}{\pi_t(z)}\mathrm{d}z \\
&= \int_{l}^{l+\eta/2} \pi_t(z)\exp(\ell_t(z+\eta/2) - \ell_t(z))\mathrm{d}z \\
&= \int_{l}^{l+\eta/2} \pi_t(z)\exp\left(\int_{z}^{z+\eta/2} \ell_t'(y)\mathrm{d}y\right)\mathrm{d}z \\
&\geq \int_{l}^{l+\eta/2} \pi_t(z)\exp\left(r\,t\eta/2\right)\mathrm{d}z = e^{r\,t\eta/2}\tilde{\mathbb{P}}_t\left[\Phi \in [l, l+\eta/2]\right]\,.
\end{aligned}$$

As $r > 0$, we have that the probability assigned to the interval $[l, l+\eta/2]$ is bounded by a term that vanishes for $t \to \infty$,

$$\tilde{\mathbb{P}}_t\left[\Phi \in [l, l+\eta/2]\right] \leq e^{-rt\eta/2}\tilde{\mathbb{P}}_t\left[\Phi \in [l+\eta/2, l+\eta]\right] \leq e^{-rt\eta/2}\,.$$

Hence, the agent must assigned zero probability to the interval $[l, l+\eta/2]$ in the long-run.

Applying this argument iteratively yields that the agent assigns zero probability to the interval $[l,h)$. The argument in the case where $m_t(\phi) \leq \overline{m} < 0$ is completely analogous. □

Step 3: Convergence in the Correctly Specified Case

Lemma 9 allows us to argue that the subjective agent's beliefs converge when her model of the world is correctly specified $\tilde{a} = A$.

Proposition 9. *If* $\tilde{a} = A$ *then the agent's posterior belief converges to a Dirac measure on the true state* Φ*.*

Proof of Proposition 9. Fix a $\eta > 0$. We have that for all $\phi \leq \Phi - \eta$,

$$m_t(\phi) = Q(e_t, A, \Phi) - Q(e_t, \tilde{a}, \phi) \geq \eta \underline{\kappa}_\phi \,.$$

Hence, by Lemma 9, we have that a.s. $\lim_{t\to\infty} \Pi_t(\Phi - \eta) = 0$ and thus a.s. $\underline{\phi}_\infty \geq \Phi - \eta$. Taking the supremum over η yields that a.s. $\underline{\phi}_\infty \geq \Phi$. An analogous argument yields that a.s. $\overline{\phi}_\infty \leq \Phi$. Since $\underline{\phi}_\infty \leq \overline{\phi}_\infty$, we conclude that a.s. $\overline{\phi}_\infty = \underline{\phi}_\infty = \Phi$. □

Step 4: Long-run Bounds on the Agent's Beliefs.

Building on Lemma 2, in this step we show that the agent's subjective belief is bounded in the long-run. The next Lemma shows that $\underline{\phi}_\infty$ and $\overline{\phi}_\infty$ are well defined and that the agent's long-run beliefs are almost surely bounded.

Lemma 10. *We have that $\Phi - {}^{\overline{\kappa}_a}/_{\underline{\kappa}_\phi}(\tilde{a} - A) \leq \underline{\phi}_\infty$ and $\overline{\phi}_\infty \leq \Phi$.*

Proof of Lemma 10. We first show $\overline{\phi}_\infty \leq \Phi$. Note that as the average output increases in ability, it follows that for every sequence of actions the output that the agent observes is smaller than

$$q_t \leq \hat{q}_t := q_t + Q(e_t, \tilde{a}, \Phi) - Q(e_t, A, \Phi) \,.$$

By construction if the agent where to observe the outputs $(\hat{q}_t)$ instead of (q_t) he would have the same belief as a correctly specified decision maker with ability $\tilde{a}$ would have if the state equals Φ. The beliefs of such a correctly specified decision maker converge to Φ almost surely for every sequence of actions by Proposition 9.

By Lemma 2, $q_t \leq \hat{q}_t$ implies that the agent's posterior belief is lower in the sense of MLR than for a sequence of beliefs that converges to Φ almost surely. As MLR implies first order stochastic dominance it follows that $\overline{\phi}_\infty \leq \Phi$.

We next show $\underline{\phi}_\infty \geq \Phi - \frac{\overline{\kappa}_a}{\underline{\kappa}_\phi}(\tilde{a} - A)$. Let $\Phi' = \Phi - \frac{\overline{\kappa}_a}{\underline{\kappa}_\phi}(\tilde{a} - A)$. We will show that

$$q_t \geq \hat{q}_t := q_t + Q(e_t, \tilde{a}, \Phi') - Q(e_t, A, \Phi) \,.$$

We have that

$$\begin{aligned} Q(e_t, \tilde{a}, \Phi') - Q(e_, A, \Phi) &= Q(e_t, \tilde{a}, \Phi') - Q(e_t, \tilde{a}, \Phi) + Q(e_t, \tilde{a}, \Phi) - Q(e_t, A, \Phi) \\ &\leq -\underline{\kappa}_\phi(\Phi - \Phi') + \overline{\kappa}_a(\tilde{a} - A) = 0. \end{aligned}$$

By construction if the agent where to observe the outputs $(\hat{q}_t)$ instead of (q_t) he would have the same belief as a correctly specified decision maker with ability $\tilde{a}$ would have if the state equals Φ', and hence the result follows from the same argument as above. $\square$

Step 5: Bounds on the Myopically Optimal Actions.

