Itś not my Fault! Self-Confidence and Experimentation

Nina Hestermann (Toulouse School of Economics)
Yves Le Yaouanq (LMU Munich)

Discussion Paper No. 124

November 2, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

It's not my fault! Self-confidence and experimentation*

Nina Hestermann † Yves Le Yaouanq ‡

First version: March 8, 2016
Current draft: October 8, 2018

Abstract

We study the inference and experimentation problem of an agent in a situation where the outcomes depend on the individual's intrinsic ability and on an external variable. We analyze the mistakes made by decision-makers who hold inaccurate prior beliefs about their ability. Overconfident individuals take too much credit for their successes and excessively blame external factors if they fail. They are too easily dissatisfied with their environment, which leads them to experiment in variable environments and revise their self-confidence over time. In contrast, underconfident decision-makers might be trapped in low-quality environments and incur perpetual utility losses.

*We are extremely grateful to Roland Bénabou, Florian Englmaier, Christian Gollier and Jean Tirole for their guidance and encouragement. We also thank Nicolas Astier, Miaomiao Dong, Daniel Garrett, George Lukyanov, Thomas Mariotti, Takeshi Murooka, Klaus Schmidt, Séverine Toussaert, and conference participants at M-BEES 2016, ECBE 2016, EEA 2016 and LMU Munich. Yves Le Yaouanq gratefully acknowledges financial support from the *Corps des Mines* and from the *Deutsche Forschungsgemeinschaft* (through CRC TRR 190). All remaining mistakes are due to external factors.

†Toulouse School of Economics. E-mail: nina.hestermann@gmail.com

‡Ludwig-Maximilians-Universität, Munich. E-mail: yves.leyaouanq@econ.lmu.de

1 Introduction

Individuals usually have imperfect knowledge about their ability to succeed in their projects. Since many studies have claimed that people tend to think too highly of their intrinsic characteristics in important dimensions—intelligence, skills, willpower—the psychology and economics literature has devoted a lot of attention to investigating the consequences of overconfidence on behavior and welfare.

In many situations, the outcomes of agents' endeavors not only depend on their intrinsic ability but also on some characteristics of their environment, which they may at first know imperfectly. In this paper, we show that overconfidence distorts the process by which individuals learn about these exogenous payoff-relevant variables. For instance, a student who initially holds confident expectations about his skills but who repeatedly fails at exams might revise his beliefs about his ability, but also conclude that the academic system is less fair than he had thought. This pessimistic inference, in turn, conditions his future decisions, such as how much effort to invest for the next exams, or even whether to drop out of the university.

An agent repeatedly performs a task and receives a binary outcome: success or failure. At each date the probability of succeeding $p(\lambda, \theta)$ is an increasing function of the agent's fixed ability θ, and of an exogenous parameter λ that summarizes the characteristics of the external contingencies in which the agent operates: the difficulty of the task, the abilities and intentions of the co-workers, the returns to human capital, etc. To understand the causal effect of self-confidence, we compare two individuals who only differ in their prior beliefs about θ, one being overconfident in the monotone likelihood ratio ordering relative to the other.

In Section 3 we study the passive inferences made by individuals who operate in a stable environment after a finite number of periods. We show that overconfident individuals are prone to a misattribution of outcomes when forming beliefs about λ. Perhaps surprisingly, the mistake takes a subtle form that depends on the degree of complementarity of θ and λ in the production function p: in particular, it is not true that overconfident individuals are always too pessimistic about the quality of their environment. We give a precise characterization of the misattribution and we show that

its interpretation is related to the *self-serving attribution bias* documented in psychology: overconfident individuals tend to overestimate the informativeness of positive outcomes about their ability, as they take too much merit for their achievements, and they underestimate the informativeness of negative outcomes, as they hold external contingencies responsible for any failures. This mistake implies a variety of related misperceptions, which we outline in Section 3. As an example, overconfidence leads successful individuals to overestimate of the productivity of investment in human capital, and leads less successful individuals to underestimate it.

In Section 4, we embed the baseline model into an active experimentation framework and we focus on the asymptotic properties of the process of updating one's beliefs. Analyzing whether endogenous learning opportunities ultimately eliminate initial misperceptions is important in knowing whether inaccurate self-assessments are a transient bias limited to inexperienced decision-makers or whether this distortion can survive in the long run. We show that the agent's initial beliefs about his ability have a long-lasting influence on his behavior and beliefs. This result contrasts with standard Bayesian models with one-dimensional uncertainty, where the influence of prior beliefs vanishes in the long run.

The agent's ability θ is fixed throughout the infinite horizon. At each period the agent decides whether to stay in the current environment or to replace it by another (randomly drawn) environment, for instance changing jobs, re-orienting one's academic career, etc. The individual is patient and faces a trade-off between exploration, that is, acquiring knowledge about himself and the current environment, and exploitation, that is, maximizing the expected reward. Our main result is that overconfidence and underconfidence have different implications for long-run beliefs, behavior, and welfare. An overconfident individual tends to be too easily dissatisfied with the external conditions and to expect (incorrectly) higher rewards elsewhere. A consequence of this is to tend to switch too early from one environment to another. This experimentation effort provides the agent with a large data set of outcomes received in variable external conditions. Accordingly, blaming external factors for his failure is no longer credible in the long run, and the agent's overconfidence is asymptotically reduced to the point where his decisions and payoffs are optimal. In contrast, under-

confident decision-makers are too easily satisfied with their environment. A consequence might be that they—wrongly—stop experimenting, and never learn the truth about their ability, as they perpetually and incorrectly attribute their surprisingly high success rate to the quality of their current conditions. Underconfident individuals might therefore be trapped in low-quality environments and incur utility losses forever due to their misperception (Proposition 4).

Contrary to common wisdom, our analysis therefore suggests that underconfidence is more problematic than overconfidence, since these two distortions have different implications for long-run learning. We believe that this finding is consistent with the existing evidence. First, our model predicts that overconfidence generates utility losses in the short run in any new environment. Second, the theory also predicts that the rate of learning about one's ability is slowed down by the identification challenge which is at the core of the model, and that complete learning is not achieved in the long run for individuals who stay in stable conditions. This might explain why even experienced decision-makers, e.g., CEOs (Malmendier and Tate, 2005), are sometimes found to be overconfident. Finally, the prediction that overconfidence is reduced by active learning while underconfidence persists is compatible with the fact that underconfident reports are rarely encountered in field data. This latter finding might be due to selection effects. Since underconfident individuals endogenously stay away from ability-intensive activities, they might be underrepresented in the samples studied by researchers: for instance, individuals who are unconfident about their skills as CEOs endogenously choose a different career path. Our model suggests that these individuals, who are absent from the field evidence, are those who incur the largest welfare costs in the long run since their decisions endogenously prevent them from correcting their beliefs.

Our paper connects the literature on self-confidence with the literature on learning with misperceptions. We use the term overconfidence to describe the inflated beliefs that many individuals hold about their own skills, talent, or personal traits, as suggested by a large literature in psychology and economics.[1] Early evidence, such as the better-than-average

[1]This definition is conceptually distinct from others used in the economics literature, for instance individuals' tendency to overestimate the precision of their information

effect (Svenson, 1981; Thaler, 2000; Weinstein, 1980) or behavioral inefficiencies in competitive environments (Camerer and Lovallo, 1999; Hoelzl and Rustichini, 2005), has led many scholars to conclude that overconfidence is widespread, an observation corroborated by field data, e.g., on CEOs (Malmendier and Tate, 2005), truck drivers (Burks et al., 2013), and professional chess or poker players (Park and Santos-Pinto, 2010).[2] A large theoretical literature has been devoted to analyzing the costs of overconfident beliefs. An unreasonably high self-confidence might lead individuals to exert too much effort with little chance of succeeding (Bénabou and Tirole, 2002), set unrealistic goals (Baumeister et al., 1993), or compete too much. As an illustration, Barber and Odean (2001) link overconfidence to excessive trading and show that men, who are known for being more overconfident, trade 45% more than women and incur important losses from this. In this literature, the effect of overconfidence generally comes down to making the individual too optimistic about future outcomes.

We go beyond this literature by showing that overconfident individuals have a tendency to attribute their achievements to their own merits, but their failures to external factors. This self-serving attribution bias has been noted by psychologists in various contexts: academic outcomes (Arkin and Maruyama, 1979), car accidents (Stewart, 2005), collective or individual performance in sport (Lau and Russell, 1980), outcome of joint projects, for instance among couples (Ross and Sicoly, 1979).[3] We give a precise

(Grubb, 2009).

[2]Two criticisms have been addressed to this literature. First, several explanations for the existing evidence based on rational individual Bayesian learning have been offered (see in particular Van den Steen, 2004; Zabojnik, 2004; Kőszegi, 2006; Santos-Pinto and Sobel, 2005; Benoît and Dubra, 2011). Recent research provides tests that overcome these limitations, (for instance Benoît et al., 2015; Eil and Rao, 2011; Möbius et al., 2013). Second, the evidence for aggregate overconfidence appears to be mixed: overconfidence is commonly observed for easy tasks, but several studies report aggregate underconfidence for difficult tasks (see for instance Moore, 2007; Moore and Healy, 2008; Kruger et al., 2008; Benoît and Dubra, 2011).

[3]To our knowledge, the only literature in economics that has explored the consequences of self-serving attribution biases has focused on financial applications. Gervais and Odean (2001) model traders who become overconfident by taking too much credit for successes; they show that the attribution bias leads them to make overconfident decisions and incur losses in the long run. Billett and Qian (2008) present empirical results consistent with self-serving attributions. Libby and Rennekamp (2012) verify experimentally that overconfident beliefs due to biased attributions influence financial decisions. In all these papers, biased attributions are the channel by which people become overconfident and—contrary to our setting—have no direct effect on decisions.

definition and characterization of the attribution bias as a function of the degree of complementarity between the ability of the individual and the quality of the environment. Our model shows that the attribution bias does not necessarily indicate motivated reasoning at the inference stage (Kunda, 1990; Zuckerman, 1979), since an individual who applies Bayes' rule to incorrect prior beliefs makes inferences that appears biased to an external observer, as documented in the experiment by Grossman and Owens (2012).

Our second contribution is to analyze the effects of overconfidence on asymptotic posterior beliefs. Our result that passive learning is not necessarily complete asymptotically complements a literature that questions and extends the standard results on the consistency of posterior beliefs. Some of the papers in this literature assume a departure from Bayesian updating (Rabin and Schrag, 1999; Schwartzstein, 2014; Gottlieb, 2017; Benjamin et al., 2016), while others analyze the updating of an agent who initially attaches a null probability to the true data-generating process (Berk, 1966; Bunke and Milhaud, 1998). Our model is instead based on an identification issue: an agent provided with an infinite number of signals received in a stable environment cannot learn about two dimensions at the same time, as several distinct theories can explain the outcomes. A related result by Acemoglu et al. (2016) shows that two Bayesian agents can disagree about the data-generating process in the long run when they have different initial beliefs about the signal likelihood ratio.

That active experimentation need not result in complete learning is already well-known (Aghion et al., 1991; Easley and Kiefer, 1988). The decision problem that we study extends the standard experimentation frameworks (Banks and Sundaram, 1992) by assuming that the decision-maker learns about two uncertain parameters, one of which (λ) influences only the value of the current arm, while the other (θ) conditions the rewards to all arms. Our model is also distinct from recent work on active learning with a misspecified model (Esponda and Pouzo, 2016; Fudenberg et al., 2017), which assumes that agents' beliefs assign zero probability to the true mapping between actions and consequences. We assume that prior beliefs have full support, which guarantees that learning is feasible and in turn implies that agents who face a stable process are no longer surprised by their outcomes in the long run. We see two benefits of this specification. First,

incompleteness in long-run learning can be attributed to endogenous data limitations resulting from the agent's own choices rather than to inconsistent prior beliefs. Second, our model circumvents the standard criticism addressed to theories of misspecified learning, according to which decision-makers should reconsider their prior beliefs after a sufficiently long history that contradicts their expectations.

The effect of overconfident beliefs on learning about exogenous variables was independently explored by Heidhues et al. (2018), who characterize the vicious circle of suboptimal actions and incorrect attributions resulting from the joint evolution of beliefs and behavior. Our model studies this question with a different angle. Heidhues et al. (2018) rule out learning about one's ability, which we allow for throughout the paper and which is the main focus of Section 4. They also restrict attention to decisions made in a stable environment while the problem that we study consists in deciding whether to opt out of one's current environment. Interestingly, the effects of overconfidence are distinct in the two models. Heidhues et al. (2018) show that, in a stable environment and under some assumptions on the technology, overconfidence results in greater utility losses than underconfidence. Our model shows that exactly the opposite is true for an agent who has the opportunity to experiment in different environments. While most of the literature has focused on the costs of overconfidence, our results suggest that in dynamic settings underconfidence is the most problematic distortion due to its self-confirming nature.[4]

The paper is organized as follows. Section 2 presents the environment. Section 3 analyzes the attribution bias in finite time. Section 4 focuses on asymptotic learning. Section 5 discusses some interpretations of the model and concludes.

[4]Dubra (2004) and Zabojnik (2004) also analyze the link between self-confidence and insufficient sampling. Dubra (2004) studies a search model and shows that optimism is less harmful than pessimism, as optimists are less likely to accept suboptimally low offers. Zabojnik (2004) assume that individuals are information-loving when their self-confidence is low but information-averse when their self-confidence is high, and thus stop sampling in the latter case. The lack of complete learning in our model is instead based on the identification issue that arises in stable environments, as individuals keep receiving information in every period.

2 Environment

Payoffs An individual is engaged in a repeated task over an infinite horizon indexed by $t \in \{1, 2, \cdots\}$. On each date t, the individual receives a binary outcome π_t: a success is denoted by $\pi_t = 1$, whereas a failure is denoted by $\pi_t = 0$. The agent's outcome at date t is stochastic and depends on two variables. The first variable is the agent's intrinsic ability at the task, written θ and drawn on the non-degenerate support $\Theta = [\underline{\theta}, \bar{\theta}]$. The second variable is a task-specific parameter λ that is exogenous to the agent. The variable λ is distributed according to the continuous full-support pdf g_0 on the non-degenerate interval $\Lambda = [\underline{\lambda}, \bar{\lambda}]$. The variable λ describes some permanent features of the task or the environment about which the agent learns by experimenting. The variables λ and θ are independent. Conditional on a pair (λ, θ), the outcomes are independently and identically distributed across periods. The agent's probability of succeeding at the task is therefore stationary and written $p(\lambda, \theta)$, and increases with the agent's ability θ and with the quality of the environment λ. The function p is of class C^2 and bounded away from 0 and 1. We write p_λ and p_θ for the partial derivatives of p, and assume that $p_\lambda > 0$ and $p_\theta > 0$.

Stability of the environment We assume that θ is fixed. In Section 3 we also assume that λ is fixed, reflecting the idea that the nature of the environment remains stable over time. In Section 4 we allow for the possibility that the environment changes, which we model as a new random draw of λ from Λ, either for exogenous reasons (automatic job rotation, beginning of a new academic year with new instructors, etc.), or as the result of the agent's own decisions.

