The Evolution of Cognitive Biases in Human Learning

Peter S. Park*

*Department of Mathematics, Harvard University

1 Oxford St., Cambridge, MA 02138 USA

Email: pspark@math.harvard.edu

Abstract. We propose that humans may not meaningfully retain high-variance environmental feedback, a behavior that likely evolved in a past environment with unfavorable fitness tradeoffs from overcommitting attention. We argue that in such settings, humans instead rely on an innate estimate of future payoffs optimized for their evolutionary past of cultural learning: learning from knowledge taught by fellow group members, rather than from the environmental feedback itself. A veridical decision-theoretic model of this evolutionary past can help explain several puzzles of human behavior: (a) the hard-easy effect, the overconfidence (respectively, underconfidence) of one's estimate of her ability in a difficult task (respectively, easy task); (b) underinference, the persistence of the aforementioned flawed belief in the face of evidence to the contrary; (c) the non-monotonicity of confidence with respect to the level of experience; (d) persistent vulnerability to charlatans; and (e) the situational effectiveness of marginal educational interventions. This finding corroborates the thesis[1,2] that dual inheritance theory is an ideal candidate to be a cumulative theoretical framework for the psychological sciences. Our evolutionary theory and its predictions also have foundational policy implications, particularly for education.

1. Introduction

We humans often hold on to biased views of the world. Moreover, our updating of these views in the face of contradicting environmental feedback is generally too small to be unbiased[3]. This is evolutionarily puzzling, given that both research and human intuition consistently indicate that environmental feedback is a foundational source of data for human learning. For example, people often learn to pay credit card bills[4] and return rented videos[5] on time after first paying late fees.

There is a compelling case for considering Bayesian inference—the algorithm that incorporates environmental feedback into one's previously held belief in an unbiased manner—as the normatively rational way to learn from environmental feedback[6]. However, human cognition deviates from Bayesian rationality in numerous ways. Such deviations are called cognitive biases. Their study began half a century ago with the seminal work of Kahneman, Tversky, and their collaborators, who designed replicable experiments that yielded variegated falsifications of the then-largely-unchallenged (and still-widely-used) assumption that human decision-making is rational[7–9].

Since then, cognitive biases have been the subject of an interdisciplinary body of research spanning the cognitive sciences, psychology, and behavioral economics that has helped resolve many long-standing puzzles of human behavior. Why do supposedly rational humans continue to wage unwinnable wars, fuel dangerously inflated financial bubbles, and underprepare in their fight against imminent pandemics?

The answer: humans are not rational. Leaders of countries regularly go to war under the overconfident impression that victory is possible or even assured, notwithstanding the available evidence to the contrary[10,11]. Traders regularly become overconfident in their high valuations of their assets—even in the face of contradicting evidence—resulting in speculative bubbles that eventually crash[12,13]. Emerging pathogens like SARS-CoV-2 regularly spread to become epidemics, due to overconfident politicians and government officials' failure to learn—despite repeated lessons from past experience[14]—how to prevent, detect, contain, and mitigate infectious disease outbreaks[15]. The hubristic combination of two cognitive biases—overconfidence, the belief that one's ability is better than it actually is; and underinference, the tendency to systematically ignore evidence that contradicts this belief—continues to cause severe, long-lasting, and largely avoidable harm to numerous people.

How did our evolutionary past select for cognitive biases, traits that systematically cause errors in judgement? At first glance, learning from environmental feedback in a biased manner seems to be evolutionarily deleterious, since an accurate understanding of one's current environment would facilitate her adaptation to it and thereby bolster her evolutionary fitness. It stands to reason that if ancestral humans learned primarily from environmental feedback, then natural selection would favor those who did so objectively—via Bayesian inference—rather than irrationally.

However, an extensive, multidisciplinary body of evidence suggests that humans primarily rely on learning from their fellow group members, rather than from the environmental feedback itself[16–18]. This is why *homo sapiens* has been so adept at thriving in diverse environments throughout the world. The key to humans' evolutionary success was not their capacity to individually learn about their environment, which is nowhere near game-changing; but their unique capacity—enabled by language and other modes of cultural learning—to draw on the collective reservoir of knowledge on it, compiled by their predecessors and peers[19–21].