Let e_t^m be the action that is myopically optimal in period t

$$e_t^m = \arg\max_e \tilde{\mathbb{E}}_{t-1}[q_t] = \arg\max_e \int_{(\underline{\phi},\overline{\phi})} Q(e, \tilde{a}, \phi)\pi_{t-1}(\phi)\mathrm{d}\phi \,.$$

We define the long-run lower and upper bound on the agent's actions

$$\check{e} = \liminf_{t\to\infty} e_t^m$$
$$\hat{e} = \limsup_{t\to\infty} e_t^m \,.$$

The next lemma shows that if the agent assigns subjective probability of almost one to the event that the state is strictly greater (smaller) than some $\underline{\phi}_\infty$ ($\overline{\phi}_\infty$) than the myopically optimal action is greater than the optimal action if the agent assigns probability one to the state $\underline{\phi}_\infty$ ($\overline{\phi}_\infty$). Recall that $e^*(\phi)$ denotes the optimal action when the agent has point beliefs on ϕ, and that $e^*(\phi)$ is increasing.

Lemma 11. *If the agent is myopic $e_t = e_t^m$ then the long run bounds on his actions satisfy*

$$e^*(\underline{\phi}_\infty) \le \check{e} \le \hat{e} \le e^*(\overline{\phi}_\infty) \,.$$

Proof of Lemma 11. Because Q is strictly concave with positive derivative at $\underline{e}$ and negative derivative at $\overline{e}$, the agent's myopically optimal action is characterized by the first order condition $\tilde{\mathbb{E}}_{t-1}[Q_e(e_t, \tilde{a}, \Phi)] = 0$.

Let $\phi' = \underline{\phi}_\infty - 2\gamma$ for some $\gamma > 0$. To show that the myopically optimal action e_t is greater $e' = e^*(\phi')$ for large t it suffices to show that the expected marginal output is positive at $e' < e^*(\underline{\phi}_\infty)$

$$\begin{aligned}\tilde{\mathbb{E}}_{t-1}[Q_e(e', \tilde{a}, \Phi)] &= \int_{(\underline{\phi},\overline{\phi})} Q_e(e', \tilde{a}, \phi)\pi_{t-1}(\phi)\mathrm{d}\phi \\ &= \int_{\phi'+\gamma}^{\overline{\phi}} Q_e(e', \tilde{a}, \phi)\pi_{t-1}(\phi)\mathrm{d}\phi + \int_{\underline{\phi}}^{\phi'+\gamma} Q_e(e', \tilde{a}, \phi)\pi_{t-1}(\phi)\mathrm{d}\phi.\end{aligned}$$

Recall that by Assumption 2 the derivative with respect to the action is bounded $|Q_e| \leq \kappa_e$ and that $Q_{e\phi} > 0$, and hence the above is greater than or equal to

$$\begin{aligned}
&\int_{\phi'+\gamma}^{\overline{\phi}} Q_e(e', \tilde{a}, \phi' + \gamma)\pi_{t-1}(\phi)\mathrm{d}\phi - \kappa_e \int_{\underline{\phi}}^{\phi'+\gamma} \pi_{t-1}(\phi)\mathrm{d}\phi \\
&= Q_e(e', \tilde{a}, \phi' + \gamma)\, [1 - \Pi_{t-1}(\phi' + \gamma)] - \kappa_e\, \Pi_{t-1}(\phi' + \gamma).
\end{aligned}$$

Furthermore, since by definition of $e' = e^*(\phi') \leq e^*(\phi'+\gamma)$, we have $Q_e(e', \tilde{a}, \phi'+\gamma) > 0$. As $\phi'+\gamma = \underline{\phi}_\infty - \gamma$, it follows from the definition of $\underline{\phi}_\infty$, which exists by Lemma 10, that $\lim_{t\to\infty} \Pi_{t-1}(\phi'+\gamma) = 0$. Hence, taking the limit $t \to \infty$ yields that $\tilde{\mathbb{E}}_{t-1}[Q_e(e', \tilde{a}, \Phi)] > 0$ for t large enough. Consequently for all $\gamma > 0$, the myopically optimal action is greater than $e' = e^*(\underline{\phi}_\infty - 2\gamma)$ for sufficiently large t. Taking the supremum over γ yields the result

$$\check{e} \geq \sup_{\gamma>0} e^*(\underline{\phi}_\infty - 2\gamma) = e^*(\underline{\phi}_\infty).$$

The proof for the upper bound $\hat{e}$ is analogous. $\square$

Step 6. Beliefs Converge to Limiting Belief.

We begin by showing that the bounds on the myopically optimal action imply bounds on the long run average output, and then complete the proof of Theorem 1 by showing that beliefs a.s. converge.

The next lemma is useful for arguing that if the agent takes an action above the action that is optimal for the state $\underline{\phi}_\infty < \Phi$ then the realized average output will be strictly greater than the average output that the agent expects if the state is $\underline{\phi}_\infty$.