Self-confidence Our analysis consists in comparing the beliefs and decisions of two agents who differ only in their initial self-confidence. Agents 1 and 2 share the same prior distribution over λ, given by the pdf g_0, but hold different initial beliefs about their ability. Agent i $(i = 1, 2)$ starts the game with a prior pdf $f_{0,i}$ that represents his beliefs about θ. The functions $f_{0,1}$ and $f_{0,2}$ are linked by a monotone likelihood ratio property, which introduces a notion of comparative self-confidence. We write $\succeq$ for

the monotone likelihood ordering applied to pdfs: if u and v are two functions of a real variable x defined on the same interval, $u \succeq v$ means that the function $x \to u(x)/v(x)$ is well-defined and nondecreasing. We assume that $f_{0,1} \succeq f_{0,2}$.

We also assume throughout the paper that $f_{0,1}$ and $f_{0,2}$ have full support on Θ. This assumption ensures that learning is not impeded by the fact that the agent's prior beliefs ascribe probability zero to the true state of the world as in misspecified learning models (Esponda and Pouzo, 2016; Fudenberg et al., 2017; Heidhues et al., 2018).

Our results can be interpreted in two different ways. In the first interpretation, $f_{0,2}$ is the "correct" prior distribution, and the behavior of $f_{0,1}$ reflects the mistakes caused by overconfidence in an absolute sense. In the second interpretation, the disagreement is considered in a relative sense only, and comparing the behaviors of the two agents informs us about the causal effect of self-confidence on attributions and experimentation behavior without taking a stance as to which of $f_{0,1}$ or $f_{0,2}$ is more correct. Under both interpretations, we refer to agent 1 as being overconfident.

For each date t, a history h_t is characterized by the identity of the environment tried out at any date s up to date t and the resulting outcome π_s. We use the subscripts t and i to write agent i's posterior beliefs at date t. For instance, $f_{t,h_t,i}$ is agent i's posterior pdf regarding θ following history h_t, and $F_{t,h_t,i}$ is the corresponding cdf.

Claim 1 establishes that the monotone likelihood ratio ordering is preserved by Bayes' rule. Our definition of comparative self-confidence is thus robust to learning: agent 1 remains more confident than agent 2 after any common sequence of observations.

Claim 1.

$$\textit{For any } (t, h_t),\ f_{t,h_t,1} \succeq f_{t,h_t,2}.$$

3 Attribution bias

We begin by analyzing the agents' inferences in a situation where they repeatedly perform the task in a stable environment. A value of λ is drawn at date 0 from Λ and remains fixed for several periods.

3.1 General results

Our first result characterizes the direction of the misperception of λ that results from overconfidence about θ. To build intuition, consider the example of a manager of ability θ who works on a project together with an employee of unknown ability λ. Manager 1 believes that he is high-skilled ($\theta = \overline{\theta}$), while manager 2 believes that he is low-skilled ($\theta = \underline{\theta}$); both believe that the employee is high-skilled ($\lambda = \overline{\lambda}$) or low-skilled ($\lambda = \underline{\lambda}$) with probability 0.5 each.

Suppose that the project fails. Will manager 1 become more or less optimistic about the type of the employee than manager 2? We show that the answer to this question depends on the degree of complementarity between θ and λ. Suppose first that θ and λ are complements in the production function p. For instance, there exists ϵ close to zero such that $p(\overline{\lambda}, \overline{\theta}) = 1 - \epsilon$ while $p(\lambda, \theta) = \epsilon$ for all other values of (λ, θ): that is (in approximation), the project succeeds if and only if both the manager and the employee are skilled. After a failure, manager 1 infers that the employee is high-skilled with probability close to zero, while manager 2 infers very little about λ: the first manager explains the failure by the low skills of the employee, while the more realistic manager takes responsibility for it. That is, manager 1 believes that he is working with an employee whose type makes his own skills irrelevant for the success of the venture, while manager 2 believes that the project would be more successful if he were himself more skilled.

Suppose instead that θ and λ are substitutes in the production function p. For instance, $p(\underline{\lambda}, \underline{\theta}) = \epsilon$ while $p(\lambda, \theta) = 1 - \epsilon$ for all other values of (λ, θ): the project succeeds if and only if at least one of the team members is high-skilled. The direction of the attribution bias is opposite to the previous case: after a failure, manager 2 infers that the employee is not skilled, while manager 1 does not update at all about λ. However, the intuition is similar: while manager 2 believes that his own ability makes a large difference for a project with this employee, manager 1 underestimates the importance of his ability for the collective venture.

These examples illustrate that the distortion implied by overconfidence depends on whether the outcomes are more informative about the agent's

ability in a favorable or in an unfavorable environment. Overconfident agents who fail have a tendency to think that their own ability is not important. Conversely, overconfident agents who succeed believe that their own skills were instrumental to the success. This distortion does not require any inferential mistake, as it is implied by Bayes' rule applied to potentially incorrect prior beliefs.

The formal characteristic of the technology that determines the nature of the attribution bias is linked to the value of the cross-partial derivative $p_{\lambda\theta}$. We say that p is *log-submodular* (or *log-sbm*) if $p_{\lambda\theta}p \leq p_\lambda p_\theta$. We say that p is *log-supermodular* (or *log-spm*) if the inequality is reversed, and use the adjective strict when the inequality is strict on $\Lambda \times \Theta$.

We assume that $p_{\lambda\theta}p - p_\lambda p_\theta$ and $p_{\lambda\theta}(1-p) + p_\lambda p_\theta$ have constant signs, which leaves three possible cases. In the first case, both p and $1-p$ are log-sbm. This implies that, for any $\overline{\lambda} > \underline{\lambda}$, the likelihood ratios $p(\overline{\lambda},\theta)/p(\underline{\lambda},\theta)$ and $(1-p(\overline{\lambda},\theta))/(1-p(\underline{\lambda},\theta))$ are both nonincreasing in θ: a success (good news) in a good environment is then less informative about θ than a success in a bad environment, while a failure (bad news) contains more information if it is obtained in a good environment. The usual additive ($p(\lambda,\theta) = u(\lambda) + v(\theta)$) and multiplicative ($p(\lambda,\theta) = u(\theta)v(\lambda)$) forms belong to this category.

In the second case, p is strictly log-spm, which implies that $1-p$ is strictly log-sbm. The likelihood ratio $p(\overline{\lambda},\theta)/p(\underline{\lambda},\theta)$ is then increasing in θ. This assumption describes situations where succeeding in a good environment is more informative about the agent's ability than succeeding in a bad environment: this can be due to the fact that even high-skilled individuals are very unlikely to succeed in an unfavorable environment, which implies that a successful outcome would be attributed to an unlikely lucky break rather than to intrinsic dispositions.

In the third case, $1-p$ is strictly log-spm, which implies that p is strictly log-sbm. The likelihood ratio $(1-p(\overline{\lambda},\theta))/(1-p(\underline{\lambda},\theta))$ is then increasing in θ: failing in a bad environment is more informative about the agent's ability than failing in a good environment. This can be due to the fact that succeeding in the good conditions is extremely likely even for low-skilled individuals, which implies that failures are attributed to an adverse random shock.

Proposition 1 summarizes the direction of the bias in these three cases. We write $h_t = n_t$ for a history composed of n_t successes out of t attempts and $g_{t,n_t,i}$ for the conditional posterior beliefs about λ. Except when p and $1-p$ are log-sbm, we can only compare $g_{t,n_t,1}$ and $g_{t,n_t,2}$ for extreme scenarios (large success rate, or small success rate), as a comparison in the monotone likelihood ordering is not possible in general for intermediate success frequencies.

Proposition 1. *1. If p and $1-p$ are log-sbm, then $g_{t,n_t,1} \preceq g_{t,n_t,2}$ for any (n_t, t).*

2. If p is strictly log-spm, there exists $\alpha_0, \beta_0 \in (0,1)$ such that $g_{t,n_t,1} \succeq g_{t,n_t,2}$ if $n_t/t \geq \alpha_0$, and $g_{t,n_t,1} \preceq g_{t,n_t,2}$ if $n_t/t \leq \beta_0$.

3. If $1-p$ is strictly log-spm, there exists $\alpha_1, \beta_1 \in (0,1)$ such that $g_{t,n_t,1} \preceq g_{t,n_t,2}$ if $n_t/t \geq \alpha_1$, and $g_{t,n_t,1} \succeq g_{t,n_t,2}$ if $n_t/t \leq \beta_1$.

Proposition 1 states that the overconfident agent misperceives his environment relative to the more realistic agent. This result does not require any correlation between θ and λ from the ex ante perspective, and follows from the correct application of Bayes' rule to heterogeneous prior beliefs.

3.2 Misperception of the informativeness

As Proposition 1 shows, the bias in inference due to overconfidence thus takes different forms depending on the shape of p. We now argue that this set of results relies on a similar intuition, which is closely linked to the notion of *self-serving attribution bias* in psychology. In this section, we give a precise definition of the attribution bias and provide a result that unifies the three cases considered in Proposition 1.

The attribution bias is commonly understood in the following way: "We are prone to alter our perception of causality [...]. We attribute success to our own dispositions and failure to external forces." (Hastorf et al., 1970, p. 73) In typical experiments on the attribution bias, participants learn about their performance at a task and are asked to formulate a causal explanation of their outcome. For instance, in the study by Johnson et al. (1964), participants teach arithmetic concepts to fourth-grade boys and

learn about the performance of the pupils at a subsequent test. Johnson et al. (1964) show that teachers tend to attribute positive performance to their own teaching skills, whereas they place the responsibility for poor performance on external factors, such as the pupil's lack of motivation for learning.

To formalize the notion of a causal explanation of success and failure, consider the following hypothetical elicitation. Suppose that subjects 1 and 2 participate in the experiment by Johnson et al. (1964). Both subjects fail at teaching arithmetic to a child, and learn that another subject 3 who was paired with the same pupil has also been unsuccessful at the task. Subjects 1 and 2 are then asked to use this information in order to form beliefs about the teaching ability of subject 3. Subject 2 has taken responsibility for the child's disappointing performance. He thus believes in a causal link between poor teaching skills and the learning outcomes of the pupil. Thus, he should update his beliefs about the ability of subject 3 downward by a large amount. In contrast, subject 1 believes that the poor performance is mostly due to the child's lack of motivation for learning: he should thus not revise his beliefs about the skills of subject 3, as he thinks that the child is responsible for the outcome. To sum up, due to his overconfidence, subject 1 has a tendency to overestimate the ability of the other participants who have been unsuccessful in the same conditions. The same intuition applies in the case where subjects 1, 2 and 3 have been successful: the overconfident subject 1 overestimates the importance of teaching ability in the learning outcome, and thus he overestimates the ability of other successful participants.

We now formally establish that this bias is predicted by Bayesian updating applied to overconfident prior beliefs, and that this observation is true irrespective of the shape of the function p. Corollary 1 below unifies the seemingly disparate results exposed in Proposition 1. Formally, the two agents 1 and 2 observe the common history (t, n_t), and are asked to form beliefs about the ability of an agent 3 who has performed the task in the same conditions (that is, with the same λ) and obtained the same history (t, n_t). Initially, agents 1 and 2 share a common prior distribution over the type $\tilde{\theta}$ of agent 3, represented by the continuous pdf $\tilde{f}_0$ defined on Θ. By

Bayes' rule, agent i $(i = 1, 2)$ estimates

$$\tilde{f}_{t,n_t,i}(\tilde{\theta}) = \frac{\tilde{f}_0(\tilde{\theta}) \int_\Lambda p(\lambda,\tilde{\theta})^{n_t}(1-p(\lambda,\tilde{\theta}))^{t-n_t} dG_{t,n_t,i}(\lambda)}{\iint_{\Lambda\times\Theta} \tilde{f}_0(\theta) p(\lambda,\theta)^{n_t}(1-p(\lambda,\theta))^{t-n_t} dG_{t,n_t,i}(\lambda) d\tilde{F}_0(\theta)}.$$

The comparison of $\tilde{f}_{t,n_t,1}$ and $\tilde{f}_{t,n_t,2}$ informs us about the theories formed by the two agents about the determinants of success and failure in their environment. Definition 1 formalizes our definition of the self-serving attribution bias: an agent is prone to this bias if he overestimates, in relative terms, the ability of other individuals who have obtained the same outcomes.

Definition 1. Agent 1 is prone to a *self-serving attribution bias* relative to agent 2 after the history (n_t, t) if $\tilde{f}_{t,n_t,1} \succeq \tilde{f}_{t,n_t,2}$.

Corollary 1 shows that overconfident prior beliefs causally generate a self-serving attribution bias. This result covers the three cases exposed in Proposition 1.[5]

Corollary 1. *Suppose that $p_{\lambda\theta}p - p_\lambda p_\theta$ and $p_{\lambda\theta}(1-p) + p_\lambda p_\theta$ have constant signs. Then there exists $\alpha_2, \beta_2 \in (0,1)$ such that, for any (t, n_t) such that $n_t/t \geq \alpha_2$ or $n_t/t \leq \beta_2$, agent 1 is prone to a self-serving attribution bias relative to agent 2 after the history (n_t, t).*

3.3 Additional implications

In this section, we briefly mention two other implications of the distortion in inferences generated by overconfident prior beliefs and uncovered by Proposition 1. First, while static models predict a positive relationship between overconfidence (overestimation of θ) and optimism about future outcomes (overestimation of $p(\lambda,\theta)$), this effect is not robust to

[5] If $f_{0,1}$ is interpreted as the most accurate prior distribution, the model also predicts an inverse attribution bias for an agent 2 who starts from an unrealistically low self-confidence. This finding resonates with casual evidence on the imposter syndrome, whereby high achievers understate their own merit and exaggerate the role of luck in their accomplishments. Consistently with the model, this mindset if found more often among women or minority groups whose self-confidence levels tend to be below the population average (Clance and Imes, 1978; Sonnak and Towell, 2001).

updating in a situation of two-dimensional uncertainty. Consider for instance the manager–employee example of subsection 3.1 and suppose that $p(\underline{\lambda}, \overline{\theta}) = p(\underline{\lambda}, \underline{\theta}) = 1/3, p(\overline{\lambda}, \underline{\theta}) = 2/3, p(\overline{\lambda}, \overline{\theta}) = 1 - \epsilon$, where ϵ is close to zero. After a failure, an overconfident manager estimates that his future probability of success, if he keeps working with the same employee, is close to 1/3, whereas a realistic manager predicts a future success rate strictly greater than 1/3. The excessive inference drawn by the overconfident manager about the ability of the employee makes him more pessimistic regarding the future productivity of the venture.

Second, the overconfident agent misperceives the productivity of human capital in the environment. To formalize this result, suppose that based on his own outcomes, the agent tries to estimate the difference in expected productivity between an individual of known ability θ_L and an individual of known ability $\theta_H > \theta_L$. Formally, the agent estimates

$$\vartheta_{t,n_t,i} = \int_{\Lambda} [p(\lambda, \theta_H) - p(\lambda, \theta_L)] dG_{t,n_t,i}(\lambda).$$

This subjective parameter potentially governs important decisions, such as how much to invest in one's own (or one's children's) human capital. We restrict attention to the second and third cases in Proposition 1, in which the role of λ has an unambiguous interpretation: if p is strictly log-spm, λ measures the extent to which human capital is important in the agent's environment, whereas if $1 - p$ is strictly log-spm, λ is an inverse measure of this variable.[6]

After a successful history, agent 1 forms the belief that talented individuals are appropriately rewarded relative to their low-skilled peers. As a consequence, he sees large benefits from investment in human capital. After failing, in contrast, agent 1 doubts that people obtain their just deserts,

[6]If p is strictly log-spm, the ratio $p(\lambda, \overline{\theta})/p(\lambda, \underline{\theta})$ is increasing in λ, and thus successes are more indicative of high skills if λ is large; the ratio $(1 - p(\lambda, \overline{\theta}))/(1 - p(\lambda, \underline{\theta}))$ is decreasing in λ, and thus failures are more indicative of low skills if λ is large. The higher is λ, the more important is intrinsic ability in the agent's outcomes. If $1 - p$ is strictly log-spm, the opposite statements are true: the higher is λ, the less θ is important. If p and $1 - p$ are log-sbm, after a failure, an observer infers more about θ in a large λ environment than in a low λ environment, but the opposite holds after a success. Thus, whether an increase in λ makes ability more or less important has no unambiguous answer in that case.

and therefore underestimates the benefits from investment in θ.