Given the secondary role that environmental feedback played in the learning process of our ancestors, cognitive biases did not constitute a death sentence in the eyes of natural selection, as one might originally assume. In fact, we will demonstrate that the optimal way to learn from fellow group members can appear cognitively biased to those under the mistaken assumption that the learning is being driven by environmental feedback. Specifically, we construct a veridical evolutionary model of human learning: one based primarily on learning from group members rather than from environmental feedback. The model is decision-theoretic, in that it parametrizes ancestral humans' environment of cultural learning and investigates the decision-making strategy that maximizes payoff acquisition from this environment (fitness) among the space of feasible strategies (fitness landscape). We proceed to demonstrate that the fitness-maximizing strategy is characterized by various cognitive biases like overconfidence and underinference.

We thus find that multiple classes of cognitive biases that are traditionally thought of as structural flaws in humans' individual learning may instead be simultaneously rooted in the primarily social nature of human learning in natural settings. This finding corroborates the thesis[1,2] that dual inheritance theory is an ideal candidate to be a cumulative theoretical framework for the psychological sciences: one that can help resolve the field's replication crisis by generating meaningful hypotheses not based on personal or folk intuitions.

2. Empirical background

A myriad of cognitive biases, including overconfidence and underinference, are apparent from the insightful experiments of Sanchez and Dunning[22,23] on human learning. In each of their experiments, subjects learned a new task possessing a payoff structure with fixed uncertainty: classifying profiles with lists of properties (for example, symptoms) into categories (for example, made-up diseases). The subjects attempted this task 60 times while simultaneously reporting their confidence: their self-estimate of the probability that their answer is correct. After each of their 60 answers, they received immediate feedback.

Four variants of this experiment were carried out originally[22], and two of these four variants were replicated while also recording time measurements[23]. The subjects' confidence graph—that of their average self-estimate as a function of trial number—behaved similarly across all six experimental variants (see Fig. 1 for a qualitative illustration). Most notably, this shared behavior was non-monotonic, in that it was comprised of three phases: a beginning phase of increase, an intermediate phase of decrease, and a final phase that returned to increase. Meanwhile, the proportion of subjects that actually answered correctly grew monotonically, as expected. Confidence was on average non-monotonic, not only as a function of the subjects' level of experience, but also as a function of the true value: the true probability of giving the correct answer.

The non-monotonicity of confidence is not surprising, given that it has been documented in previous studies. The experiments of Dunning and Kruger[24] on confidence as a function of true ability—as well as replications[25,26]—have often yielded a non-monotonic relationship between the two measured variables. This literature has injected into the public consciousness a famous non-monotonic graph (Fig. 2) attributed to the namesake Dunning–Kruger effect: the cognitive bias in which incompetent people are unable to recognize their incompetence.

Also, the work of Hoffman and Burks[27] investigating truckers' self-estimates of the number of miles driven each week (Fig. 3) has found their average to be non-monotonic with respect to the level of experience, with an initial phase of increase followed by a phase of decrease. Note that the experience level at which the final phase reverting to increase could occur is not included in the empirical data, so this final phase is neither confirmed nor contradicted. Meanwhile, the average of the true value is monotonically increasing in the level of experience: consistent with the findings of Sanchez and Dunning.

There are other puzzling patterns in the Sanchez–Dunning experimental data. For one thing, the average difference between confidence and the environmental feedback—a value that quantifies overconfidence—eventually became positive, and proceeded to increase instead of decaying to zero. Consistently becoming overconfident compared to the environmental feedback, by itself, constitutes an immediate contradiction of Bayesian rationality[28]. For another, the confidence graphs from all six variants of the experiment were essentially indistinguishable from each other, even though the subjects of each experimental variant on average performed differently and thus received different environmental feedback. The confidence graph in essence only depended on the number of past observations, the level of experience.

Overall, the subjects of the Sanchez–Dunning experiments did not learn from their environmental feedback in a Bayesian-rational manner. First, they consistently developed overconfident beliefs that persisted, even in the face of contradicting environmental feedback. Second, confidence somehow always behaved in the same qualitative way regardless of differences in the environmental feedback. Finally, confidence was non-monotonic with respect to the level of experience, even when the true value was monotonic. These three patterns robustly replicated in all six variants of the experiment, including the variant that used the incentive-compatible Becker–DeGroot–Marschak method[29] to monetarily incentivize accurate answers. The non-Bayesian inaccuracy of subjects' learning and the persistence of this inaccuracy in the face of monetary incentivization have also been documented in replications of the Dunning–Kruger experiment[30,31], as well as in the aforementioned work of Hoffman and Burks[27] on truckers' self-estimates of productivity.