Lemma 12. *The long-run average surprise in output satisfies*

$$\liminf_{t\to\infty} Q(e_t, A, \Phi) - Q(e_t, \tilde{a}, \underline{\phi}_\infty) \geq \Gamma(\underline{\phi}_\infty), \text{ and} \tag{13}$$

$$\limsup_{t\to\infty} Q(e_t, A, \Phi) - Q(e_t, \tilde{a}, \overline{\phi}_\infty) \leq \Gamma(\overline{\phi}_\infty). \tag{14}$$

Proof of Lemma 12. Let $e' = e^*(\phi')$ for some ϕ'. We have that

$$\begin{aligned}
Q(e_t, A, \Phi) - Q(e_t, \tilde{a}, \phi') - \Gamma(\phi') &= [Q(e_t, A, \Phi) - Q(e_t, \tilde{a}, \phi')] - [Q(e', A, \Phi) - Q(e', \tilde{a}, \phi')] \\
&= -[Q(e', A, \Phi) - Q(e_t, A, \Phi)] + [Q(e', \tilde{a}, \phi') - Q(e_t, \tilde{a}, \phi')] \\
&= \int_{e_t}^{e'} Q_e(z, \tilde{a}, \phi') - Q_e(z, A, \Phi)\mathrm{d}z. \qquad (15)
\end{aligned}$$

We first establish that (13) holds. To show this, we first show that (15) is non-negative for $\phi' = \underline{\phi}_\infty$ and $e_t \geq \check{e}$. For $e' = e^*(\underline{\phi}_\infty)$, Lemma 11 implies that $e^*(\underline{\phi}_\infty) \leq \check{e}$, and hence the term (15) equals

$$\int_{e_t}^{e^*(\underline{\phi}_\infty)} Q_e(z, \tilde{a}, \underline{\phi}_\infty) - Q_e(z, A, \Phi)\mathrm{d}z = \int_{e^*(\underline{\phi}_\infty)}^{e_t} Q_e(z, A, \Phi) - Q_e(z, \tilde{a}, \underline{\phi}_\infty)\mathrm{d}z \geq 0\,,$$

where the last inequality follows from the facts that: $Q_{e\phi} > 0$ and $\underline{\phi}_\infty \leq \Phi$ by Lemma 10, and $Q_{ea} \leq 0$ and $\tilde{a} > A$. That (13) holds follows as Q is continuous in e and $\liminf_{t\to\infty} e_t \geq \underline{e} \geq e^*(\underline{\phi}_\infty)$.

We finally show that (14) holds. To show this, we first show that (15) is non-positive for $\phi' = \overline{\phi}_\infty$ and $e_t \leq \hat{e}$. In this case $e' = e^*(\overline{\phi}_\infty)$, and Lemma 11 implies $e^*(\overline{\phi}_\infty) \geq \hat{e}$. Hence, the term (15) equals

$$\int_{e_t}^{e^*(\overline{\phi}_\infty)} Q_e(z, \tilde{a}, \overline{\phi}_\infty) - Q_e(z, A, \Phi)\mathrm{d}z \leq 0\,,$$

where the last inequality follows from the facts that: $Q_{e\phi} > 0$ and $\overline{\phi}_\infty \leq \Phi$ by Lemma 10, and $Q_{ea} \leq 0$ and $\tilde{a} > A$. That (14) holds follows as Q is continuous in e and $\limsup_{t\to\infty} e_t \geq \hat{e} \geq e^*(\overline{\phi}_\infty)$. □

Lemma 12 shows that if $\Gamma(\underline{\phi}_\infty) > 0$ and $\Gamma(\overline{\phi}_\infty) < 0$ the output will be on average higher than the output he would expect at the state $\underline{\phi}_\infty$ and lower than the output he would expect at the state $\overline{\phi}_\infty$. Intuitively, this should lead the agent to assign probability zero to states around $\underline{\phi}_\infty$ and $\overline{\phi}_\infty$, which contradicts the definition of $[\underline{\phi}_\infty, \overline{\phi}_\infty]$ as the smallest interval to which the agent assigns probability one in the long-run and hence implies that $\Gamma(\underline{\phi}_\infty) = \Gamma(\overline{\phi}_\infty) = 0$. We use the next lemma to formalize this intuition.

The next Lemma shows that the condition of Lemma 9 is satisfied whenever the surprise function is positive at $\underline{\phi}_\infty$ or negative at $\overline{\phi}_\infty$.

Lemma 13. *(a) If $\Gamma(\underline{\phi}_\infty) > 0$ then there exists $\beta, \underline{m} > 0$ such that a.s. for all $\phi \in [\underline{\phi}_\infty, \underline{\phi}_\infty + \beta]$,*

$$\liminf_{t\to\infty} m_t(\phi) \geq \underline{m}\,.$$

.

(b) If $\Gamma(\overline{\phi}_\infty) < 0$ then there exists $\beta, \overline{m} < 0$ such that a.s. for all $\phi \in [\overline{\phi}_\infty - \beta, \overline{\phi}_\infty]$,

$$\limsup_{t\to\infty} m_t(\phi) \leq \overline{m}\,.$$

Proof of Lemma 13. We show (a) the proof of (b) is analogous. Lemma 12 implies that almost surely

$$\liminf_{t\to\infty} m_t(\underline{\phi}_\infty) = \liminf_{t\to\infty} Q(e_t, A, \Phi) - Q(e_t, \tilde{a}, \underline{\phi}_\infty) \geq \Gamma(\underline{\phi}_\infty) > 0\,.$$