Corollary 2. *Suppose that either p or $1-p$ is strictly log-spm. There exists $\alpha_3, \beta_3 \in (0,1)$ such that $\vartheta_{t,n_t,1} \geq \vartheta_{t,n_t,2}$ if $n_t/t \geq \alpha_3$, and $\vartheta_{t,n_t,1} \leq \vartheta_{t,n_t,2}$ if $n_t/t \leq \beta_3$.*

4 Asymptotic learning

We now turn to analyzing the individuals' beliefs after they receive infinite sequences of outcomes. Our objective is to analyze the conditions under which the initial miscalibrations in self-confidence are eliminated asymptotically by Bayesian learning. We first consider a passive learning situation in subsection 4.1 before making the agents' experimentation decisions endogenous in subsection 4.2.

4.1 Passive learning

We start by analyzing the agents' asymptotic beliefs in situations where they perform the task in every period in an environment which is exogenously imposed on them. The following results are a useful preliminary to studying the active experimentation decision in subsection 4.2. They are also of independent interest for those applications where individuals do not make active experimentation decisions. Our main result is that whether Bayesian individuals eventually learn the truth about themselves crucially depends on the stability of their external conditions.

We first analyze the case where the value of λ is drawn at the beginning of the game and fixed thereafter, reflecting the assumption that the external conditions are uncertain but stable. We write $f_{t,i}$ and $g_{t,i}$ for the unconditional posterior beliefs at date t, and $k_{t,i}$ for the posterior pdf formed over the probability $p(\lambda, \theta)$ of succeeding in this environment. We use capital letters (F, G and K) for the cdfs. The true parameters of the data-generating process are written λ_0 and θ_0. We assume that $p(\lambda_0, \theta_0)$ belongs to the interior of $[\inf p, \sup p]$.

Our full-support assumptions ensure that the learning process is correctly specified, in the sense that the agents' prior beliefs regarding the

probability of success attribute a positive probability to any open neighborhood of the true value $p(\lambda_0, \theta_0)$.

The individuals receive an infinite sequence of informative signals. Standard statistical learning theorems prove that the sequence of posterior beliefs about p is consistent: almost surely, $K_{t,i}$ converge weakly to the Dirac measure $\delta_{p(\lambda_0,\theta_0)}$ centered at $p(\lambda_0, \theta_0)$, which is approximated by the empirical success rate.

Nevertheless, the information received is not sufficient to extract the true values of λ and θ individually: since several pairs (λ, θ) predict the same success rate, neither parameter is identifiable separately. Since the agents initially, and at each point in time, have different beliefs about θ, they form two different theories that both correctly explain their observations. In the limit, the two individuals agree on the future empirical success rate, but the overconfident agent keeps overestimating θ and forms more pessimistic beliefs about the quality of the environment.

Proposition 2. *Suppose that $\lambda = \lambda_0$ remains fixed and that the true probability of success is $p(\lambda_0, \theta_0) \in (\inf p, \sup p)$. With probability one the posterior beliefs $K_{t,i}$, $G_{t,i}$ and $F_{t,i}$ converge weakly to limit distributions $K_{\infty,i}$, $G_{\infty,i}$ and $F_{\infty,i}$ such that*

1. *$K_{\infty,1} = K_{\infty,2} = \delta_{p(\lambda_0,\theta_0)}$;*
2. *$G_{\infty,1}$ and $G_{\infty,2}$ admit densities $g_{\infty,1}$ and $g_{\infty,2}$ that satisfy $g_{\infty,1} \preceq g_{\infty,2}$.*
3. *$F_{\infty,1}$ and $F_{\infty,2}$ admit densities $f_{\infty,1}$ and $f_{\infty,2}$ that satisfy $f_{\infty,1} \succeq f_{\infty,2}$.*

Note that, unlike Proposition 1, Proposition 2 is independent of the nature of the interaction between λ and θ.

Proposition 2 has the following behavioral implications. Individuals who perform a task in a stable environment form correct limiting beliefs about their future outcomes in this environment. All the behavioral distortions associated with initial overconfidence (e.g., incorrect effort investment, excess entry) therefore vanish asymptotically: in the limit, the individuals make decisions based on accurate expectations of the consequences and obtain the maximum possible payoffs. However, since overconfidence is not eliminated by experimentation, an incorrect self-assessment affects

decisions in any new environment, irrespective of the amount of experimentation previously performed: agent 1 is more optimistic than agent 2 regarding the future outcomes if a new value of λ is drawn while θ is kept constant.

Let us now contrast this result to the case where a new value for λ is drawn every m periods according to the density g_0 and independently of the past history. This assumption represents situations where individuals are regularly exposed to new external conditions for reasons that are outside their control: automatic job rotation, turnover in a team, beginning of a new academic year with different professors, evaluation of their performance by different individuals, etc. In the long run, blaming external factors for the disappointing empirical success rate is no longer credible since the individual has been operating in many different environments, and he must therefore admit that he was himself responsible for the outcomes all along. Overconfidence is therefore entirely eliminated asymptotically.

Proposition 3. *Suppose that a new value of λ is drawn independently every m periods and that the true ability of the agent is $\theta_0 \in (\underline{\theta}, \bar{\theta})$. With probability one the posterior beliefs $F_{t,i}$ converge weakly to limit distributions $F_{\infty,i}$ such that $F_{\infty,1} = F_{\infty,2} = \delta_{\theta_0}$.*

Together, Propositions 2 and 3 establish that the possibility of overconfidence in the long run depends on the stability of the environment. Unrealistic levels of self-confidence can persist even for Bayesian learners: for instance, a worker who performs the same job for a long time can remain overconfident by blaming his colleagues, or the firm more generally, for the disappointing success rate. The model predicts that exogenous variation in the external conditions fosters realism about one's ability. In the next subsection we study the joint evolution of beliefs and behavior in a situation where the stability of external conditions is an endogenous feature resulting from the agents' decisions.

4.2 Active learning

We now study the joint evolution of beliefs and experimentation decisions. Our objective is to analyze the conditions under which individuals'

decisions endogenously generate enough data to allow them to overcome the identification challenge faced in stable environments and to learn their true ability in the long run.

We incorporate the model into an infinite-horizon bandit problem (Berry and Fristedt, 1985). On each date t, the agent performs the task, observes the outcome, and decides whether to stay in the current conditions or to opt out and start performing the activity in a randomly selected new environment. For instance, a manager decides whether to replace the current employees; a worker chooses whether to look for a new position in another company; a married individual decides whether to divorce and marry a new partner; a student chooses whether to persevere in their current field or reorient their educational choices, etc. This decision is conditioned by the decision-maker's beliefs about his ability and by his beliefs about the type of his current environment, since both dimensions determine the payoff that the agent expects from switching to a new environment.

We impose the simplest information structure that keeps the analysis of the two-dimensional bandit tractable while maintaining the key properties of the updating problem considered in the general model. Agents are either high-skilled ($\theta = \overline{\theta}$) or low-skilled ($\theta = \underline{\theta}$), and environments are either favorable ($\lambda = \overline{\lambda}$) or unfavorable ($\lambda = \underline{\lambda}$). We maintain the identification issue at the core of the model by assuming that $p(\underline{\lambda}, \overline{\theta}) = p(\overline{\lambda}, \underline{\theta})$.[7] To simplify the notation we write $p_l = p(\underline{\lambda}, \underline{\theta}), p_m = p(\underline{\lambda}, \overline{\theta}) = p(\overline{\lambda}, \underline{\theta})$ and $p_h = p(\overline{\lambda}, \overline{\theta})$.

The agent initially attaches a weight $q_0 \in (0, 1)$ to the state $\overline{\theta}$. An increase in q_0 can thus be interpreted as an increase in the individual's initial self-confidence. The individual faces an infinite and countable number of different environments. All conditions look similar ex ante: the qualities of the environments are independent and identically distributed, so that any given environment has a probability $\nu \in (0, 1)$ of being of quality $\overline{\lambda}$. We make no assumptions on p other than $0 < p_l < p_m < p_h < 1$.

On each date t, the agent chooses an environment and obtains the outcome π_t. We assume that quitting an environment is irreversible: if an environment has been tried and discarded by the agent, it is no longer available. This assumption entails a loss of generality since individuals might

[7] Footnote 8 states how our results are modified if one relaxes this assumption.

find it optimal to come back to a previously tried environment, but this restriction is inessential for our main result and simplifies the exposition. Since all untried environments look identical to the agent, we therefore formulate the experimentation problem as a two-armed bandit: arm 1 consists in staying in one's current conditions, while arm 2 consists in switching to a new environment. We say that the agent *experiments* if he decides to pull arm 2.

The agent is a risk-neutral discounted expected-utility maximizer with a discount factor $\delta < 1$. A history is a finite sequence $h_t = [(\sigma_0, \pi_0), \cdots, (\sigma_t, \pi_t)]$, where $\sigma_t \in \{1, 2\}$ denotes the identity of the arm selected at date t and $\pi_t \in \{0, 1\}$ denotes the Bernoulli outcome at date t. A strategy is an infinite sequence $\sigma = [\sigma_0, \sigma_1(\pi_0 = 1), \sigma_1(\pi_0 = 0), \cdots]$ that specifies which arm is selected by the agent initially and after any finite history.

The individual faces a trade-off between exploration and exploitation. The choice of an arm at date t is governed by two concerns: first, maximizing the immediate probability of success; second, gaining information about the quality of the current environment λ and about the agent's own ability θ.

In the Appendix we show with standard arguments that an optimal strategy exists and that the value function V of the decision problem is well-defined and characterized by a Bellman equation. However, solving this decision problem with two-dimensional learning is not feasible with the standard tools and results from the literature on bandit problems (e.g., Gittins indices), as the arms are correlated: the uncertain parameter θ governs the rewards to both arms. We therefore characterize the optimal behavior only in the case of a myopic agent ($\delta = 0$) who does not value experimentation, and we then proceed to show that our main result on asymptotic beliefs and behavior extends to the general case of a patient decision-maker ($\delta > 0$).

4.2.1 Myopic behavior

In this section we assume that $\delta = 0$, and thus the agent maximizes the immediate expected reward. Let $B_t^{n_t}(p) = p^{n_t}(1-p)^{t-n_t}$.

Experimentation decisions Suppose that the agent selects arm 2 at some date t_0. The weight q that he attaches to $\overline{\theta}$ at date t_0 is a sufficient statistic for the information acquired so far. Suppose that the agent then stays t periods and receives n_t successes in the new conditions.

Since the agent is myopic, it is optimal to select the arm which delivers the greatest probability of success. After some algebra, arm 2 is optimal if and only if

$$(1-q)(p_m-p_l)[B_t^{n_t}(p_m)-B_t^{n_t}(p_l)]+q(p_h-p_m)[B_t^{n_t}(p_h)-B_t^{n_t}(p_m)] \leq 0. \quad (1)$$

Condition 1 has two important properties. First, for fixed q, it is satisfied if and only if n_t is lower than some threshold. As the intuition suggests, the decision-maker thus opts out when a disappointing sequence of outcomes has led him to form pessimistic beliefs regarding the quality of his current environment relative to the average external conditions. Second, for fixed n_t it is satisfied if and only if q is larger than some threshold. A decision-maker who initially perceives a larger q than what is realistic tends to see the grass as being greener on the other side of the fence: this belief endogenously encourages opting out. In contrast, underconfident decision-makers are too easily satisfied with their external conditions and experiment too little relative to the payoff-maximizing behavior.

Overconfidence We now turn to analyzing the possible asymptotic scenarios resulting from these endogenous experimentation choices. Suppose first that the agent's true type is $\underline{\theta}$, and that he starts the game with a confident prior belief q_0 close to but different from one. Suppose that the first environment tried is of type $\underline{\lambda}$. If the agent stays long enough in this environment, his success rate converges almost surely to p_l. The agent then learns his own ability and the type of environment perfectly. Knowing that the current conditions are unfavorable, he therefore opts out in finite time. Suppose now that the first environment is instead of type $\overline{\lambda}$, in which case the success rate converges almost surely to p_m. Asymptotically, Condition 1 is then equivalent to

$$p_m \leq q_0 p_h + (1-q_0)p_l. \quad (2)$$

Equation 2 is satisfied if and only if q_0 is large enough. Intuitively, by performing the task infinitely often in the current environment, the agent progressively learns that the future success rate in that environment equals p_m; since q_0 is close to one, the agent attributes his empirical success rate to the fact that the current conditions are of type $\underline{\lambda}$. Since the agent expects a probability of success larger than p_m in the average environment, he therefore decides to switch from the current environment in finite time. In both cases ($\lambda = \underline{\lambda}, \lambda = \overline{\lambda}$), he opts out in finite time after forming pessimistic beliefs about the environment but also revising his level of self-confidence downwards.

The same arguments apply to the analysis of the continuation history that the agent receives after switching to a new environment. Environments of type $\underline{\lambda}$ are thus always left in finite time, whereas environments of type $\overline{\lambda}$ are left in finite time if the agent's self-confidence is large enough for Condition 2 to be satisfied. Yet, over time, the individual's endogenous experimentation effort provides more information about the true value of θ, in line with Proposition 3. By performing the task in variable external conditions, the individual gradually realizes that his ability is lower than he thought, until his level of self-confidence q becomes small enough to satisfy $p_m > qp_h + (1-q)p_l$. At that point, if the individual stays long enough in an environment of type $\overline{\lambda}$, he expects a reward close to p_m in the current environment, and a reward lower than p_m in an average environment, due to his low self-confidence. He therefore prefers to stop experimenting in the current conditions. As we argue in Proposition 4 below, this scenario happens with probability one. In the long run, learning is incomplete since the individual stops experimenting in finite time and therefore cannot disentangle the states $(\underline{\lambda}, \overline{\theta})$ and $(\overline{\lambda}, \underline{\theta})$. However, learning is adequate in the sense that the individual eventually settles into an environment of type $\overline{\lambda}$ and receives the highest possible long-run payoff p_m for an agent of ability $\underline{\theta}$.

Underconfidence Suppose now that the individual has true ability $\overline{\theta}$ and initial self-confidence $q_0 \in (0,1)$. If the first environment is of type $\overline{\lambda}$, Condition 1 can be violated at each period with positive probability. Asymptotically, the individual learns that he is high-skilled and that the

environment is favorable. Learning is thus complete in both dimensions.

However, if the first environment is of type $\underline{\lambda}$ and if q_0 is small enough to revert Equation 2, Condition 1 can also be violated at each period. The individual then stops experimenting since he attributes his high success rate p_m to the quality of the external conditions rather than to his own merit. This decision prevents him from receiving further information, and from revising his beliefs about his ability upwards. Learning is then not only incomplete but also inadequate: by experimenting more, this individual would have achieved a long-run payoff equal to p_h, but instead finds himself trapped in an inferior environment, receiving a suboptimal long-run success rate equal to p_m. In contrast to the case of an overconfident agent, the initial miscalibration in prior beliefs thus generates asymptotic inefficiencies and incorrect decisions.