Note that the Dunning–Kruger experiment is similar in objective and design to the Sanchez–Dunning experiment. A crucial difference, however, is that accurate environmental feedback is immediately provided by the experimenter in the latter, but not in the former. The Sanchez–Dunning experiment thus compellingly raises the question of why humans have evolved to underinfer from freely available environmental feedback, even when meaningfully learning from it is made easy and monetarily advantageous.

3. Theory

3.1. Learning environment of ancestral humans

Given their role in various and potentially catastrophic errors in human judgement, the three non-Bayesian patterns in the Sanchez–Dunning experimental data are generally considered suboptimal cognitive biases today. However, we propose that they may be products of a payoff-estimation strategy optimized for a past environment. In it, ancestral humans faced unfavorable fitness tradeoffs from overcommitting attention to retain observations of high-variance payoffs, and consequently evolved to neglect them in favor of knowledge learned from their fellow group members. Indeed, the expected benefit of retaining environmental feedback would have been negligible if it came with high statistical noise. The expected cost due to external risks from overcommitting attention—such as increased vulnerability to ambushes by other humans, predation, and accidental injury—would then have been large in comparison. After all, ancestral humans are estimated to have experienced a 14% rate of mortality due to inter-group conflict[32]. This can help explain why a person's estimate of future payoffs may not meaningfully update with high-variance observations of previous payoffs.

This phenomenon is particularly apparent in gamblers. If a gambler who repeatedly tries her luck at a lottery meaningfully retained her environmental feedback, then she would eventually be able to conclude—from the negative mean of her payoffs—that the lottery is most likely rigged against her. In reality, however, human judgement is often clouded by the high variance in the outcome of such lotteries. As a result, many gamblers continue betting their money on the same rigged lotteries they have played numerous times in the past.

When environmental feedback occurs with high variance, the updating process of a human learner's belief may not be due to a meaningful incorporation of her environmental feedback; it may instead be due to a different payoff structure from the evolutionary past. This can help explain why subjects of the Sanchez–Dunning experiments who were monetarily incentivized (by the Becker–DeGroot–Marschak method) to accurately estimate their future payoff—an expected value between a correct answer's full payoff and an incorrect answer's zero payoff—systematically underinferred from their previous environmental feedback. Indeed, not having meaningfully retained it, the subjects instead defaulted to using an innate estimate of the expected payoff optimized for their evolutionary past. This innate estimate can be greater than the estimate based on the environmental feedback, if the underlying task is difficult; or less than it, if the task is easy.

The existence of this hard-easy effect—overconfidence on difficult tasks and underconfidence on easy tasks—is well-established in the empirical literature[33,34]. It is not surprising that if a human learner's process of updating her belief does not meaningfully incorporate environmental feedback, then her belief would diverge from the environmental feedback, and the difference between the two would persist rather than decay to zero.

Our hypothesis—that ancestral humans faced evolutionary pressures to rely on knowledge shared by group members, rather than on high-variance environmental feedback—

can also help explain why humans have evolved to estimate their future payoff as a non-monotonic function of their level of experience. Specifically, this non-monotonicity can be explained by the uncertainty ancestral humans faced regarding whether a given task was learnable. Constructing an evolutionary model of human learning that faithfully represents its collaborative nature—learning not only by innovation, but also by imitation—makes this uncertainty apparent.

3.2. The model

In our evolutionary model (detailed in the Supplementary Information), a student learns how to perform a given task from a fellow group member who claims to be an expert on it. We broadly label such group members "teachers." For the sake of tractability, our model does not distinguish between learning from observing others and learning from direct teaching, unlike past ethnographic studies on hunter-gatherer learning[18,35,36].

As the student in our evolutionary model spends time learning the task from her teacher, she gains more and more knowledge on it. However, it is unclear whether the knowledge is truly coming from imitating her teacher's guidance, as would be the case if the teacher is an actual expert. There is also the possibility that the teacher is a charlatan whose advice does not meaningfully help. In that case, the student learns from epiphanies obtained through the slow process of innovation, even though she unknowingly credits them to the teacher's guidance.