As $0 < Q_\phi < \overline{\kappa}_\phi$ it follows that $m_t(\phi) \geq m_t(\underline{\phi}_\infty) - \overline{\kappa}_\phi(\phi - \underline{\phi}_\infty) \geq \Gamma(\underline{\phi}_\infty) - \overline{\kappa}_\phi(\phi - \underline{\phi}_\infty)$, and hence that

$$m_t(\phi) \geq {}^{1}\!/_{2}\,\Gamma(\underline{\phi}_\infty)$$

for all $\phi \in [\underline{\phi}_\infty, \underline{\phi}_\infty + \beta]$ with $\beta = {}^{\Gamma(\underline{\phi}_\infty)}\!/_{2\overline{\kappa}_\phi}$. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. We first show that $\Gamma(\underline{\phi}_\infty) \leq 0$. Suppose for the sake of a contradiction that $\Gamma(\underline{\phi}_\infty) > 0$. By Lemma 13 there exists a $\beta, \underline{m} > 0$ such that $\liminf_{t\to\infty} m_t(\phi) \geq \underline{m} > 0$ for all $\phi \in [\underline{\phi}_\infty, \underline{\phi}_\infty + \beta]$. By Lemma 9 almost surely

$$\lim_{t\to\infty} \tilde{\mathbb{P}}_t[\Phi \in [\underline{\phi}_\infty, \underline{\phi}_\infty + \beta)] = 0\,.$$

Hence, the agent assigns zero probability to the interval $[\underline{\phi}_\infty, \underline{\phi}_\infty + \beta)$ in the long-run, which contradicts the definition of $\underline{\phi}_\infty$. Consequently, $\Gamma(\underline{\phi}_\infty) \leq 0$.

An analogous argument yields that $\Gamma(\overline{\phi}_\infty) \geq 0$. Since Γ crosses zero from above by Lemma 5, it follows that $\underline{\phi}_\infty \geq \phi_\infty$ and $\overline{\phi}_\infty \leq \phi_\infty$. Since $\underline{\phi}_\infty \leq \overline{\phi}_\infty$ by definition, $\underline{\phi}_\infty = \phi_\infty = \overline{\phi}_\infty$. By definition of $\underline{\phi}_\infty$ and $\overline{\phi}_\infty$, this implies the statement of the theorem. □

A.2.3 Belief Concentration in the Linear Non-Myopic Case

Next, we argue that beliefs concentrate and converge in distribution to the root of the surprise function Γ even when the agent is non-myopic, i.e. experiments, if Q is linear in ϕ. If Q is linear in ϕ there exist functions G, H such that

$$Q(e, a, \phi) = \phi\, H(e, a) + G(e, a)\,.$$

In order to satisfy the Assumption that $Q_\phi > 0$, we need to assume that Φ does not change sign. We will thus henceforth consider the case where subjectively as well as objectively $\Phi > 0$. As $Q_\phi \geq \underline{\kappa}_\phi > 0$ by Assumption 2, it follows that $H(e, a) \geq \underline{\kappa}_\phi > 0$. We impose this linear structure until the end of the proof of Theorem 2.

Lemma 14. *There exists constants* $\underline{\kappa}_\ell \leq \overline{\kappa}_\ell < 0$ *and* $\tau > 0$ *such that for every sequence of signals* $(q_s)_{s\leq t}$ *and actions* $(e_s)_{s\leq t}$ *and all* ϕ

$$\underline{\kappa}_\ell \cdot t \leq \ell_t''(\phi) \leq \overline{\kappa}_\ell \cdot t \ \ \textit{for } t \geq \tau.$$

Proof of Lemma 14. Note that $g' \in [\underline{\kappa}_f, \overline{\kappa}_f]$ by our assumption of bounded log-concavity and Assumption 3. The first and the second derivative of the log-likelihood function is given by

$$\begin{aligned}
\ell_t'(\phi) &= \sum_{s\leq t} -g\big(q_s - Q(e_s, \tilde{a}, \phi)\big) H(e_s, \tilde{a}) + \frac{\partial}{\partial \phi} \log \pi_0(\phi)\,, \\
\ell_t''(\phi) &= \sum_{s\leq t} g'\big(q_s - Q(e_s, \tilde{a}, \phi)\big) H^2(e_s, \tilde{a}) + \frac{\partial^2}{\partial \phi^2} \log \pi_0(\phi) \\
&\leq t\, \overline{\kappa}_f\, \underline{\kappa}_\phi^2 + \overline{\kappa}_\pi\,.
\end{aligned}$$

Observe that for large enough t, $t\, \overline{\kappa}_f\, \underline{\kappa}_\phi^2 + \overline{\kappa}_\pi < 0$, which establishes that there exists a $\overline{\kappa}_\ell < 0$ such that $\ell_t''(\phi) \leq \overline{\kappa}_\ell \cdot t < 0$ for large enough t.

By essentially the same argument $\ell_t''(\phi) \geq t \cdot \underline{\kappa}_f\, \overline{\kappa}_\phi^2 + \underline{\kappa}_\pi$, and hence there exists a $\underline{\kappa}_\ell < 0$ such that $\ell_t''(\phi) \geq \underline{\kappa}_\ell \cdot t$ for large enough t. $\square$