Note that this scenario happens with positive probability for the first environment if q_0 is small, but it also happens with positive probability asymptotically for any value of $q_0 \in (0,1)$: even an initially confident individual might fall into the underconfidence trap if his first attempts are, unluckily, unsuccessful, up to the point where his self-confidence q falls below the threshold defined by Condition 2.

4.2.2 Limiting beliefs

We now generalize these asymptotic results by relaxing the assumption $\delta = 0$. We write q_t for the posterior weight that the individual ascribes to the state $\overline{\theta}$.

Proposition 4. *The individual stops experimenting in finite time almost surely. In addition, for any $q_0 \in (0,1)$,*

1. *If the true ability of the agent is $\underline{\theta}$, then with probability one the last environment is of type $\overline{\lambda}$. Moreover, there exists a threshold $\bar{q} \in (0,1)$ such that q_t converges almost surely to a limit $q_\infty \in (0, \bar{q}]$.*

2. *If the true ability of the agent is $\overline{\theta}$, there exists a threshold $\underline{q} \in (0,1)$ such that only the two following scenarios have a positive probability:*

   (a) *The last environment is of type $\overline{\lambda}$ and q_t converges to 1.*

(b) The last environment is of type $\underline{\lambda}$ and q_t converges to some limit $q_\infty \in (0, \underline{q}]$.

Proposition 4 establishes two main results. First, for any value of θ, with probability one the agent decides in finite time to stop experimenting, a common finding from the literature on active experimentation (Aghion et al., 1991; Easley and Kiefer, 1988; Brezzi and Lai, 2000). Intuitively, infinite experimentation would lead the individual's self-confidence q_t to converge to zero or one, depending on his true ability. In the limit where $q_t \sim 0$ or $q_t \sim 1$, the individual's problem consists of sampling the possible external conditions until finding an environment of type $\overline{\lambda}$ in which to stay. This happens in finite time almost surely. The assumption that the prior is correctly specified is crucial for this result. Indeed, an individual with initial self-confidence $q_0 = 1$ but true ability $\underline{\theta}$ would perpetually experiment with probability one.

Second, if the agent starts with overconfident beliefs, his learning process is incomplete but adequate, meaning that the agent eventually finds good conditions and obtains the maximum possible asymptotic payoff. In contrast, an agent who starts with underconfident prior beliefs faces a positive probability of making suboptimal choices forever, attributing his surprisingly large success rate to extrinsic characteristics instead of taking credit for it. The mistakes induced by miscalibrated prior beliefs are therefore not symmetric: overconfidence inflicts a transient cost to the agent by inducing him to over-experiment, but this distortion disappears in the long run. Underconfidence generates a persistent distortion that might survive endogenous experimentation.[8]

[8] If the agent does not face any identification issue, i.e., if $p(\underline{\lambda}, \overline{\theta}) \neq p(\overline{\lambda}, \underline{\theta})$, then it remains true that experimentation stops in finite time almost surely. However, asymptotic learning is then both adequate and complete since the long-run outcomes obtained in any stable environment perfectly inform the decision-maker about his true ability.

5 Discussion and conclusion

5.1 Interpretation of the parameters

In this section we discuss the interpretation and the predictions of the model in different contexts.

The nature of the activity The variable λ can be viewed as the nature of the task, or its intrinsic difficulty. The form of p then reflects whether easier activities are more or less informative than difficult activities about the ability of the agent. The model predicts that overconfident individuals misperceive the difficulty of the task, and that they are also more prone to experimenting variable activities since they are easily disappointed with the outcomes received at a given task.

Just world The variable λ can also be viewed as a measure of the extent to which people are responsible for their own outcomes, as opposed to luck or other uncontrollable factors. If p is strictly log-spm, a low-λ environment can for instance refer to a situation where some social groups are discriminated against because of fixed individual traits (gender, ethnicity, socio-economic background), in which case their talent can do little to compensate for the fundamental inequity. This contrasts with a high-λ environment, which describes a society where people obtain their just deserts. Individuals' beliefs about λ can then be understood as their *locus of control*.

The model predicts that successful individuals understate the importance of socio-economic rigidities: believing in a "just world" (Lerner, 1980; Bénabou and Tirole, 2006), they attribute others' misfortunes to their dispositions, such as their supposed lack of ability or willpower (Corollary 1). Conversely, they overestimate the merits of their high-achieving peers. Less successful individuals underestimate the fairness of the social mobility system and believe that the outcomes of others do not reflect their dispositions.

If citizens factor distributive-justice concerns into judgments about redistributive policies (Alesina and Angeletos, 2005), their view of the nature of social competition determines their political preferences. Our model pre-

dicts that, even supposing that they are only motivated by concerns for social justice, the rich are too prone to advocate pro-market policies and low levels of redistribution, whereas the reverse holds for the poor. The experiment by Deffains et al. (2016) offers evidence consistent with this theory. After performing a real effort task whose returns are uncertain, subjects tend to choose lower redistribution levels for their peers if their own performance lies in the top half of the distribution.

The model can also account for the effect of "role models" whose accomplishments in various domains (sport, science, business, etc.) are frequently showcased by popular culture as a source of inspiration. The exposure to success stories is thought of as a way to promote faith in the long-term returns to effort, especially for groups who face unfavorable conditions (ethnic minorities, female scientists, etc.). Interestingly, this strategy sometimes backfires (Lockwood and Kunda, 1997). In our model, an individual who observes a successful role model revises his beliefs about λ upwards, which tends to fosters his belief that succeeding in his conditions is possible. However, if the agent already has some experience at the task, receiving information about λ also leads him to reexamine his own history and to update his self-confidence. The direction of this effect depends on the success ratio and specific history, as Proposition 1 suggests. As an example, if $1-p$ is log-sbm, an individual who has received disappointing outcomes so far realizes, upon observing a successful peer, that the environment is more favorable than he thought, but that his own ability is lower than he thought. The overall impact on the perceived probability of success depends on which of these two effects dominates.

Production externalities In a team production context, λ can describe the performance, intentions, or skills of the decision-maker's co-workers. The model predicts that attributions of merit and blame in teams depend on the nature of the strategic relationship between the co-workers' contributions, that exogenous variation in external conditions fosters learning about one's ability, and that overconfident workers are more prone to change jobs or teams if they have the opportunity to do so.

Information structure Suppose that the agent receives a sequence of informative signals about his own ability, and that the correlation between signals is uncertain ex ante. The model predicts that an overconfident individual overestimates the correlation when receiving a series of bad news, and underestimates it when receiving a sequence of good news. For instance, consider a student or worker who receives feedback on a project from two advisers. The advisers might form their judgment independently, or the second adviser might simply consult the first adviser's opinion without properly analyzing the project on his own. The informativeness of the feedback is greater in the former case. The model predicts that the student overestimates the independence of his advisers' judgments if they both report favorably on the project, and overestimates their correlation if they both express adverse opinions.

Self-control In another interpretation of the model, θ represents the individual's capacity to exert self-control and to stick to his contingent plans, while λ measures the extent to which external conditions are intrinsically tempting. The model predicts that naive agents who repeatedly succumb to temptation blame persistent external conditions instead of acknowledging their self-control issues. For instance, an individual who fails at quitting smoking might explain his difficulties by the fact that he has recently gone through a stressful period at work. Such an individual might therefore maintain the optimistic belief that quitting smoking will be easy once he faces more favorable conditions. Becoming sophisticated about one's self-control and making correct predictions about one's behavior in new situations requires exposure to a variety of external conditions. Interestingly, our model predicts that naiveté is self-correcting through endogenous experimentation while pessimism is self-confirming, which adds to the puzzle of the persistence of naiveté in the field.[9]

[9]This observation parallels a result by Ali (2011) obtained with a different mechanism. Ali (2011) shows that naiveté is self-correcting due to the insufficient take-up of commitment devices, while underconfidence is self-confirming as it can lead the individual to overcommit and stop receiving information about his self-control.

5.2 Conclusion

This paper shows that overconfidence generates distortions in the process by which individuals learn about their environment. Overconfidence causes a self-serving bias in the attribution of blame and merit (Proposition 1). This distortion leads individuals to make incorrect causal attributions of others' outcomes (Corollary 1) and to form incorrect beliefs about the returns to human capital (Corollary 2). Trying out different environments is a necessary and sufficient condition for overconfidence to vanish in the long run (Propositions 2 and 3). Since overconfidence fosters experimentation in different environments, it is self-correcting in the long run whereas underconfidence is self-confirming (Proposition 4).

We conclude by mentioning two directions in which the analysis can be extended. First, in some situations, observing the outcomes received by peers exposed to the same external conditions would provide additional information to the individual about λ. In general, social learning might therefore mitigate the identification challenge, but our comparative statics results would survive, provided that the individual does not have access to an infinite quantity of observations generated in his environment. More importantly, the agent's own inferences might prevent him from learning efficiently from observing his peers. For instance, an overconfident individual who has failed repeatedly would attribute others' outcomes to luck rather than to their own merit, a belief which would not easily be dismissed by subsequent observations.

Second, the individual decision problem can be used as a foundation to study the strategic interaction between an agent and a principal or an audience. The agent might be motivated by the opportunity to signal his ability to third parties, which would influence the type of environment or tasks into which he strategically self-selects. Principals might strategically release information about the difficulty of the task to maintain the agent's optimism and motivation, or sabotage the agents' self-esteem to reduce their willingness to opt out of the relation. More generally, the interaction between a principal who can influence the nature of the task or the environment and an agent prone to misperceptions raises interesting and important questions, that we leave for future research.

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2016). Fragility of asymptotic agreement under Bayesian learning. *Theoretical Economics 11*(1), 187–225.

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991). Optimal learning by experimentation. *Review of Economic Studies 58*(4), 621–654.

Alesina, A. and G.-M. Angeletos (2005). Fairness and redistribution. *American Economic Review 95*(4), 960–980.

Ali, S. N. (2011). Learning self-control. *Quarterly Journal of Economics 126*(2), 857–893.

Arkin, R. M. and G. M. Maruyama (1979). Attribution, affect, and college exam performance. *Journal of Educational Psychology 71*(1), 85.

Banks, J. S. and R. K. Sundaram (1992). Denumerable-armed bandits. *Econometrica 60*(5), 1071–1096.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics 116*(1), 261–292.

Baumeister, R. F., T. F. Heatherton, and D. M. Tice (1993). When ego threats lead to self-regulation failure: Negative consequences of high self-esteem. *Journal of Personality and Social Psychology 64*(1), 141.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3), 871–915.

Bénabou, R. and J. Tirole (2006). Belief in a just world and redistributive politics. *Quarterly Journal of Economics 121*(2), 699–746.

Benjamin, D., M. Rabin, and C. Raymond (2016). A model of non-belief in the law of large numbers. *Journal of the European Economic Association 14*(2), 515–544.

Benoît, J. and J. Dubra (2011). Apparent overconfidence. *Econometrica 79*(5), 1591–1625.

Benoît, J.-P., J. Dubra, and D. A. Moore (2015). Does the better-than-average effect show that people are overconfident?: Two experiments. *Journal of the European Economic Association 13*(2), 293–329.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics 37*(1), 51–58.

Berry, D. A. and B. Fristedt (1985). *Bandit Problems: Sequential Allocation of Experiments*. Berlin: Springer-Verlag.

Billett, M. T. and Y. Qian (2008). Are overconfident CEOs born or made? evidence of self-attribution bias from frequent acquirers. *Management Science 54*(6), 1037–1051.

Brezzi, M. and T. L. Lai (2000). Incomplete learning from endogenous data in dynamic allocation. *Econometrica 68*(6), 1511–1516.

Bunke, O. and X. Milhaud (1998). Asymptotic behavior of Bayes estimates under possibly incorrect models. *The Annals of Statistics 26*(2), 617–644.

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini (2013). Overconfidence and social signalling. *Review of Economic Studies 80*(3), 949–983.

Camerer, C. F. and D. Lovallo (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review 89*(1), 306–318.

Clance, P. R. and S. A. Imes (1978). The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice 15*(3), 241.

Deffains, B., R. Espinosa, and C. Thöni (2016). Political self-serving bias and redistribution. *Journal of Public Economics 134*, 67–74.

Dubra, J. (2004). Optimism and overconfidence in search. *Review of Economic Dynamics 7*(1), 198–218.

Easley, D. and N. M. Kiefer (1988). Controlling a stochastic process with unknown parameters. *Econometrica 56*(5), 1045–1064.

Eil, D. and J. Rao (2011). The good news–bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics 3*(2), 114–48.

Esponda, I. and D. Pouzo (2016). Berk–Nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica 84*(3), 1093–1130.

Fudenberg, D., G. Romanyuk, and P. Strack (2017). Active learning with a misspecified prior. *Theoretical Economics 12*(3), 1155–1189.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis* (Third ed.). London: CRC Press.

Gervais, S. and T. Odean (2001). Learning to be overconfident. *Review of Financial Studies 14*(1), 1–27.

Gottlieb, D. (2017). Will you never learn? self deception and biases in information processing. Working paper.

Grossman, Z. and D. Owens (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization 84*(2), 510–524.

Grubb, M. D. (2009). Selling to overconfident consumers. *American Economic Review 99*(5), 1770–1807.

Hastorf, A. H., D. J. Schneider, and J. Polefka (1970). *Person perception.* Addison-Wesley.

Heidhues, P., B. Kőszegi, and P. Strack (2018). Unrealistic expectations and misguided learning. *Econometrica 86*(4), 1159–1214.

Hoelzl, E. and A. Rustichini (2005). Overconfident: Do you put your money on it? *The Economic Journal 115*(503), 305–318.

Johnson, T. J., R. Feigenbaum, and M. Weiby (1964). Some determinants and consequences of the teacher's perception of causation. *Journal of Educational Psychology 55*(5), 237.

Kőszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association 4*(4), 673–707.

Kruger, J., P. D. Windschitl, J. Burrus, F. Fessel, and J. R. Chambers (2008). The rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology 44*(2), 220–232.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin 108*(3), 480.

Lau, R. R. and D. Russell (1980). Attributions in the sports pages. *Journal of Personality and Social Psychology 39*(1), 29.

Lerner, M. J. (1980). *The belief in a just world*. Berlin: Springer-Verlag.

Libby, R. and K. Rennekamp (2012). Self-serving attribution bias, overconfidence, and the issuance of management forecasts. *Journal of Accounting Research 50*(1), 197–231.

Lockwood, P. and Z. Kunda (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology 73*(1), 91.

Malmendier, U. and G. Tate (2005). CEO overconfidence and corporate investment. *Journal of Finance 60*(6), 2661–2700.

Möbius, M. M., M. Niederle, P. Niehaus, and T. Rosenblat (2013). Managing self-confidence: Theory and experimental evidence. Working paper.

Mitrinovic, D. S., J. Pecaric, and A. M. Fink (2013). *Classical and New Inequalities in Analysis*, Volume 61. Berlin: Springer-Verlag.

Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes 102*(1), 42–58.

Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological Review 115*(2), 502.

Park, Y. J. and L. Santos-Pinto (2010). Overconfidence in tournaments: evidence from the field. *Theory and Decision 69*(1), 143–166.

Rabin, M. and J. L. Schrag (1999). First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics 114*(1), 37–82.

Ross, M. and F. Sicoly (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology 37*(3), 322.

Santos-Pinto, L. and J. Sobel (2005). A model of positive self-image in subjective assessments. *American Economic Review 95*(5), 1386–1402.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association 12*(6), 1423–1452.