The student hopes to eventually master the task. However, depending on whether her teacher is an actual expert or a charlatan, she may not be able to do so. Learning through innovation under the wings of a charlatanical teacher is not guaranteed to complete in finite time, because—contrary to the charlatan's promises—the task may be impossible. On the other hand, this risk does not exist when the student learns from imitating an actual expert who has already mastered the task. This is analogous to the comparison between solving an exam problem and solving a research problem. The former—imitation learning—is guaranteed to complete in finite time, because the teacher has solved the problem herself before assigning it as an exam question. However, the latter—innovation learning—is not guaranteed to complete in finite time. Indeed, a research problem, by definition, is one that has not yet been solved by anyone, so it may *a priori* be impossible to solve.

The student's expected marginal payoff is a function of her level of experience. This function is monotonically increasing when the current task is guaranteed to be learnable in finite time, but eventually decays to zero when the task may instead be impossibly difficult. We hypothesize that the fitness advantage conferred by accurately measuring future payoffs selected for ancestral humans whose confidence best tracked the evolutionary past's non-monotonic expected payoff function. Its non-monotonicity is due to its piecewise definition (Fig. 4). The increasing, then decreasing portion of the expected payoff function is conditional on the fact that the current task's teacher is either an actual expert or a charlatan. The final increasing portion is conditional on having ruled out the latter possibility, because these tasks should optimally be quit at an intermediate level of experience.

It would be better, if possible, to quit these tasks soon after the student starts to learn them, but she cannot initially differentiate between innovation and imitation learning, because their speeds are too similar. Fortunately, the speed of learning increases over time[23]. We hypothesize that it increases faster for imitation learning than it does for innovation learning, which enables the learner to eventually differentiate between the two types of tasks by a time-measurement experiment. Thus, the student performs a time-measurement experiment at the aforementioned intermediate level of experience, quitting if the speed of her learning is slow—innovation—and not doing so if it is fast—imitation.

It is likely that our hunter-gatherer ancestors relied primarily on learning cooperatively from fellow group members, rather than on learning individually from environmental feedback[16–21,37]. This was especially true when the environmental feedback occurred with high variance, which made it difficult to retain without experiencing unfavorable evolutionary tradeoffs. As a result, the ancestral humans who best survived were not those who learned the most objectively from their environmental feedback; they were those who learned the most effectively from their group members. The optimal strategy for this purpose is a speculative one that cannot initially distinguish between expert and charlatanical teachers, but hopes to learn from the former while mitigating the risk of the latter by quitting at an intermediate level of experience. This strategy appears to suffer from various cognitive biases when one assumes, as in the premise of the Sanchez–Dunning experiments, that it should have been optimized for the purpose of learning from environmental feedback.

4. Predictions

Our evolutionary theory makes several predictions on how humans learn in settings of high-variance environmental feedback. Note that these predictions do not apply to humans' excellent learning in settings of low-variance environmental feedback, enabled by their capacity for causal reasoning[38]. Nor do they apply to humans who have the training and the will to learn from high-variance environmental feedback in a statistically unbiased manner, rather than via their default intuition.

Empirical studies of hunter-gatherer learning[18,35,36,39–42] have yielded valuable insights on the mechanics and evolution of human learning. We hypothesize that when analogous empirical studies investigate the variables relevant to our evolutionary theory—subjects' confidence estimates, teacher selection, and quitting behavior over the course of learning a given task—their findings will corroborate its divergent sets of predictions, depending on the variance of the environmental feedback and the subjects' level of statistical sophistication. Human learning will approximate Bayesian rationality in the absence of causal opacity: when either the given task's environmental feedback (e.g., foraging yield per unit of time) or the environmental feedback regarding evidently important mechanisms of the task (e.g., hunters' accuracy and pull strength with the bow and arrow, or honey-gatherers' tree-climbing ability) has low variance. However, human learning will systematically and persistently deviate from Bayesian rationality when the observed data relevant to the task are generally characterized by high variance (e.g., shamanistic predictions of complex phenomena) and the learner lacks either the training or the will to aggregate this data in a statistically unbiased way.

Quantitative predictions on the causal interactions and statistical relationships between the parameters of our evolutionary model—including the student's level of experience (denoted in the Supplementary Information by the level of knowledge b), teacher/task turnover (denoted by the quitting strategy $b^*_{n,i}$), confidence (denoted by $g_{b^*_{n,i}}(b)$), task learning as a function of time (denoted by $L_1(t)$ and $L_2(t)$), the ancestral environment's distribution of task difficulties (denoted by μ), this environment's expected marginal payoff conditional on the task's difficulty (denoted by $f_a(b)$), the proportion in this environment of tasks that are impossible to meaningfully learn (denoted by p), and the proportion in this environment of learning from imitating an actual expert (denoted by q) —can be found in the Supplementary Information.