Since $\overline{\kappa}_f < 0$, for large enough t the agent's posterior log-likelihood is strictly concave for every sequence of signals, and hence there exists a unique log-likelihood maximizer (or modal belief) of the agent when t is large enough:

$$\phi_t^{ML} := \arg\max_\phi \ell_t(\phi)\,.$$

Proposition 10 (Concentration)**.** *There exists a constant* k *such that for all large enough* t

$$\tilde{\mathbb{E}}_t[(\phi - \phi_t^{ML})^2] \leq k\frac{1}{t}\,. \tag{16}$$

Proof of Proposition 10. We consider large enough t such that the agent's posterior log-likelihood is strictly concave for every sequence of signals, and hence the log-likelihood maximizer ϕ_t^{ML} is unique. Furthermore, because $\pi_0(\underline{\phi}) = 0$, we have that $\ell_t(\underline{\phi}) = -\infty$. Thus, the maximizer ϕ_t^{ML} is interior. Since ℓ_t is strictly concave and twice differentiable, ϕ_t^{ML} is implicitly defined by

$$0 = \ell_t'(\phi_t^{ML})\,.$$

The loss in log-likelihood relative to this maximizer is bounded from below by the squared distance from the maximizer

$$\ell_t(\phi_t^{ML}) - \ell_t(\phi) = \int_\phi^{\phi_t^{ML}} \ell_t'(z) - \ell_t'(\phi_t^{ML})\mathrm{d}z = -\int_\phi^{\phi_t^{ML}}\int_z^{\phi_t^{ML}} \ell_t''(y)\mathrm{d}y\,\mathrm{d}z \geq t\,|\overline{\kappa}_\ell|\frac{1}{2}\left(\phi - \phi_t^{ML}\right)^2 .$$

By an analogous argument $\ell_t(\phi_t^{ML}) - \ell_t(\phi) \leq t\,|\underline{\kappa}_\ell|\,\frac{1}{2}\left(\phi - \phi_t^{ML}\right)^2$. The expected distance of the true state from the log-likelihood maximizer is given by

$$\begin{aligned}
\tilde{\mathbb{E}}_t\left[\left(\phi - \phi_t^{ML}\right)^2\right] &= \int_{(\underline{\phi},\overline{\phi})} \left(\phi_t^{ML} - \phi\right)^2 \frac{e^{\ell_t(\phi)}}{\int_{(\underline{\phi},\overline{\phi})} e^{\ell_t(z)}\mathrm{d}z}\mathrm{d}\phi \\
&= \int_{(\underline{\phi},\overline{\phi})} \left(\phi_t^{ML} - \phi\right)^2 \frac{e^{-[\ell_t(\phi_t^{ML})-\ell_t(\phi)]}}{\int_{(\underline{\phi},\overline{\phi})} e^{-[\ell_t(\phi_t^{ML})-\ell_t(z)]}\mathrm{d}z}\mathrm{d}\phi \\
&\leq \frac{|\underline{\kappa}_\ell|}{|\overline{\kappa}_\ell|}\frac{\int_{(\underline{\phi},\overline{\phi})} \frac{t|\overline{\kappa}_\ell|}{\sqrt{2\pi}}\left(\phi - \phi_t^{ML}\right)^2 e^{-t|\overline{\kappa}_\ell|\frac{1}{2}\left(\phi_t^{ML}-\phi\right)^2}\mathrm{d}\phi}{\int_{(\underline{\phi},\overline{\phi})} \frac{t|\underline{\kappa}_\ell|}{\sqrt{2\pi}} e^{-t|\underline{\kappa}_\ell|\frac{1}{2}\left(\phi_t^{ML}-\phi\right)^2}\mathrm{d}z} \\
&= \frac{|\underline{\kappa}_\ell|}{|\overline{\kappa}_\ell|^2}\cdot\frac{1}{t} .
\end{aligned}$$

In the last step we use that the term above the numerator is the variance of a normal distribution with variance $\frac{1}{|\overline{\kappa}_\ell|\,t}$, and the term in the denominator is the integral over a normal density (with variance $\frac{1}{|\underline{\kappa}_\ell|\,t}$) and hence equal to one . □

As a consequence for large enough t, the agent's posterior expected squared distance between the state and the log-likelihood maximizer decays at the speed of $1/t$ for any sequence of signals he observes.

Define $\tilde{\phi}_t$ as the agent's subjective posterior mean

$$\tilde{\phi}_t := \tilde{\mathbb{E}}_t[\phi] .$$

The result of Proposition 10 immediately implies that the agent's subjective beliefs also concentrate around his posterior mean.

Lemma 15. *There exists a constant k such that for all large enough t*

$$\tilde{\mathbb{E}}_t[(\phi - \tilde{\phi}_t)^2] \leq k\frac{1}{t} . \tag{17}$$

Proof of Lemma 15. The agent's subjective posterior mean minimizes the squared distance from the agent's point of view, i.e. for any $\hat{\phi}$

$$0 = \frac{\partial}{\partial\hat{\phi}}\tilde{\mathbb{E}}_t[(\hat{\phi} - \phi)^2] = 2\tilde{\mathbb{E}}_t[\hat{\phi} - \phi] = 2(\hat{\phi} - \tilde{\phi}_t) .$$

Hence, the posterior variance must be less than the expected distance between the state and the maximum likelihood estimate

$$\tilde{\mathbb{E}}_t\left[\left(\phi-\tilde{\phi}_t\right)^2\right] \le \tilde{\mathbb{E}}_t\left[\left(\phi-\phi_t^{ML}\right)^2\right] \le \frac{k}{t}\,.$$

□

Denote by e_t^m the action that is *myopically* optimal given the agent's posterior belief

$$e_t^m \in \arg\max_e \tilde{\mathbb{E}}_{t-1}[Q(e,\tilde{a},\phi)]\,.$$

Recall that we denote by $e^*(\hat{\phi})$ the action that is subjectively optimal when the agent assigns probability one to some state $\hat{\phi}$. As the output function is linear in ϕ, the myopically optimal action is implicitly given by the first order condition

$$0=\tilde{\mathbb{E}}_{t-1}[\phi]\,H_e(e,\tilde{a})+G_e(e,\tilde{a})\,.$$

This immediately implies the following:

Lemma 16. *The myopically optimal action e_t^m equals the optimal action when the agent assigns probability one to the state $\tilde{\phi}_t$*

$$e_t^m=e^*(\tilde{\phi}_{t-1})\,.$$

In the next step, we show that the change in the optimal action is locally Lipschitz continuous in the subjective average belief.