Sonnak, C. and T. Towell (2001). The impostor phenomenon in British university students: Relationships between self-esteem, mental health, parental rearing style and socioeconomic status. *Personality and Individual Differences 31*(6), 863–874.

Stewart, A. E. (2005). Attributions of responsibility for motor vehicle crashes. *Accident Analysis & Prevention 37*(4), 681–688.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica 47*(2), 143–148.

Thaler, R. H. (2000). From homo economicus to homo sapiens. *Journal of Economic Perspectives 14*(1), 133–141.

Van den Steen, E. (2004). Rational overoptimism (and other biases). *American Economic Review 94*(4), 1141–1151.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology 39*(5), 806.

Zabojnik, J. (2004). A model of rational bias in self-assessments. *Economic Theory 23*(2), 259–282.

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality 47*(2), 245–287.

Appendix

We write $\mathcal{L}_{t,n}(\lambda,\theta) = p(\lambda,\theta)^n(1-p(\lambda,\theta))^{t-n}$ for the (normalized) likelihood function and we skip the variables (λ,θ) when it is not confusing. We will make extensive use of the continuous version of Chebyshev's sum inequality, restated below (see Mitrinovic et al., 2013, , chapter 9).

Lemma A.1. *Consider a compact interval $X \subset \mathbb{R}$. If $f, g : X \to \mathbb{R}$ are integrable functions, both nondecreasing or both nonincreasing, and $h : X \to \mathbb{R}_+$ is integrable, then*

$$\int_X f(x)g(x)h(x)dx \int_X h(x)dx \geq \int_X f(x)h(x)dx \int_X g(x)h(x)dx. \tag{A.1}$$

If f is nonincreasing and g is nondecreasing, inequality A.1 is reversed.

A Proofs of Section 3

A.1 Proof of Claim 1

Supposed that the agent has tried out the activity in m_t different environments up to date t, and let $j = 1, \cdots, m_t$ be an index for the identity of the environments. Let n_j be the number of successes in the environment indexed by j and t_j the total number of attempts in this environment. Given the history $h_t = (t_1, n_1, \cdots, t_{m_t}, n_{m_t})$, Bayes' rule yields

$$f_{t,h_t,i}(\theta) = \frac{f_{0,i}(\theta)\prod_{j=1}^{m_t}\int_\Lambda \mathcal{L}_{t_j,n_j}(\lambda_j,\theta)dG_0(\lambda_j)}{\int_\Theta dF_{0,i}(\theta')\prod_{j=1}^{m_t}\int_\Lambda \mathcal{L}_{t_j,n_j}(\lambda_j,\theta')dG_0(\lambda_j)},$$

and therefore

$$\frac{f_{t,h_t,1}(\theta)}{f_{t,h_t,2}(\theta)} = \frac{f_{0,1}(\theta)}{f_{0,2}(\theta)}\frac{\int_\Theta dF_{0,2}(\theta')\prod_{j=1}^{m_t}\int_\Lambda \mathcal{L}_{t_j,n_j}(\lambda_j,\theta')dG_0(\lambda_j)}{\int_\Theta dF_{0,1}(\theta')\prod_{j=1}^{m_t}\int_\Lambda \mathcal{L}_{t_j,n_j}(\lambda_j,\theta')dG_0(\lambda_j)}$$

is nondecreasing in θ.

A.2 Proof of Proposition 1

In the main text we prove Proposition 1.2. In the footnotes we explain how to adapt the arguments to prove Propositions 1.1 and 1.3.

The proof proceeds in two steps. First, we show that the likelihood ratio $\mathcal{L}_{t,n_t}(\lambda_1,\theta)/\mathcal{L}_{t,n_t}(\lambda_2,\theta)$ is nondecreasing in θ for any $\lambda_1 > \lambda_2$ whenever the success rate n_t/t is large enough, and nonincreasing whenever the success rate is small enough.[10,11] This property is straightforward for fixed (λ_1,λ_2); the crux of the proof is to obtain bounds that are uniform in (λ_1,λ_2). The second step consists of an application of Lemma A.1.

Claim A.2. *Suppose that p is strictly log-spm and consider the domain $\mathcal{D} = \{(\lambda_1,\lambda_2,\theta) \in \Lambda^2 \times \Theta \mid \lambda_1 > \lambda_2\}$ and the function ψ defined on $\mathcal{D}$ by*

$$\psi(\lambda_1,\lambda_2,\theta) = \frac{\mathcal{L}_{t,n_t}(\lambda_1,\theta)}{\mathcal{L}_{t,n_t}(\lambda_2,\theta)}.$$

There exist $\alpha_0,\beta_0 \in (0,1)$ such that if $n_t/t \geq \alpha_0$ (respectively $n_t/t \leq \beta_0$), ψ is nondecreasing (respectively nonincreasing) in θ for any $\lambda_1 > \lambda_2$.

Proof. The function ψ is continuously differentiable and ψ_θ satisfies

$$\frac{\psi_\theta(\lambda_1,\lambda_2,\theta)}{\psi(\lambda_1,\lambda_2,\theta)} = n_t\Big[\frac{p_\theta(\lambda_1,\theta)}{p(\lambda_1,\theta)} - \frac{p_\theta(\lambda_2,\theta)}{p(\lambda_2,\theta)}\Big] - (t-n_t)\Big[\frac{p_\theta(\lambda_1,\theta)}{1-p(\lambda_1,\theta)} - \frac{p_\theta(\lambda_2,\theta)}{1-p(\lambda_2,\theta)}\Big]. \tag{A.2}$$

Consider the functions ζ and ξ defined on $\mathcal{D}$ by

$$\zeta(\lambda_1,\lambda_2,\theta) = \frac{p_\theta(\lambda_1,\theta)}{p(\lambda_1,\theta)} - \frac{p_\theta(\lambda_2,\theta)}{p(\lambda_2,\theta)}$$

and

$$\xi(\lambda_1,\lambda_2,\theta) = \frac{p_\theta(\lambda_1,\theta)}{1-p(\lambda_1,\theta)} - \frac{p_\theta(\lambda_2,\theta)}{1-p(\lambda_2,\theta)}.$$

Since p is strictly log-spm the functions $\lambda \to p_\theta(\lambda,\theta)/p(\lambda,\theta)$ and $\lambda \to p_\theta(\lambda,\theta)/(1-p(\lambda,\theta))$ are increasing in λ for any θ. Thus on the domain $\mathcal{D}$ the functions ζ and ξ take only positive values. In addition (we drop the dependence

[10] If p and $1-p$ are log-sbm we show that the likelihood ratio is nonincreasing in θ for any $\lambda_1 > \lambda_2$ and any sequence of outcomes.

[11] If $1-p$ is strictly log-spm we show that the likelihood ratio is nonincreasing in θ for any $\lambda_1 > \lambda_2$ whenever the success rate is large enough, and nondecreasing whenever the success rate is small enough.

in (λ, θ) of all functions to lighten the notational burden),

$$\lim_{\epsilon \to 0} \frac{\xi(\lambda+\epsilon, \lambda, \theta)}{\zeta(\lambda+\epsilon, \lambda, \theta)} = \frac{\frac{p_{\lambda\theta}(1-p) + p_\lambda p_\theta}{(1-p)^2}}{\frac{p_{\lambda\theta} p - p_\lambda p_\theta}{p^2}} > 0.$$

Hence the function ξ/ζ can be extended by continuity to the compact domain $\bar{\mathcal{D}} = \{(\lambda_1, \lambda_2, \theta) \in \Lambda^2 \times \Theta \mid \lambda_1 \geq \lambda_2\}$ and its extension takes positive values only. This proves that ξ/ζ admits a positive lower bound $\inf \xi/\zeta$ and a positive upper bound $\sup \xi/\zeta$ on $\mathcal{D}$.

Let

$$\alpha_0 = \frac{\sup \frac{\xi}{\zeta}}{1 + \sup \frac{\xi}{\zeta}} \text{ and } \beta_0 = \frac{\inf \frac{\xi}{\zeta}}{1 + \inf \frac{\xi}{\zeta}}.$$

It is clear that $\alpha_0 \in (0,1)$ and $\beta_0 \in (0,1)$. In addition, for any (n_t, t) such that $n_t/t \geq \alpha_0$ we have

$$\frac{n_t}{t - n_t} \geq \sup \frac{\xi}{\zeta},$$

which implies by Equation A.2 that $\psi_\theta \geq 0$ on $\mathcal{D}$, i.e. that ψ is nondecreasing in θ. If $n_t/t \leq \beta_0$ we have $n_t/(t-n_t) \leq \inf \xi/\zeta$ and therefore ψ is nonincreasing in θ.[12,13] □

To complete the proof, suppose first that $n_t/t \geq \alpha_0$ defined in Claim A.2. Take any $\lambda_1 > \lambda_2$. The function $f_{0,1}/f_{0,2}$ is nondecreasing in θ, and, by Claim A.2, the function $\psi(\lambda_1, \lambda_2, \theta)$ is also nondecreasing in θ. Lemma A.1 delivers

$$\Big[\int_\Theta \frac{\mathcal{L}_{t,n_t}(\lambda_1,\theta)}{\mathcal{L}_{t,n_t}(\lambda_2,\theta)} \frac{f_{0,1}(\theta)}{f_{0,2}(\theta)} \mathcal{L}_{t,n_t}(\lambda_2,\theta) dF_{0,2}(\theta)\Big]\Big[\int_\Theta \mathcal{L}_{t,n_t}(\lambda_2,\theta) dF_{0,2}(\theta)\Big] \geq \quad \text{(A.3)}$$
$$\Big[\int_\Theta \frac{\mathcal{L}_{t,n_t}(\lambda_1,\theta)}{\mathcal{L}_{t,n_t}(\lambda_2,\theta)} \mathcal{L}_{t,n_t}(\lambda_2,\theta) dF_{0,2}(\theta)\Big]\Big[\int_\Theta \frac{f_{0,1}(\theta)}{f_{0,2}(\theta)} \mathcal{L}_{t,n_t}(\lambda_2,\theta) dF_{0,2}(\theta)\Big].$$

[12] If p and $1-p$ are log-sbm, ζ is nonpositive whereas ξ is nonnegative, and thus ψ_θ is nonpositive for any (n_t, t), which proves that ψ is nonincreasing in θ.

[13] If $1-p$ is strictly log-spm, the functions ζ and ξ take negative values only and ξ/ζ can be extended by continuity to $\bar{\mathcal{D}}$, where it takes positive values only. Define (α_1, β_1) similarly as (α_0, β_0) above. For any $n_t/t \geq \alpha_1$ the function ψ_θ is nonpositive, i.e. ψ is nonincreasing in θ. For any $n_t/t \leq \beta_1$ the function ψ is nondecreasing in θ.

Rearranging A.3 yields

$$\frac{\int_\Theta \mathcal{L}_{t,n_t}(\lambda_1,\theta)dF_{0,1}(\theta)}{\int_\Theta \mathcal{L}_{t,n_t}(\lambda_1,\theta)dF_{0,2}(\theta)} \geq \frac{\int_\Theta \mathcal{L}_{t,n_t}(\lambda_2,\theta)dF_{0,1}(\theta)}{\int_\Theta \mathcal{L}_{t,n_t}(\lambda_2,\theta)dF_{0,2}(\theta)},$$

which is simply

$$\frac{g_{t,n_t,1}(\lambda_1)}{g_{t,n_t,2}(\lambda_1)} \geq \frac{g_{t,n_t,1}(\lambda_2)}{g_{t,n_t,2}(\lambda_2)}. \tag{A.4}$$

Since Equation A.4 is true for any $\lambda_1 > \lambda_2$, $g_{t,n_t,1} \succeq g_{t,n_t,2}$. If $n_t/t \leq \beta_0$ then by Claim A.2 ψ is nonincreasing in θ, which implies that inequalities A.3 and A.4 are reversed, i.e. that $g_{t,n_t,1} \preceq g_{t,n_t,2}$.[14,15]

A.3 Proof of Corollary 1

We first observe that, whenever $p_\lambda p - p_\lambda p_\theta$ and $p_\lambda(1-p) + p_\lambda p_\theta$ have constant signs, then there exists α_2, β_2 such that $n_t/t \geq \alpha_2$ or $n_t/t \leq \beta_2$ implies that, for any $\tilde{\theta}_1 > \tilde{\theta}_2$, the functions $\lambda \to g_{t,n_t,1}(\lambda)/g_{t,n_t,2}(\lambda)$ and $\lambda \to \mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_1)/\mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_2)$ are both nonincreasing or both nondecreasing.

Case 1: p and $1-p$ are log-sbm Then Proposition 1 and a claim analogous to Claim A.2 (inverting the roles of λ and θ) show that, for any (n_t, t), $\lambda \to g_{t,n_t,1}(\lambda)/g_{t,n_t,2}(\lambda)$ and $\lambda \to \mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_1)/\mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_2)$ are both nonincreasing.

Case 2: p is strictly log-spm Then by a reasoning analogous to Claim A.2 it is possible to find α_2, β_2 such that

$$\frac{n_t}{t} \geq \alpha_2 \Rightarrow \frac{\mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_1)}{\mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_2)} \text{ and } \frac{g_{t,n_t,1}(\lambda)}{g_{t,n_t,2}(\lambda)} \text{ are nondecreasing}$$

$$\text{and } \frac{n_t}{t} \leq \beta_2 \Rightarrow \frac{\mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_1)}{\mathcal{L}_{t,n_t}(\lambda,\tilde{\theta}_2)} \text{ and } \frac{g_{t,n_t,1}(\lambda)}{g_{t,n_t,2}(\lambda)} \text{ are nonincreasing.}$$

[14] If p and $1-p$ are log-sbm the function ψ is nonincreasing in θ for any (n_t, t), and thus Equation A.4 is reversed for any (n_t, t).

[15] If $1-p$ is strictly log-spm, for any $n_t/t \geq \alpha_1$ the function ψ is nondecreasing in θ. Therefore by Lemma A.1 inequality A.3 is reversed, and hence condition A.4 is reversed as well, which proves that $g_{t,n_t,1} \preceq g_{t,n_t,2}$. The case $n_t/t \leq \beta_1$ is symmetric.

Case 3: $1-p$ is strictly log-spm Then by a reasoning analogous to Claim A.2 it is possible to find α_2, β_2 such that

$$\frac{n_t}{t} \geq \alpha_2 \Rightarrow \frac{\mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_1)}{\mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_2)} \text{ and } \frac{g_{t,n_t,1}(\lambda)}{g_{t,n_t,2}(\lambda)} \text{ are nonincreasing}$$
$$\text{and } \frac{n_t}{t} \leq \beta_2 \Rightarrow \frac{\mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_1)}{\mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_2)} \text{ and } \frac{g_{t,n_t,1}(\lambda)}{g_{t,n_t,2}(\lambda)} \text{ are nondecreasing.}$$

Suppose that $n_t/t \geq \alpha_2$ or $n_t/t \leq \beta_2$, where α_2 and β_2 are constructed above. Then by Lemma A.1, for any $\tilde{\theta}_1 > \tilde{\theta}_2$ we have

$$\int_\Lambda \mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_1) dG_{t,n_t,1}(\lambda) \int_\Lambda \mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_2) dG_{t,n_t,2}(\lambda) \geq$$
$$\int_\Lambda \mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_2) dG_{t,n_t,1}(\lambda) \int_\Lambda \mathcal{L}_{t,n_t}(\lambda, \tilde{\theta}_1) dG_{t,n_t,2}(\lambda)$$

which simplifies to

$$\frac{\tilde{f}_{t,n_t,1}(\tilde{\theta}_1)}{\tilde{f}_{t,n_t,2}(\tilde{\theta}_1)} \geq \frac{\tilde{f}_{t,n_t,1}(\tilde{\theta}_2)}{\tilde{f}_{t,n_t,2}(\tilde{\theta}_2)}$$

This proves that $\tilde{f}_{t,n_t,1} \succeq \tilde{f}_{t,n_t,2}$.