Below, we discuss several of the most counterintuitive predictions, which enjoy various degrees of empirical corroboration.

4.1. Underinference from high-variance environmental feedback

We propose that when ancestral humans learned in settings of high-variance environmental feedback, unfavorable evolutionary tradeoffs from overcommitting attention forbade them from meaningfully retaining it. We thus argue that humans have evolved to largely disregard the statistical information contained in sequences of high-variance environmental feedback observed over time. While humans can incorporate into their beliefs easy-to-remember descriptors of these sequences—such as the maximum value, the minimum value, and the number of observations—they may not by default be able to incorporate the statistical information needed to approximate a Bayesian aggregate of this data: most notably, the mean.

Consequently, we hypothesize that human confidence—and revealed preference in general—is best modelled not as a Bayesian estimate, but as a function of the subset of information available to humans' cognitively constrained decision-making. This subset includes low-variance observations, easily retained descriptors of high-variance observations, and even information in the complement of environmental observations: for example, theoretical knowledge and social cues.

The latter dependence should cause the innate confidence function—like other components of human cognition and behavior[43]—to vary cross-culturally. Persistent non-Bayesianness[22,23,27,28,30] and the hard-easy effect[33,34] have been robustly documented in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies. However, WEIRD individuals in many ways constitute outliers on the spectrum of human psychology[43]. It is thus not *a priori* clear whether a psychological pattern observed in WEIRD individuals will generalize[43,44].

We nonetheless hypothesize that the broad patterns predicted by our evolutionary theory will generalize to all statistically unsophisticated humans, including those of various hunter-gatherer societies. A human learner's aggregate of the information cognitively available to her—for example, her innate confidence function—will generically differ from the Bayesian aggregate of all observed information when it has high variance. The difference between the two will generically persist, and moreover, will follow the pattern prescribed by the hard-easy effect.

Despite these predicted universalities, we stress that they are nevertheless best studied in a contextual manner rather than a monolithic one. This necessitates investigating the dependence of such psychological aspects on not just subjects' environmental observations, but on a much larger class of treatment effects that exhibit significant cross-contextual variation[44]. We thus propose replicating the Sanchez–Dunning experiment in a wide variety of contexts, such as with subjects from contemporary hunter-gatherer groups and other non-WEIRD societies. Doing so would yield valuable insights on both the universalities and idiosyncrasies in human learning.

4.2. Non-monotonicity of confidence with respect to the level of experience

Our evolutionary theory proposes that the empirically observed non-monotonicity of human confidence with respect to experience[22,23,27] evolved to track a piecewise-defined payoff structure from our evolutionary past of cooperative learning. Specifically, this piecewise definition ultimately arose from the risk of wasting time learning from a charlatanical teacher.

A deleterious consequence of this non-monotonicity is that inexperienced teachers, leaders, and pundits may express more confidence in their explanatory frameworks and predictions than their more experienced counterparts, thereby worsening their followers' decision-making.

4.3. Persistent vulnerability to charlatans and maladaptive social norms

A temporary vulnerability to charlatanical teachers, leaders, and pundits may be unavoidable and understandable, as illustrated by the proverb "Fool me once, shame on you; fool me twice, shame on me." However, humans' vulnerability to charlatans is often surprisingly persistent[45–47]. One example of this phenomenon is the unfortunate credence that many people have given to pseudoscientific treatments, tests, and vaccines, even with their lives at stake[48]. Another is humans' widespread and persistent vulnerability to cults, even those engaged in coercive or maladaptive practices[49].

Many instances of charlatanical teaching occur due to social norms. Their persistence is due to the fact that social mechanisms such as punishment and status can entrench a wide variety of social norms, even unproductive or maladaptive ones[50,51]. Examples of the latter include shamanism, which consistently develops in hunter-gatherer societies despite the general invalidity of its predictions[52]; funerary cannibalism, which causes prion disease[53]; and female genital cutting, which increases the rates of birthing complications[54] and of contracting sexually transmitted illnesses[55].

Dual inheritance theory proposes that maladaptive social norms can survive group-selection pressures for extended periods of time if they are part of the same cultural package as adaptive social norms, such that the group cannot separate the former from the latter without sacrificing its overall capacity for coordination[56–58]. However, this leaves the question of why humans, for all their abilities in social reasoning and cooperation[59,60], often cannot surgically coordinate away from maladaptive parts of their cultural package during equilibrium selection.