Lemma 17. *For every campact interval I there exists k_I such that*

$$0\le (e^*)'(\phi)\le k_I\,.$$

Proof of Lemma 17. As Q is concave in the action e, the optimal action $e^*(\phi)$ when the agent assigns a point belief to ϕ satisfies $0=Q_e(e(\phi),\tilde{a},\phi)$. By the implicit function theorem

$$(e^*)'(\phi)=-\frac{Q_{e\phi}(e^*(\phi),\tilde{a},\phi)}{Q_{ee}(e^*(\phi),\tilde{a},\phi)}>0\,.$$

As $(e^*)'$ is continuous it follows that it is bounded on I by

$$k_I=\max_{\phi\in I}\frac{Q_{e\phi}(e^*(\phi),\tilde{a},\phi)}{|Q_{ee}(e^*(\phi),\tilde{a},\phi)|}.$$

□

In the next step, we show that the agent's gain from learning vanishes as t increases, and hence the optimal action approaches the myopically optimal one. An easy upper bound on the gain from learning is the change in payoffs when the agent uses the myopically optimal payoff.

Lemma 18. *As $t \to \infty$, the optimal action e_t and myopically optimal action e_t^m converges, i.e*

$$\lim_{t\to\infty} (e_t^m - e_t)^2 = 0$$

.

Proof of Lemma 18. Fix an interval of beliefs $I_\phi = [\underline{\phi}_\infty - \gamma, \overline{\phi}_\infty + \gamma]$ for some $\gamma > 0$ and a corresponding set of actions $I_e = [e^*(\underline{\phi}_\infty - \gamma), e^*(\overline{\phi}_\infty + \gamma)]$. Let κ_{ee} be given by

$$\kappa_{ee} = \sup_{\phi\in I_\phi,\, e\in I_e} Q_{ee}(e, \tilde{a}, \phi)\,.$$

As $I_\phi \times I_e$ is compact and Q continuous with $Q_{ee} < 0$ it follows that $\kappa_{ee} < 0$. By Lemma 11 the myopic action will be in the interval I_e after some period T_e.

Define the projection Pe of an action e to I_e by $Pe = \arg\min_{\hat{e}\in I_e} |\hat{e} - e|$. We have that the subjectively expected contemporaneous loss in period t from taking an action e other than the myopically optimal one is given by:

$$\begin{aligned}
\tilde{\mathbb{E}}_{t-1}[Q(e_t^m, \tilde{a}, \phi) - Q(e, \tilde{a}, \phi)] &= \int_{(\underline{\phi},\overline{\phi})} \{Q(e_t^m, \tilde{a}, \phi) - Q(e, \tilde{a}, \phi)\}\, \pi_{t-1}(\phi)\mathrm{d}\phi \\
&= \int_{(\underline{\phi},\overline{\phi})} \int_e^{e_t^m} Q_e(z, \tilde{a}, \phi)\mathrm{d}z\, \pi_{t-1}(\phi)\, \mathrm{d}\phi \\
&= \int_{(\underline{\phi},\overline{\phi})} \int_e^{e_t^m} \left\{ Q_e(e_t^m, \tilde{a}, \phi) - \int_z^{e_t^m} Q_{ee}(y, \tilde{a}, \phi)\mathrm{d}y \right\} \mathrm{d}z\, \pi_{t-1}(\phi)\, \mathrm{d}\phi \\
&= \int_{(\underline{\phi},\overline{\phi})} \left\{ (e_t^m - e) Q_e(e_t^m, \tilde{a}, \phi) - \int_e^{e_t^m} \int_z^{e_t^m} Q_{ee}(y, \tilde{a}, \phi)\mathrm{d}y\, \mathrm{d}z \right\} \pi_{t-1}(\phi)\, \mathrm{d}\phi
\end{aligned}$$

Using that $\int_{(\underline{\phi},\overline{\phi})} Q(e_t^m, \tilde{a}, \phi)\pi_{t-1}(\phi)\mathrm{d}\phi = 0$ and that the integral bounds of the second term in curly brackets are ordered the same way, we have that