A.4 Proof of Corollary 2

Suppose first that p is strictly log-spm, which implies that the difference $p(\lambda, \theta_H) - p(\lambda, \theta_L)$ is nondecreasing in λ. Consider α_0, β_0 defined in Proposition 1. Suppose first that $n_t/t \geq \alpha_0$. By Proposition 1, $g_{t,n_t,1} \succeq g_{t,n_t,2}$, which implies that $g_{t,n_t,1}$ first-order stochastically dominates $g_{t,n_t,2}$. Thus,

$$\int_\Lambda [p(\lambda, \theta_H) - p(\lambda, \theta_L)] dG_{t,n_t,1}(\lambda) \geq \int_\Lambda [p(\lambda, \theta_H) - p(\lambda, \theta_L)] dG_{t,n_t,2}(\lambda), \quad \text{(A.5)}$$

which is simply $\vartheta_{t,n_t,1} \geq \vartheta_{t,n_t,2}$. If $n_t/t \leq \beta_0$, $g_{t,n_t,1}$ is first-order stochastically dominated by $g_{t,n_t,2}$ and therefore the inequality is reversed. Defining $(\alpha_3, \beta_3) = (\alpha_0, \beta_0)$ completes the proof.

If $1-p$ is strictly log-spm the difference $p(\lambda, \theta_H) - p(\lambda, \theta_L)$ is nonincreasing in λ. For any $n_t/t \geq \alpha_1$ we have $g_{t,n_t,1} \preceq g_{t,n_t,2}$, and thus inequality A.5 remains true. For any $n_t/t \leq \beta_1$, $g_{t,n_t,1} \succeq g_{t,n_t,2}$ and the inequality is reversed. The result follows from defining $(\alpha_3, \beta_3) = (\alpha_1, \beta_1)$.

B Proofs of Section 4

B.1 Proof of Proposition 2

In the following we write $p_0 = p(\lambda_0, \theta_0)$ for the true success rate.

B.1.1 Proof of Proposition 2.1

Both agents are learning the value of a one-dimensional parameter p_0 from a sequence of i.i.d. Bernoulli trials. In addition, the full-support assumptions guarantee that both agents' prior beliefs put positive mass on a neighborhood of p_0. Standard statistical learning theorems (see for instance Gelman et al., 2013) prove that with probability one $K_{t,i}(p) \to 0$ for any $p < p_0$ and $K_{t,i}(p) \to 1$ for any $p > p_0$. Thus, with probability one $K_{t,i}$ converges pointwise to the cdf of δ_{p_0} at any $p \neq p_0$, i.e. at any point where the limit cdf is continuous. This proves that with probability one the distribution $K_{t,i}$ converges in distribution to δ_{p_0}.

B.1.2 Proof of Proposition 2.2

Let

$$\Omega(p) = \{\lambda \in \Lambda \mid p(\lambda, \underline{\theta}) \leq p \leq p(\lambda, \bar{\theta})\}.$$

Let us define the function $\theta : \{(p, \lambda) \in (0,1) \times \Lambda \mid \lambda \in \Omega(p)\} \to \Theta$ by

$$p(\lambda, \theta(p, \lambda)) = p.$$

By the implicit function theorem $\theta(., \lambda)$ is continuously differentiable and its partial derivative θ_p is positive.

Let $h_i : (0,1) \times \Lambda \to \mathbb{R}$ be defined by

$$h_i(p, \lambda) = \begin{cases} f_{0,i}(\theta(p, \lambda))\theta_p(p, \lambda) & \text{if } \lambda \in \Omega(p) \\ 0 & \text{otherwise.} \end{cases}$$

The proof relies on the following lemma.

Lemma A.2. *With probability one the sequence $G_{t,i}$ converges weakly to a limit distribution $G_{\infty,i}$ characterized by the density*

$$g_{\infty,i}(\lambda) = \frac{g_0(\lambda) h_i(p_0, \lambda)}{\int_\Lambda g_0(\lambda') h_i(p_0, \lambda') d\lambda'}.$$

Proof lemma A.2. We show that with probability one $g_{t,i}(\lambda) \to g_{\infty,i}(\lambda)$ for all

$\lambda \in \Lambda$ that does not belong to the boundary of $\Omega(p_0)$. Since the boundary of $\Omega(p_0)$ has measure zero, Scheffe's lemma then implies that with probability one $G_{t,i}$ converges weakly to $G_{\infty,i}$.

For all $\lambda \in \Lambda$, a change of variables delivers

$$\int_\Theta f_{0,i}(\theta)p(\lambda,\theta)^{n_t}(1-p(\lambda,\theta))^{t-n_t}d\theta = \int_0^1 h_i(p,\lambda)p^{n_t}(1-p)^{t-n_t}dp.$$

Fix λ and a history (t, n_t). By Bayes' rule,

$$\begin{aligned} g_{t,n_t,i}(\lambda) &= \frac{g_0(\lambda)\int_\Theta f_{0,i}(\theta)p(\lambda,\theta)^{n_t}(1-p(\lambda,\theta))^{t-n_t}d\theta}{\int_\Lambda g_0(\lambda')[\int_\Theta f_{0,i}(\theta)p(\lambda',\theta)^{n_t}(1-p(\lambda',\theta))^{t-n_t}d\theta]d\lambda'} \\ &= \frac{g_0(\lambda)\int_0^1 h_i(p,\lambda)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 [\int_\Lambda g_0(\lambda')h_i(p,\lambda')d\lambda']p^{n_t}(1-p)^{t-n_t}dp}. \end{aligned} \tag{A.6}$$

Lemma A.3. *Let $u : [0,1] \to [0,\infty)$ and $v : [0,1] \to [0,\infty)$ be integrable and bounded functions. Suppose that u and v are continuous on a neighborhood of p_0 and that $v(p_0) > 0$. Then for any sequence n_t such that $n_t/t \to p_0$,*

$$\lim_{t\to+\infty} \frac{\int_0^1 u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp} = \frac{u(p_0)}{v(p_0)}.$$

Proof. Suppose first that $u(p_0) = 0$. Let

$$I_t = \frac{\int_0^1 u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp}.$$

Fix $\epsilon > 0$. By continuity of u and v on a neighborhood of p_0 there exists $\delta > 0$ and a constant $m > 0$ such that $v(p) > m$ and $u(p) < m\epsilon/2$ for any $p \in (p_0 - \delta, p_0 + \delta)$. Let us decompose the integral in three regions $[0, p_0 - \delta], [p_0 - \delta, p_0 + \delta], [p_0 + \delta, 1]$.

First, note that for any $t \in \mathbb{N}$

$$\frac{\int_{p_0-\delta}^{p_0+\delta} u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp} \leq \frac{\int_{p_0-\delta}^{p_0+\delta} u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_{p_0-\delta}^{p_0+\delta} v(p)p^{n_t}(1-p)^{t-n_t}dp} \leq \frac{2\delta m\frac{\epsilon}{2}}{2\delta m} = \frac{\epsilon}{2}. \tag{A.7}$$

Since $n_t/t \to p_0$ there exists $t_0 \in \mathbb{N}$ such that $p_0 - \delta/4 < n_t/t$ for any $t \geq t_0$. If $t \geq t_0$ the function $x \to x^{n_t}(1-x)^{t-n_t}$ is increasing on $(0, p_0 - \delta/4)$. Thus,

$$\frac{\int_0^{p_0-\delta} u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp} \leq \frac{\int_0^{p_0-\delta} u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_{p_0-\delta/2}^{p_0-\delta/4} v(p)p^{n_t}(1-p)^{t-n_t}dp} \leq \frac{(p_0-\delta)\sup u}{\frac{\delta m}{4}} \frac{(p_0-\delta)^{n_t}(1-p_0+\delta)^{t-n_t}}{(p_0-\frac{\delta}{2})^{n_t}(1-p_0+\frac{\delta}{2})^{t-n_t}}.$$

Note that the expression on the right-hand side converges to zero. Therefore there exists $t_1 \geq t_0$ such that for all $t \geq t_1$,

$$\frac{\int_0^{p_0-\delta} u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp} \leq \frac{\epsilon}{4}. \tag{A.8}$$

Similarly, there exists t_2 in $\mathbb{N}$ such that for any $t \geq t_2$,

$$\frac{\int_{p_0+\delta}^1 u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp} \leq \frac{\epsilon}{4}. \tag{A.9}$$

Combining inequalities A.7, A.8 and A.9 shows that for any $t \geq \max(t_1, t_2)$,

$$I_t \leq \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon.$$

This proves that $\lim_{t\to+\infty} I_t = 0$.

Suppose now that $u(p_0) > 0$. Since $|u(p)v(p_0) - v(p)u(p_0)| = 0$ for $p = p_0$ we

have

$$\lim_{t\to+\infty} \frac{\int_0^1 |u(p)v(p_0) - v(p)u(p_0)|p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp} = 0,$$

which implies

$$\lim_{t\to+\infty} \frac{\int_0^1 u(p)p^{n_t}(1-p)^{t-n_t}dp}{\int_0^1 v(p)p^{n_t}(1-p)^{t-n_t}dp} = \frac{u(p_0)}{v(p_0)}.$$

□

To complete the proof of Lemma A.2, note by the law of large numbers that the sequence n_t/t converges almost surely to p_0. Consider any such sequence and any $\lambda \in \Lambda$ that does not belong to the boundary of $\Omega(p_0)$. Let $u(p) = h_i(p, \lambda)$ and

$$v(p) = \int_\Lambda g_0(\lambda')h_i(p, \lambda')d\lambda'$$

The functions u and v are integrable and bounded. If λ belongs to the interior of $\Omega(p_0)$ then $u(p) = f_{0,i}(\theta(p,\lambda))\theta_p(p,\lambda)$ on a neighborhood of p_0 and thus u is continuous on this neighborhood. If λ does not belong to the interior of $\Omega(p_0)$ then $u(p) = 0$ on a neighborhood of p_0, and thus u is continuous on this neighborhood. Furthermore, since $\inf p < p_0 < \sup p$ the function $h_i(p_0, \lambda')$ is positive on a subset of Λ of positive measure. Thus, $v(p_0) > 0$ and v is continuous on a neighborhood of p_0. Therefore u and v satisfy all the assumptions of Lemma A.3, which by A.6 implies

$$\lim_{t\to+\infty} g_{t,n_t,i}(\lambda) = g_{\infty,i}(\lambda).$$

This completes the proof of Lemma A.2. □

To conclude the proof of Proposition 2.2, note that $g_{\infty,1}$ and $g_{\infty,2}$ have the same support $\Omega(p_0)$. Take $\lambda_1, \lambda_2 \in \Omega(p_0)$ such that $\lambda_1 > \lambda_2$. We have

$$\begin{aligned}
&\theta(\lambda_1, p_0) < \theta(\lambda_2, p_0) \text{ since p is increasing} \\
\Rightarrow &\frac{f_{0,1}(\theta(\lambda_1, p_0))}{f_{0,2}(\theta(\lambda_1, p_0))} \leq \frac{f_{0,1}(\theta(\lambda_2, p_0))}{f_{0,2}(\theta(\lambda_2, p_0))} \text{ since } f_{0,1} \succeq f_{0,2} \\
\Rightarrow &\frac{h_1(p_0, \lambda_1)}{h_2(p_0, \lambda_1)} \leq \frac{h_1(p_0, \lambda_2)}{h_2(p_0, \lambda_2)} \\
\Rightarrow &\frac{g_{\infty,1}(\lambda_1)}{g_{\infty,2}(\lambda_1)} \leq \frac{g_{\infty,1}(\lambda_2)}{g_{\infty,2}(\lambda_2)}.
\end{aligned}$$

This proves that $g_{\infty,1} \preceq g_{\infty,2}$.

Proof of Proposition 2.3 We omit the details for the sake of brevity. With arguments similar to the proof of Proposition 2.2 it is possible to prove that $f_{\infty,1}$ and $f_{\infty,2}$ have the same (non-empty) support and that $f_{\infty,1}/f_{\infty,2}$ is proportional to $f_{0,1}/f_{0,2}$ on that support.

B.2 Proof of Proposition 3

We prove the result for the subsequence that consists only of dates that are multiples of m, as extending the result to intermediate dates is straightforward. To simplify the notation we therefore write $F_{t,h_t,i}$ for the beliefs held after trying t environments, i.e. mt periods in total.

For any θ and any $k \in \{0, \cdots, m\}$, let

$$q_k(\theta) = \binom{m}{k} \int_\Lambda \mathcal{L}_{m,k}(\lambda, \theta) dG_0(\lambda)$$

be the probability of succeeding k times out of m trials in a stable an environment randomly drawn according to g_0 and conditional on ability being equal to θ.

For any date t and any $k \in \{0, \cdots, m\}$, let $n_{t,k} \in \{0, \cdots, t\}$ be the number of environments up to date t at which the individual has succeeded k times and failed $m-k$ times, and $h_t = (n_{t,0}, \cdots, n_{t,m})$.

Fix ϵ and take any $\tilde{\theta} < \theta_0$ and any δ such that $\tilde{\theta} < \theta_0 - \delta < \theta_0$. Bayes' rule delivers

$$\frac{F_{t,h_t,i}(\tilde{\theta})}{1 - F_{t,h_t,i}(\tilde{\theta})} = \frac{\int_{\underline{\theta}}^{\tilde{\theta}} \prod_{k=0}^{m} q_k(\theta)^{n_{t,k}} dF_{0,i}(\theta)}{\int_{\tilde{\theta}}^{\bar{\theta}} \prod_{k=0}^{m} q_k(\theta)^{n_{t,k}} dF_{0,i}(\theta)}. \tag{A.10}$$

Take $\theta_1 \leq \tilde{\theta}$.

$$\frac{1}{t} \ln \left[\frac{\prod_{k=0}^{m} q_k(\theta_1)^{n_{t,k}}}{\prod_{k=0}^{m} q_k(\tilde{\theta})^{n_{t,k}}} \right] = \sum_{k=0}^{m} \frac{n_{t,k}}{t} \ln \left[\frac{q_k(\theta_1)}{q_k(\tilde{\theta})} \right]. \tag{A.11}$$

By the law of large numbers, $n_{t,k}/t \to q_k(\theta_0)$ almost surely for any k. For

any such sequence, the right-hand side of A.11 converges to

$$D_{KL}(Q_{\theta_0}||Q_{\tilde{\theta}}) - D_{KL}(Q_{\theta_0}||Q_{\theta_1})$$

where for any θ, $D_{KL}(Q_{\theta_0}||Q_\theta)$ is the Kullback-Leibler divergence from Q_θ to the true distribution Q_{θ_0} defined by

$$D_{KL}(Q_{\theta_0}||Q_\theta) = \sum_{k=0}^{m} q_k(\theta_0) \ln \frac{q_k(\theta_0)}{q_k(\theta)}.$$

Since $\theta_1 \leq \tilde{\theta} < \theta_0$ it is easy to see that

$$D_{KL}(Q_{\theta_0}||Q_{\tilde{\theta}}) \leq D_{KL}(Q_{\theta_0}||Q_{\theta_1})$$

and thus by Equation A.11,

$$\prod_{k=0}^{m} q_k(\theta_1)^{n_{t,k}} \leq \prod_{k=0}^{m} q_k(\tilde{\theta})^{n_{t,k}}$$

when t is large enough.