Why do people not abandon the guidance of an overconfident, exploitative, or conformist charlatan when it continues to yield unfavorable results? A possible explanation is provided by one of our learning theory's implications: that quitting due to a lack of meaningful teaching may be triggered not by inauspicious environmental feedback, but by reaching a specific level of experience without having completed learning the given task. In the former case, a charlatan whose advice yields unfavorable results is vulnerable to coordinated discrediting from the students who observe them and thereby decide to quit. This occurs when the outcome is always unfavorable; humans are good at incorporating consistent environmental feedback into their beliefs. In the latter case—which occurs when the outcome is uncertain, in that it can be favorable or unfavorable—the points of time at which her students quit may differ, since they have different levels of experience in general. Consequently, charlatans may be much less vulnerable to coordinated discrediting from their students when environmental feedback occurs with high variance.

Our theory of learning may thus help explain charlatans' influence on human society that has continued long after its adoption of the scientific method: the realization that collecting environmental data should not rely on human cognition's biased process of aggregation. What results from humans' continued vulnerability to charlatans is the lasting evolutionary pressure posed by the risk of being forever led astray by their unhelpful advice. This evolutionary pressure creates a positive feedback loop that both shapes and is shaped by humans' speculative evolutionary strategy: that of learning from teachers without verifying from the get-go whether they are true experts or charlatans.

4.4. Situational effectiveness of marginal educational interventions

When a student learns from a teacher, but does not retain the resulting environmental feedback, our theory predicts the point in time at which she will quit due to a lack of meaningful teaching. Specifically, it predicts that this quitting is triggered when the student reaches a certain level of experience without having completed learning the given task.

We have previously shown that this implication can help explain humans' persistent vulnerability to charlatans in settings of high-variance environmental feedback. It can also help explain researchers' disagreement on the efficacy of educational interventions, such as additional schooling or decreased class size. Such educational interventions marginally lengthen the time students spend meaningfully learning from a teacher. There is a large body of empirical evidence suggesting that compared to their marginal nature, these educational interventions are unusually effective[61–69]. However, there is also a large body of evidence suggesting the opposite: that marginal educational interventions are essentially ineffective[70–74].

The puzzling bimodality of these findings can be resolved by our theory's prediction that quitting due to a lack of meaningful teaching occurs at a singular level of experience. Indeed, this implies that the efficacy of marginal educational interventions is situational: it can either be similarly marginal or unusually high. The outcome depends on whether the students' level of experience is located just below the distinguished quitting point, since passing it while learning from a teacher ensures that they will not give up in the future.

5. Discussion

Many institutions of modern society—such as markets, federalist political systems, and the Internet—predominantly rely on decentralized decision-making. This is justified by the influential argument that decentralized decisions benefit from locally made, locally relevant observations that are generally inaccessible to a centralized decision-maker[75]. However, our theory proposes that humans may not meaningfully retain high-variance environmental observations, a consequence of which is that these observations are then inaccessible not only to the centralized decision-maker, but also—for all intents and purposes—to the local observer herself. This is illustrated by many people's persistent skepticism and denial of climate change, an unrelenting barrier to the implementation of urgently needed climate policies[76]. In contrast, a

Bayesian learner would have eventually deduced from her temperature observations—despite their high variance—that the climate is warming, at least in her local area.

Overall, there is no guarantee that a person's private information, even when truthfully revealed by an incentive-compatible mechanism, constitutes a Bayesian aggregate of her local observations. Policymakers and researchers may hope to at least infer the latter from the former: the sufficient statistic approach. But there is no guarantee that this is possible, either. To illustrate, suppose that we are tasked with predicting a person's ability in a given task from her own estimate of this ability—her confidence—in a setting of high-variance environmental feedback. Unfortunately, we cannot achieve this in general because the statistic is not sufficient. Due to the non-monotonicity of human confidence, an inexperienced person's confidence may be at the same level as that of an expert; consequently, we would not be able to distinguish between the two just from observing their confidence estimates.