$$\begin{aligned}
\tilde{\mathbb{E}}_{t-1}[Q(e_t^m, \tilde{a}, \phi) - Q(e, \tilde{a}, \phi)] &= \int_{(\underline{\phi},\overline{\phi})} \Big| \int_e^{e_t^m} \int_z^{e_t^m} |Q_{ee}(y, \tilde{a}, \phi)| \mathrm{d}y\, \mathrm{d}z \Big| \, \pi_{t-1}(\phi)\, \mathrm{d}\phi \\
&\geq \int_{(\underline{\phi},\overline{\phi})} \Big| \int_e^{e_t^m} \int_z^{e_t^m} \mathbf{1}_{\{y\in I_e\}} \, |Q_{ee}(y, \tilde{a}, \phi)| \mathrm{d}y\, \mathrm{d}z \Big| \, \mathbf{1}_{\{\phi\in I_\phi\}} \pi_{t-1}(\phi)\, \mathrm{d}\phi \\
&\geq \int_{(\underline{\phi},\overline{\phi})} \Big| \int_e^{e_t^m} \int_z^{e_t^m} \mathbf{1}_{\{y\in I_e\}} \, |\kappa_{ee}| \, \mathrm{d}y\, \mathrm{d}z \Big| \, \mathbf{1}_{\{\phi\in I_\phi\}} \pi_{t-1}(\phi)\, \mathrm{d}\phi \\
&\geq \frac{|\kappa_{ee}|}{2} \mathbf{1}_{\{t\geq T_e\}} \int_{\mathbb{R}} (e_t^m - Pe)^2 \, \mathbf{1}_{\{\phi\in I_\phi\}} \pi_{t-1}(\phi)\, \mathrm{d}\phi \\
&= \frac{|\kappa_{ee}|}{2} \mathbf{1}_{\{t\geq T_e\}} \tilde{\mathbb{P}}_{t-1}[\Phi \in I_\phi](e_t^m - Pe)^2.
\end{aligned}$$

We next derive a (rough) upper bound on the gain of learning. We calculate the upper bound on the per-period gain by taking the difference in expected payoff between an agent who gets to know the state of the world perfectly minus one who learns nothing over and above what he knows at the beginning of period t. To state the bound, we use that by Lemma 17 e^* is Lipschitz continuous on I_ϕ and denote the corresponding Lipschitz constant by κ_o. One has,

$$\begin{aligned}
&\tilde{\mathbb{E}}_{t-1}[Q(e^*(\phi), \tilde{a}, \phi) - Q(e_t^m, \tilde{a}, \phi)] = \int_{(\underline{\phi},\overline{\phi})} \{Q(e^*(\phi), \tilde{a}, \phi) - Q(e_t^m, \tilde{a}, \phi)\} \pi_{t-1}(\phi)\mathrm{d}\phi \\
&\quad \leq \kappa_e \int_{(\underline{\phi},\overline{\phi})} \mathbf{1}_{\{\phi\in I_\phi\}} |e^*(\phi) - e^*(\phi_{t-1})| \, \pi_{t-1}(\phi)\, \mathrm{d}\phi + \kappa_e \int_{(\underline{\phi},\overline{\phi})} \mathbf{1}_{\{\phi\notin I_\phi\}} |e^*(\phi) - e^*(\phi_{t-1})| \, \pi_{t-1}(\phi)\, \mathrm{d}\phi \\
&\quad \leq \kappa_e \, \kappa_o \int_{(\underline{\phi},\overline{\phi})} \mathbf{1}_{\{\phi\in I_\phi\}} |\phi - \phi_{t-1}| \, \pi_{t-1}(\phi)\, \mathrm{d}\phi + \kappa_e \, \tilde{\mathbb{P}}_{t-1}[\Phi \notin T_\phi](e_{max} - e_{min}).
\end{aligned}$$

By Jensen's Inequality, we can bound the above term by

$$\begin{aligned}
&\tilde{\mathbb{E}}_{t-1}[Q(e^*(\phi), \tilde{a}, \phi) - Q(e_t^m, \tilde{a}, \phi)] \leq \kappa_e \, \kappa_o \sqrt{\int_{(\underline{\phi},\overline{\phi})} (\phi - \phi_{t-1})^2 \, \pi_{t-1}(\phi)\, \mathrm{d}\phi} + \kappa_e \, \tilde{\mathbb{P}}_{t-1}[\Phi \notin I_\phi](e_{max} - e_{min}) \\
&\quad = \kappa_e \, \kappa_o \sqrt{\tilde{\mathbb{E}}_{t-1}[(\Phi - \phi_{t-1})^2]} + \kappa_e \, \tilde{\mathbb{P}}_{t-1}[\Phi \notin I_\phi](e_{max} - e_{min}) \,.
\end{aligned}$$

As δ times the current loss must be smaller than $(1-\delta)$ the future gains, we have that

$$\begin{aligned}
&\delta \, \tilde{\mathbb{E}}_{t-1}[Q(e_t^m, \tilde{a}, \phi) - Q(e_t, \tilde{a}, \phi)] \leq (1-\delta)\tilde{\mathbb{E}}_{t-1}[Q(e^*(\phi), \tilde{a}, \phi) - Q(e_t^m, \tilde{a}, \phi)] \\
\Rightarrow\ &\frac{|\kappa_{ee}|}{2} \mathbf{1}_{\{t\geq T_e\}} \tilde{\mathbb{P}}_{t-1}[\Phi \in I_\phi](e_t^m - Pe_t)^2 \\
&\leq \frac{1-\delta}{\delta} \Big\{ \kappa_e \, \kappa_o \sqrt{\tilde{\mathbb{E}}_{t-1}[(\Phi - \phi_{t-1})^2]} + \kappa_e \, \tilde{\mathbb{P}}_{t-1}[\Phi \notin I_\phi](e_{max} - e_{min}) \Big\} \,.
\end{aligned}$$

Consequently, we have that for $t > T$

$$\Rightarrow (e_t^m - Pe_t)^2 \leq \frac{1-\delta}{\delta} \frac{2\,\kappa_e\,\kappa_o}{|\kappa_{ee}|} \frac{\sqrt{\tilde{\mathbb{E}}_{t-1}[(\Phi-\phi_{t-1})^2]}}{\tilde{\mathbb{P}}_{t-1}[\Phi \in I_\phi]} + \frac{1-\delta}{\delta} \frac{2\,\kappa_e}{|\kappa_{ee}|} \frac{\tilde{\mathbb{P}}_{t-1}[\Phi \notin I_\phi]}{\tilde{\mathbb{P}}_{t-1}[\Phi \in I_\phi]} (e_{max} - e_{min})\,.$$