Similar arguments prove that for any $\theta_2 \in [\theta_0 - \delta, \theta_0]$,

$$\prod_{k=0}^{m} q_k(\theta_2)^{n_{t,k}} \geq \prod_{k=0}^{m} q_k(\theta_0 - \delta)^{n_{t,k}}$$

when t is large enough.

Hence by Equation A.10, there exists t_0 such that for any $t \geq t_0$,

$$\frac{F_{t,h_t,i}(\tilde{\theta})}{1 - F_{t,h_t,i}(\tilde{\theta})} \leq \frac{\tilde{\theta} - \underline{\theta}}{\delta} \frac{\prod_{k=0}^{m} q_k(\tilde{\theta})^{n_{t,k}}}{\prod_{k=0}^{m} q_k(\theta_0 - \delta)^{n_{t,k}}}. \tag{A.12}$$

Note that

$$\frac{1}{t} \ln \left[\frac{\prod_{k=0}^{m} q_k(\tilde{\theta})^{n_{t,k}}}{\prod_{k=0}^{m} q_k(\theta_0 - \delta)^{n_{t,k}}} \right] = \sum_{k=0}^{m} \frac{n_{t,k}}{t} \ln \left[\frac{q_k(\tilde{\theta})}{q_k(\theta_0 - \delta)} \right]. \tag{A.13}$$

The right-hand side of A.13 converges to $D_{KL}(Q_{\theta_0}||Q_{\theta_0 - \delta}) - D_{KL}(Q_{\theta_0}||Q_{\tilde{\theta}})$

which is negative since $\tilde{\theta} < \theta_0 - \delta < \theta_0$. Thus, the left-hand side of A.13 converges to a negative limit, which implies that the argument of the logarithm tends to zero. As a consequence there exists t_1 such that $t \geq t_1$ implies

$$\frac{\prod_{k=0}^{m} q_k(\tilde{\theta})^{n_{t,k}}}{\prod_{k=0}^{m} q_k(\theta_0 - \delta)^{n_{t,k}}} < \frac{\epsilon\delta}{\tilde{\theta} - \underline{\theta}}.$$

Equation A.12 implies that for any $t \geq \max(t_0, t_1)$,

$$\frac{F_{t,h_t,i}(\tilde{\theta})}{1 - F_{t,h_t,i}(\tilde{\theta})} < \epsilon.$$

This proves that $F_{t,h_t,i}(\tilde{\theta}) \to 0$ almost surely for any $\tilde{\theta} < \theta_0$. Similar arguments show that $F_{t,h_t,i}(\tilde{\theta}) \to 1$ almost surely for any $\tilde{\theta} > \theta_0$. Thus with probability one $F_{t,h_t,i}$ converges in distribution to $F_{\infty,i}$ defined by $F_{\infty,i} = \delta_{\theta_0}$.

B.3 Proof of Section 4.2.1

We first explain Equation 1. By Bayes' rule, the agent's subjective probability of success from selecting arm 1 at the next trial equals

$$\frac{(1-q)(1-\nu)p_l B_t^{n_t}(p_l) + [(1-q)\nu + q(1-\nu)]p_m B_t^{n_t}(p_m) + q\nu p_h B_t^{n_t}(p_h)}{(1-q)(1-\nu)B_t^{n_t}(p_l) + [(1-q)\nu + q(1-\nu)]B_t^{n_t}(p_m) + q\nu B_t^{n_t}(p_h)}. \tag{A.14}$$

His subjective probability of success from selecting arm 2 equals

$$\begin{aligned} &\frac{(1-q)[(1-\nu)B_t^{n_t}(p_l) + \nu B_t^{n_t}(p_m)][(1-\nu)p_l + \nu p_m]}{(1-q)(1-\nu)B_t^{n_t}(p_l) + [(1-q)\nu + q(1-\nu)]B_t^{n_t}(p_m) + q\nu B_t^{n_t}(p_h)} \\ &+ \frac{q[(1-\nu)B_t^{n_t}(p_m) + \nu B_t^{n_t}(p_h)][(1-\nu)p_m + \nu p_h]}{(1-q)(1-\nu)B_t^{n_t}(p_l) + [(1-q)\nu + q(1-\nu)]B_t^{n_t}(p_m) + q\nu B_t^{n_t}(p_h)}. \end{aligned} \tag{A.15}$$

The agent strictly prefers selecting arm 2 if and only if expression A.14 is smaller than expression A.15. After some algebra, this condition simplifies to condition 1.

We now prove that if condition 1 is satisfied for some parameter values (q, n_t, t), it is also satisfied in (q', n_t, t) for any $q' > q$ and in (q, n', t) for any $n' < n_t$.

Indeed, given that $p_l < p_m < p_h$ it is easy to check that, for any (n_t, t), $B_t^{n_t}(p_h) > B_t^{n_t}(p_m)$ implies that $B_t^{n_t}(p_m) \geq B_t^{n_t}(p_l)$. Hence, if condition 1 is

satisfied for some $q < 1$ we must have $B_t^{n_t}(p_h) \leq B_t^{n_t}(p_m)$, which implies that the condition is satisfied in $q' = 1$; since Equation 1 is affine in q, it is therefore satisfied for any $q' > q$. In addition, if condition 1 is satisfied after $n_t \geq 1$ successes then

$$
\begin{aligned}
&(1-q)(p_m - p_l)[B_t^{n_t-1}(p_m) - B_t^{n_t-1}(p_l)] + q(p_h - p_m)[B_t^{n_t-1}(p_h) - B_t^{n_t-1}(p_m)] \\
=&\,(1-q)(p_m - p_l)[\frac{1-p_m}{p_m} B_t^{n_t}(p_m) - \frac{1-p_l}{p_l} B_t^{n_t}(p_l)] \\
&+ q(p_h - p_m)[\frac{1-p_h}{p_h} B_t^{n_t}(p_h) - \frac{1-p_m}{p_m} B_t^{n_t}(p_m)] \\
\leq&\frac{1-p_m}{p_m}\Big[(1-q)(p_m - p_l)[B_t^{n_t}(p_m) - B_t^{n_t}(p_l)] + q(p_h - p_m)[B_t^{n_t}(p_h) - B_t^{n_t}(p_m)]\Big] \\
&\text{since } -\frac{1-p_l}{p_l} \leq -\frac{1-p_m}{p_m} \text{ and } \frac{1-p_h}{p_h} \leq \frac{1-p_m}{p_m} \\
\leq&0
\end{aligned}
$$

and thus the condition is satisfied after $n_t - 1$ successes. By induction it is satisfied for any $n' < n_t$.

B.4 Proof of Proposition 4

We start by establishing some general properties of the decision problem and of the value function. We then prove Propositions 4.1 and 4.2 in turn. The proof that the agent stops experimenting in finite time almost surely is included in both parts.

B.4.1 Preliminaries

In all this subsection we assume that the agent's true ability is $\underline{\theta}$.

The agent's beliefs about his own ability and the current environment are summarized by the probability distribution $A = (\alpha, \beta, \gamma, \omega)$ over the two-dimensional variable (λ, θ): $(\alpha, \beta, \gamma, \omega)$ are the weights assigned to the states $(\underline{\lambda}, \underline{\theta})$, $(\overline{\lambda}, \underline{\theta})$, $(\underline{\lambda}, \overline{\theta})$, and $(\overline{\lambda}, \overline{\theta})$. If the agent tries a new environment while his current beliefs assign a weight q to $\overline{\theta}$, then $A = [(1-\nu)(1-q), \nu(1-q), (1-\nu)q, \nu q]$.

Let

$$p(A) = \alpha p_l + (\beta + \gamma) p_m + \omega p_h$$

be the immediate expected reward from the current environment under beliefs A.

Let

$$W(A,\sigma) = \mathbb{E}_A \sum_{t=0}^{+\infty} \delta^t \pi_t(\sigma_t)$$

be the *value* of a strategy σ given initial beliefs A. The value function of the problem is

$$V(A) = \sup_{\sigma} W(A,\sigma).$$

For any $A = (\alpha, \beta, \gamma, \omega)$, let

$$\psi A = \frac{1}{\alpha p_l + (\beta+\gamma) p_m + \omega p_h} [\alpha p_l, \beta p_m, \gamma p_m, \omega p_h]$$

be the updated distribution after a success in the current environment. Let ϕA be defined similarly, as the updated distribution after a failure in the current environment. Lastly, let

$$hA = \Big[(\alpha+\beta)(1-\nu), (\alpha+\beta)\nu, (\gamma+\delta)(1-\nu), (\gamma+\delta)\nu\Big]$$

be the distribution of states corresponding to arm 2.

The space of possible distributions A is endowed with the Euclidean topology on the three-dimensional simplex.

Lemma A.4. *There exists an optimal strategy. The value function V is continuous in A and satisfies*

$$V(A) = \max\Big[p(A) + \delta p(A) V(\psi A) + \delta(1-p(A)) V(\phi A), V(hA)\Big].$$

Proof. The existence of an optimal policy and the Bellman equation follow from standard arguments since the value of any strategy is bounded between 0 and $1/(1-\delta)$.

To prove the continuity of V, fix a distribution A, an optimal strategy σ under A and $\epsilon > 0$. Fix T such that $\delta^{T+1}/(1-\delta) \leq \epsilon/4$.

Since the rewards are Bernoulli, there exists a constant $a > 0$ such that, for any distribution B such that $\|A - B\| < a$, the probabilities of any history up to date T under A and under B differ by at most $\epsilon/T(T+1)$. This implies that

$$|\mathbb{E}_{\sigma,A} \sum_{t=0}^{T} \delta^t \pi_t - \mathbb{E}_{\sigma,B} \sum_{t=0}^{T} \delta^t \pi_t| \leq \sum_{n=0}^{T} n \frac{\epsilon}{T(T+1)} = \frac{\epsilon}{2}.$$

Hence

$$
\begin{aligned}
&|W(A,\sigma) - W(B,\sigma)| \\
\leq &|W(A,\sigma) - \mathbb{E}_{\sigma,A}\sum_{t=0}^{T}\delta^t\pi_t| + |\mathbb{E}_{\sigma,A}\sum_{t=0}^{T}\delta^t\pi_t - \mathbb{E}_{\sigma,B}\sum_{t=0}^{T}\delta^t\pi_t| + |\mathbb{E}_{\sigma,B}\sum_{t=0}^{T}\delta^t\pi_t - W(B,\sigma)| \\
\leq &\frac{\epsilon}{4} + \frac{\epsilon}{2} + \frac{\epsilon}{4} = \epsilon.
\end{aligned}
$$

Hence, in the B-bandit the strategy σ delivers a value at least equal to $V(A) - \epsilon$, which implies $V(B) \geq V(A) - \epsilon$. The symmetric reasoning shows that $V(A) \geq V(B) - \epsilon$, and thus $|V(B) - V(A)| \leq \epsilon$ for any B such that $\|A - B\| < a$. This proves the continuity of V. □

Let us write $V_1(A) = p(A) + \delta p(A)V(\psi A) + \delta(1 - p(A))V(\phi A)$ and $V_2(A) = V(hA)$ for the expected payoffs obtained after pulling arm 1 or arm 2 respectively, and playing optimally thereafter.

For any $q \in [0,1]$, let $A_{0,q} = [(1-q)(1-\nu), (1-q)\nu, q(1-\nu), q\nu]$ be the agent's beliefs if he tries a new environment with a self-confidence q.

Lemma A.5. *$V(A_{0,q})$ is strictly increasing in q.*

Proof. Consider $q < q'$ and let $A_{0,q}$ and $A_{0,q'}$ be the corresponding initial distributions. Consider any date t and a history h_t of outcomes up to date t, possibly in different environments. Let $f_{\underline{\lambda},\underline{\theta}}(h_t)$ be the (ex ante) probability of observing the history h_t conditional on the true type being $\underline{\theta}$ and the current environment at date t being of type $\underline{\lambda}$; let $f_{\underline{\lambda},\overline{\theta}}(h_t)$, $f_{\overline{\lambda},\underline{\theta}}(h_t)$, and $f_{\overline{\lambda},\overline{\theta}}(h_t)$ be defined similarly.

Starting from the distribution $A_{0,q}$ the agent's posterior beliefs $A_{t,q}$ at date t are proportional to

$$\big[(1-q)(1-\nu)f_{\underline{\lambda},\underline{\theta}}(h_t), (1-q)\nu f_{\overline{\lambda},\underline{\theta}}(h_t), q(1-\nu)f_{\underline{\lambda},\overline{\theta}}(h_t), q\nu f_{\overline{\lambda},\overline{\theta}}(h_t)\big].$$

Thus, the agent's subjective distribution over the immediate success probability of arm 1 is strictly increasing in q in the monotone likelihood ratio ordering. Hence, $p(A_{t,q'}) > p(A_{t,q})$ for any $q' > q$.

The agent's posterior beliefs $hA_{t,q}$ over arm 2 are proportional to

$$
\begin{aligned}
&\big[(1-q)(1-\nu)[(1-\nu)f_{\underline{\lambda},\underline{\theta}}(h_t) + \nu f_{\overline{\lambda},\underline{\theta}}(h_t)], (1-q)\nu[(1-\nu)f_{\underline{\lambda},\underline{\theta}}(h_t) + \nu f_{\overline{\lambda},\underline{\theta}}(h_t)], \\
&q(1-\nu)[(1-\nu)f_{\underline{\lambda},\overline{\theta}}(h_t) + \nu f_{\overline{\lambda},\overline{\theta}}(h_t)], q\nu[(1-\nu)f_{\underline{\lambda},\overline{\theta}}(h_t) + \nu f_{\overline{\lambda},\overline{\theta}}(h_t)]\big].
\end{aligned}
$$

Similarly, the agent's subjective distribution over the immediate success probability of arm 2 is strictly increasing in q in the monotone likelihood ratio ordering. Hence, $p(hA_{t,q'}) > p(hA_{t,q})$ for any $q' > q$.

Hence at any date t and for any history h_t, the reward from each arm is strictly greater under distribution $A_{t,q'}$ than under distribution $A_{t,q}$. If σ is an optimal strategy for the q-bandit the value of σ in the q'-bandit is therefore strictly greater than $V(A_{0,q})$. This implies $V(A_{0,q'}) > V(A_{0,q})$. □

Lemma A.6 relies on arguments similar to the proof of Proposition 3 and is provided without proof.

Lemma A.6. *On any path on which the agent experiments an infinite number of environments, $q_t \to 0$ almost surely.*

Lemma A.7. *There exists $\pi > 0$ and $q^* > 0$ such that for any $q \leq q^*$, if the agent tries a new environment with initial self-confidence q then the agent's probability of staying in this environment forever is greater than π.*

Proof. For any $q \in [0,1]$, let $A_{t,n_t,q}$ be the agent's posterior distribution updated from the prior $A_{0,q}$ following n_t successes and $t - n_t$ failures in the same environment.