By Lemma 15 the subjective posterior variance is bounded $\tilde{\mathbb{E}}_{t-1}[(\Phi - \phi_{t-1})^2] \leq k/(t-1)$ and thus for $t > T$

$$(e_t^m - Pe_t)^2 \leq \frac{1-\delta}{\delta} \frac{2\,\kappa_e\,\kappa_o\sqrt{k}}{|\kappa_{ee}|} \frac{1}{\tilde{\mathbb{P}}_{t-1}[\Phi \in I_\phi]} \cdot \frac{1}{\sqrt{t-1}} + \frac{1-\delta}{\delta} \frac{2\,\kappa_e}{|\kappa_{ee}|} \frac{\tilde{\mathbb{P}}_{t-1}[\Phi \notin I_\phi]}{\tilde{\mathbb{P}}_{t-1}[\Phi \in I_\phi]} (e_{max} - e_{min})\,. \quad (18)$$

As $\tilde{\mathbb{P}}_{t-1}[\Phi \in I_\phi]$ converges to 1 by the definition of $I_\phi = (\underline{\phi}_\infty - \gamma, \overline{\phi}_\infty + \gamma)$, it follows that the right-hand side converges to zero. By Lemma 11 the myopically optimal action e_t^m is strictly inside I_e. Consequently, (18) implies $\liminf_{t\to\infty} e_t$ and $\limsup_{t\to\infty} e_t$ are strictly inside I_e and (18) implies that $\lim_{t\to\infty}(e_t^m - e_t)^2 = 0$. □

Proof of Theorem 2. By Lemma 18, the limit inferior and the limit superior over actions are the same if the agent behaves strategically and if he behaves myopically. Hence, it follows from the proof for myopic actions (Theorem 1) that the agent's belief converges in distribution to a Dirac measure on ϕ_∞. □

A.2.4 Further Proofs on Stable Beliefs

Proof of Proposition 5. Analogously to $\hat{F}_i$ defined for the perceived error in the text, we let F_i be the empirical frequency of the true error ϵ_t at the pre-specified time periods $t_1, t_2, \ldots,$ i.e. $F_i(x) = |\{i' \leq i | \epsilon_{t_{i'}} \leq x\}|/i$.

For an infinite sample, the agent believes must equal his stable belief ϕ_∞. Hence, once the agent observed an infinite sample,

$$\begin{aligned}
\tilde{\epsilon}_{t_i} &= Q(e_{t_i}, A, \Phi) - Q(e_{t_i}, \tilde{a}, \phi_\infty) + \epsilon_{t_i} \\
&= Q(e^*(\phi_\infty), A, \Phi) - Q(e^*(\phi_\infty), \tilde{a}, \phi_\infty) + \epsilon_{t_i} - \int_{e_{t_i}}^{e^*(\phi_\infty)} Q_e(s, A, \Phi) ds + \int_{e_{t_i}}^{e^*(\phi_\infty)} Q_e(s, \tilde{a}, \phi_\infty) ds \\
&= \epsilon_{t_i} - \int_{e_{t_i}}^{e^*(\phi_\infty)} Q_e(s, A, \Phi) ds + \int_{e_{t_i}}^{e^*(\phi_\infty)} Q_e(s, \tilde{a}, \phi_\infty) ds.
\end{aligned}$$

Now because Q is twice continuously differentiable and $e_{t_i} \to e^*(\phi_\infty)$, for every $\eta > 0$ there exists a $\hat{t}$ such that for all $t_i > \hat{t}$, $\tilde{\epsilon}_{t_i} \in (\epsilon_{t_i} - \eta, \epsilon_{t_i} + \eta)$. Choose a time period τ such that the fraction of observations in the sequence $t_1, t_2, ...$ the agent observed before $\hat{t}$ is less than η. Then for all $t > \tau$, $\hat{F}_t(x) \in (F_t(x-\eta) - \eta, F_t(x+\eta) + \eta)$. This implies that $\hat{F}_t(x) \to F_t(x)$ for all x, for otherwise there exists an $x \in (0,1)$ and an $\eta > 0$ such that for all τ there exists some $t > \tau$ for which $\hat{F}_t(x) \notin (F_t(x-\eta) - \eta, F_t(x+\eta) + \eta)$, a contradiction. Now because the true errors are i.i.d., $F_t(x) \to F(x)$ a.s., and since $\hat{F}_t(x) \to F_t(x)$, we conclude that $\hat{F}_t(x) \to F(x)$ a.s..

Proof of Proposition 6. In the text.

Proof of Proposition 7. For the loss-function specification, the surprise function

$$\Gamma(\phi) = -(\tilde{a} - A) + (\Phi - \phi) - L(\Phi - \phi).$$

The stable beliefs are the Dirac measure on the unique root ϕ_∞ of Γ. Using that this root $\phi_\infty > \Phi$ and that $A > \tilde{a}$ to rewrite $\Gamma(\phi_\infty) = 0$ gives

$$L(\phi_\infty - \Phi) + (\phi_\infty - \Phi) = |\Delta|.$$

Thus $\phi_\infty - \Phi < |\Delta|$ and $L(\phi_\infty - \Phi) < \Delta$. $\square$