Claim A.3. *There exists $\kappa_1 > 0$, $\epsilon > 0$ such that*

$$\frac{n_t}{t} \geq p_m - \kappa_1 \Rightarrow V_1(A_{t,n_t,0}) - V_2(A_{t,n_t,0}) > \epsilon.$$

Proof. Take $\kappa_1, \iota > 0$ such that

$$(p_m - \kappa_1) \ln \frac{p_m}{p_l} + (1 - p_m + \kappa_1) \ln \frac{1 - p_m}{1 - p_l} > \iota. \tag{A.16}$$

Such a pair (κ_1, ι) exists by continuity since the left-hand side of A.16 is positive for $\kappa_1 = 0$.

Consider any (n_t, t) such that $n_t/t \geq p_m - \kappa_1$. Then

$$\frac{n_t}{t} \ln \frac{p_m}{p_l} + (1 - \frac{n_t}{t}) \ln \frac{1 - p_m}{1 - p_l} > \iota$$

which implies

$$p_m^{n_t}(1 - p_m)^{t - n_t} > e^{t\iota} p_l^{n_t}(1 - p_l)^{t - n_t} \geq p_l^{n_t}(1 - p_l)^{t - n_t}.$$

Note that

$$p(A_{t,n_t,0}) = \frac{(1-\nu)p_l^{n_t+1}(1-p_l)^{t-n_t} + \nu p_m^{n_t+1}(1-p_m)^{t-n_t}}{(1-\nu)p_l^{n_t}(1-p_l)^{t-n_t} + \nu p_m^{n_t}(1-p_m)^{t-n_t}}$$

and

$$p(hA_{t,n_t,0}) = (1-\nu)p_l + \nu p_m$$

which implies that

$$p(A_{t,n_t,0}) - p(hA_{t,n_t,0}) = \frac{\nu(1-\nu)(p_m-p_l)[p_m^{n_t}(1-p_m)^{t-n_t} - p_l^{n_t}(1-p_l)^{t-n_t}]}{(1-\nu)p_l^{n_t}(1-p_l)^{t-n_t} + \nu p_m^{n_t}(1-p_m)^{t-n_t}}.$$

Hence, for any (n_t, t) such that $n_t/t \geq p_m - \kappa_1$, we have

$$\begin{aligned} p(A_{t,n_t,0}) - p(hA_{t,n_t,0}) &> \frac{\nu(1-\nu)(p_m-p_l)(1-e^{-t\iota})p_m^{n_t}(1-p_m)^{t-n_t}}{p_m^{n_t}(1-p_m)^{t-n_t}} \\ &\geq \epsilon \end{aligned}$$

where $\epsilon = \nu(1-\nu)(p_m-p_l)(1-e^{-\iota})$ is positive since $\iota > 0$.

In addition, arguments similar to those used in the proof of lemma A.5 show that if $p(A_{t,n_t,0}) \geq p(hA_{t,n_t,0})$ then the continuation value after pulling arm 1 is greater than the continuation value after pulling arm 2. Hence, for any (n_t, t) such that $n_t/t \geq p_m - \kappa_1$ we get

$$V_1(A_{t,n_t,0}) - V_2(A_{t,n_t,0}) \geq p(A_{t,n_t,0}) - p(hA_{t,n_t,0}) > \epsilon.$$

This completes the proof of claim A.3. □

Claim A.4. *There exists $\kappa_2 > 0$ and $M > 0$ such that*

$$\forall q \in [0,1], \frac{n_t}{t} \leq p_m + \kappa_2 \Rightarrow \|hA_{t,n_t,q} - hA_{t,n_t,0}\| < Mq.$$

Proof. The posterior distribution $hA_{t,n_t,q}$ is given by

$$hA_{t,n_t,q} = [(1-\nu)(1-q_{t,n_t}), \nu(1-q_{t,n_t}), (1-\nu)q_{t,n_t}, \nu q_{t,n_t}]$$

where q_{t,n_t} is characterized by

$$\frac{q_{t,n_t}}{1-q_{t,n_t}} = \frac{q}{1-q}\frac{\nu p_h^{n_t}(1-p_h)^{t-n_t} + (1-\nu)p_m^{n_t}(1-p_m)^{t-n_t}}{\nu p_m^{n_t}(1-p_m)^{t-n_t} + (1-\nu)p_l^{n_t}(1-p_l)^{t-n_t}}. \tag{A.17}$$

Take $\kappa_2 > 0$ such that

$$(p_m + \kappa_2) \ln \frac{p_h}{p_m} + (1 - p_m - \kappa_2) \ln \frac{1 - p_h}{1 - p_m} < 0. \tag{A.18}$$

Such a number κ_2 exists by continuity since the left-hand side of A.18 is negative for $\kappa_2 = 0$.

Consider any (n_t, t) such that $n_t/t \leq p_m + \kappa_2$. Then

$$p_h^{n_t}(1 - p_h)^{t-n_t} \leq p_m^{n_t}(1 - p_m)^{t-n_t}$$

which together with A.17 implies

$$\begin{aligned} \frac{q_{t,n_t}}{1 - q_{t,n_t}} &\leq \frac{q}{1-q} \frac{p_m^{n_t}(1 - p_m)^{t-n_t}}{\nu p_m^{n_t}(1 - p_m)^{t-n_t} + (1 - \nu) p_l^{n_t}(1 - p_l)^{t-n_t}} \\ &\leq \frac{q}{1-q} \frac{1}{\nu} \end{aligned}$$

which implies $q_{t,n_t} \leq q/\nu$.

Then for any (n_t, t) such that $n_t/t \leq p_m - \kappa_2$,

$$\begin{aligned} \|hA_{t,n_t,q} - hA_{t,n_t,0}\| &= q_{t,n_t} \sqrt{2} \sqrt{\nu^2 + (1 - \nu)^2} \\ &\leq \frac{q}{\nu} \sqrt{2} \sqrt{\nu^2 + (1 - \nu)^2}. \end{aligned}$$

The proof follows from defining $M = \frac{1}{\nu} \sqrt{2} \sqrt{\nu^2 + (1 - \nu)^2}$. □

Claim A.5. *There exists $t_0 \in \mathbb{N}$ such that, conditional on staying in an environment of type $\overline{\lambda}$ forever, the condition*

$$[\forall t, \frac{n_t}{t} \geq p_m - \kappa_1] \text{ and } \left[\forall t \geq t_0, \frac{n_t}{t} \leq p_m + \kappa_2\right]$$

is satisfied with positive probability.

Proof. We write $\overline{A}$ for the complement of an event A. Let Ω be the event $\{\forall t, n_t/t \geq p_m - \kappa_1\}$. Suppose that the environment is of type $\overline{\lambda}$ and consider the martingale $Y_t = \sum_{s=1}^{t} (n_t - p_m)$ and the stopping time $\iota \in \mathbb{N} \cup \{+\infty\}$ defined by $\iota = \inf\{t \in \mathbb{N} \mid Y_t < 0\}$. Suppose that ι is finite with probability one. The optional stopping theorem implies that $\mathbb{E}[Y_\iota] = \mathbb{E}[Y_1] = 0$. But since ι is finite with probability one we also have $\mathbb{E}[Y_\iota] < 0$, which is a contradiction. Hence,

with some positive probability ι is infinite, i.e.

$$\sum_{s=1}^{t} n_t \geq p_m t \geq (p_m - \kappa_1)t \text{ for all } t$$

This implies that the event Ω has positive probability. Let $\upsilon = \mathbb{P}(\Omega) > 0$.

Let E_t be the event $\{n_t > (p_m + \kappa_2)t\}$. By Hoeffding's inequality,

$$\mathbb{P}(E_t) \leq \exp\left(-2\kappa_2^2 t\right)$$

and thus $\sum_t \mathbb{P}(E_t) < +\infty$. The Borel-Cantelli lemma implies that

$$\mathbb{P}\Big(\bigcap_{t=1}^{+\infty} \bigcup_{s \geq t} E_s\Big) = 0$$

and thus

$$\lim_{t \to +\infty} \mathbb{P}\Big(\bigcup_{s \geq t} E_s\Big) = 0.$$

Take t_0 such that

$$\mathbb{P}\Big(\bigcup_{s \geq t_0} E_s\Big) < \upsilon.$$

We have

$$\mathbb{P}\Big(\Omega \cap \bigcap_{s \geq t_0} \overline{E_s}\Big) = \underbrace{\mathbb{P}(\Omega)}_{=\upsilon} + \underbrace{\mathbb{P}\Big(\bigcap_{s \geq t_0} \overline{E_s}\Big)}_{>1-\upsilon} - \underbrace{\mathbb{P}\Big(\Omega \cup \bigcap_{s \geq t_0} \overline{E_s}\Big)}_{\leq 1} > 0.$$

This completes the proof. □

Claim A.6. *There exists $q^* > 0$ such that, for all $q \leq q^*$,*

$$\forall t < t_0, \forall n_t, V_2(A_{t,n_t,q}) < V_2(A_{t,n_t,0}) + \epsilon \tag{A.19}$$

and

$$\forall t \geq t_0, \frac{n_t}{t} \leq p_m + \kappa_2 \Rightarrow V_2(A_{t,n_t,q}) < V_2(A_{t,n_t,0}) + \epsilon. \tag{A.20}$$

Proof. By claim A.4, the distance between $hA_{t,n_t,q}$ and $hA_{t,n_t,0}$ is uniformly bounded by an expression of the form Mq when $t \geq t_0$ and $n_t/t \leq p_m + \kappa_2$. Since t_0 is fixed, it is easy to see that the distance between $hA_{t,n_t,q}$ and $hA_{t,n_t,0}$ can also be uniformly bounded by an expression of the form $M'q$ for all (n_t, t) such that $t < t_0$. Since V is continuous, it is thus possible to select q^* small

enough to make sure that $q \leq q^*$ implies that $V(hA_{t,n_t,0}) - V(hA_{t,n_t,q}) < \epsilon$ for all (n_t, t) that satisfy one of the two above conditions. Conditions A.19 and A.20 follow from $V_2 = V \circ h$. □

To complete the proof of lemma A.7, take $\epsilon, \kappa_1, \kappa_2, t_0$ as defined in claims A.3—A.6, and let $\pi' > 0$ be the probability with which the condition of claim A.5 is satisfied in an environment of type $\overline{\lambda}$. By claims A.3 and A.6, for all t we have $V_1(A_{t,n_t,0}) > V_2(A_{t,n_t,0}) + \epsilon$ and $V_2(A_{t,n_t,q}) < V_2(A_{t,n_t,0}) + \epsilon$. In addition, arguments similar to those used in the proof of lemma A.5 show that $V_1(A_{t,n_t,q}) \geq V_1(A_{t,n_t,0})$ for all $q \geq 0$. Hence, for any t we have

$$\begin{aligned} V_1(A_{t,n_t,q}) - V_2(A_{t,n_t,q}) &\geq V_1(A_{t,n_t,0}) - V_2(A_{t,n_t,q}) \\ &\geq \underbrace{V_1(A_{t,n_t,0}) - V_2(A_{t,n_t,0})}_{>\epsilon} - \underbrace{[V_2(A_{t,n_t,q}) - V_2(A_{t,n_t,0})]}_{<\epsilon} \\ &> 0. \end{aligned}$$

Hence, at any date t the agent finds it optimal to stay in the current environment. Thus if the environment is of type $\overline{\lambda}$ the probability with which the agent stays in this environment forever is at least π'. Note that π' is independent of q. Defining $\pi = \nu\pi'$ completes the proof. □

B.4.2 Proof of Proposition 4.1

First step We first prove that the agent stops experimenting in finite time almost surely. Let us proceed by contradiction and assume that the agent experiments forever. Suppose first that q_t converges to 0. There exists t_0 such that $q_t \leq q^*$ for any $t \geq t_0$. Thus, by lemma A.7 for any new environment tried at a date $t \geq t_0$ the probability of staying in this environment forever is at least $\pi > 0$. This implies that the agent stops experimenting in finite time with probability one. The other case in which q_t does not converge to zero also happens with probability zero due to lemma A.6. This shows that the agent stops experimenting in finite time almost surely.

In addition, suppose that the last environment is of type $\underline{\lambda}$. If A_t converges to $(1,0,0,0)$, by continuity of the value function the value of arm 1 converges to $V[(1,0,0,0)]$ whereas the value of arm 2 converges to $V[(1-\nu),\nu,0,0]$ which is strictly greater. Thus for t sufficiently large it must be optimal to leave the environment, which is a contradiction. Hence, if the last environment is of type $\underline{\lambda}$ the sequence A_t does not converge to $(1,0,0,0)$, which is a zero-probability

event. This proves that with probability one the agent stops experimenting in finite time in an environment of type $\overline{\lambda}$.

Second step Now, let us prove that q is bounded above. Let q be the agent's initial self-confidence when he tried his last environment for the first time. With probability one q_t converges to

$$q_\infty = \frac{q(1-\nu)}{q(1-\nu)+(1-q)\nu}.$$

To prove that q_∞ is bounded above, note that $V(A_{0,q})$ is a strictly increasing and continuous function of q that satisfies $V(A_{0,0}) < p_m/(1-\delta) < V(A_{0,1})$. Thus there exists $\bar{q} \in (0,1)$ such that

$$V(A_{0,\bar{q}}) = \frac{p_m}{1-\delta}.$$

Suppose that $q_\infty > \bar{q}$. A_t converges to $(0, 1-q_\infty, q_\infty, 0)$ almost surely, in which case the value of arm 1 converges to $p_m/(1-\delta)$ whereas the value of arm 2 converges to $V(A_{0,q_\infty}) > p_m/(1-\delta)$; thus, the agent must find it optimal to leave in finite time, which is a contradiction. Hence $q_\infty \leq \bar{q}$ almost surely.

B.4.3 Proof of Proposition 4.2

Lemma A.7 implies that there exists $q^* > 0$ such that for any $q \leq q^*$, if the agent tries an environment of type $\underline{\lambda}$ with initial self-confidence q then the agent stays forever in this environment with positive probability.[16]

First step That the agent stops experimenting in finite time almost surely and that q_t converges relies on arguments similar to the proof of Proposition 4.1.

Second step We first show that with probability one either of cases 4.2a or 4.2b is realized. If the last environment is of type $\overline{\lambda}$, then with probability one the asymptotic success rate converges to p_h and thus q_t converges to 1. If the last environment is of type $\underline{\lambda}$, a reasoning similar to the proof of Proposition 1 shows that with probability one q_∞ must be bounded above by some constant $\underline{q}$, otherwise the agent would leave the environment in finite time.

[16] Indeed, lemma A.7 only requires that with positive probability the agent selects an environment in which the expected success rate is p_m. This is the case here if $\theta = \overline{\theta}$ for an environment of type $\underline{\lambda}$.

Our final step is to show that both cases 4.2a and 4.2b happen with positive probability. First, we argue that for any initial self-confidence q there exists an integer $t_0 \geq 1$ such that, if the first t_0 attempts made by the agent in a new environment are unsuccessful, the agent's optimal action is to switch to a new environment. Thus, there exists an integer t such that, if the agent has failed at every period up to date t (and switched optimally on that path), then $q_t \leq q^*$. Such a number t exists since $q_t \to 0$ when $t \to \infty$ if the agent fails at each period. In addition, t can be chosen to make sure that the agent switches to a new environment at date t. Failing t consecutive times is a positive-probability event, and the agent stays forever in an environment of type $\underline{\lambda}$ with positive probability thereafter. This proves that case 4.2a happens with positive probability.

Arguments similar to those used in the proof of lemma A.7 show that we can construct a path on which the agent succeeds at every period until q exceeds some threshold $\tilde{q}$, after which the agent has a positive probability of staying in his current environment if this environment is of type $\overline{\lambda}$. Case 4.2b thus also happens with positive probability.