Given this statistical insufficiency, tapping into the informational benefit provided by all local observations—rather than just the low-variance ones—would in general require a structural improvement in people's ability to meaningfully learn from these observations. Such a structural improvement can be achieved by education. Students should be taught to avoid placing unconditional reliance on their intuition: their brains' default process of aggregating their observations. Instead, we should take advantage of the various economical ways to objectively record them: technologies—such as writing—that did not exist in our hunter-gatherer past but are widely available today. Even without immediate access to writing, however, we can still rely solely on our cognition, provided that it has learned how to objectively aggregate high-variance observations. Improvements in society's mathematical knowledge have equipped us with a simple and effective way of doing so: keeping track of the mean of the observations, a skill that should be overtly taught in schools rather than taken for granted.

Educators should provide each of their students with the statistical training necessary to retain high-variance observations. They themselves should also avoid relying on their intuition for aggregating the data of their students' progress, and of whether it is due to their actual guidance or the students' individual epiphanies. We further argue that educators should consider relying on statistical analyses of their students' progress over the students' own evaluations, since the latter may be based on flawed intuition. While controversial, our proposition is informed by many students' considerable inaccuracy in assessing the help provided by their teachers, as found by various empirical studies of teacher quality[77,78].

Theoretical arguments suggest that modernizing education in these ways would enhance individual quality of life and societal productivity. At the micro level, it would improve people's decision-making, and at the macro level, it would increase the societal rate of innovation by broadening and diversifying the collective knowledge base of local observations[79–82]. We thus hypothesize that the benefits of educationally enhancing people's ability to learn from their high-variance local observations, while difficult to notice in the short term, will more than make up for its costs in the long term.

Acknowledgements. The author thanks Eric Maskin for suggesting the problem and for giving invaluable guidance throughout the course of this work. He also thanks Chris Bakerlee, Holly Elmore, Ben Golub, Joe Henrich, Mark Kisin, Graham Noblit, Philipp Schoenegger, Suzanne Smith, Suren Tavakalov, and Alice Zhu for providing many helpful comments on this paper. Finally, he thanks Will MacAskill for introducing him to the reference [21], Barbara Finlay for introducing him to [44] and [46], Graham Noblit for introducing him to [50], and Joe Henrich for introducing him to [52]. This work was supported by the National Science Foundation through the Graduate Research Fellowship Program (grant number DGE1745303), the Centre for Effective Altruism through the Global Priorities Fellowship, and Harvard University through graduate student fellowships.

Competing Interests Statement. The author declares no conflict of interests.

The decision-theoretic model of cultural learning is detailed in the **Supplementary Information**.

References

1. Gintis, H. A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences* **30**, 1–16 (2007).
2. Muthukrishna, M. & Henrich, J. A problem in theory. *Nature Human Behaviour* **3**, 221–229 (2019).
3. Benjamin, D. J. Errors in probabilistic reasoning and judgment biases. in *Handbook of Behavioral Economics—Foundations and Applications 2* (eds. Bernheim, B. D., DellaVigna, S. & Laibson, D.) 69–186 (North-Holland, Amsterdam, 2019).
4. Agarwal, S., Driscoll, J. C., Gabaix, X. & Laibson, D. Learning in the credit card market. Preprint at https://www.nber.org/papers/w13822 (2008).
5. Haselhuhn, M. P., Pope, D. G., Schweitzer, M. E. & Fishman, P. The impact of personal experience on behavior: Evidence from video-rental fines. *Management Science* **58**, 52–61 (2012).
6. Corner, A. & Hahn, U. Normative theories of argumentation: Are some norms better than others? *Synthese* **190**, 3579–3610 (2012).
7. Kahneman, D. & Tversky, A. On the psychology of prediction. *Psychological Review* **80**, 237–251 (1973).
8. Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science* **185**, 1124–1131 (1974).
9. Kahneman, D., Slovic, P. & Tversky, A. *Judgment under Uncertainty: Heuristics and Biases*. (Cambridge University Press, 1982).
10. Dixon, N. F. *On the Psychology of Military Incompetence*. (Cape, 1976).
11. Johnson, D. D. P. *Overconfidence and War: The Havoc and Glory of Positive Illusions*. (Harvard University Press, 2004).
12. Scheinkman, J. A. & Xiong, W. Overconfidence and speculative bubbles. *Journal of Political Economy* **111**, 1183–1220 (2003).
13. Akerlof, G. A. & Shiller, R. J. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism.* (Princeton University Press, 2009).
14. Institute of Medicine Forum on Microbial Threats. Learning from pandemics past. in *Ethical and Legal Considerations in Mitigating Pandemic Disease* 31–60 (National Academies Press, 2007).

15. Lincoln, M. Study the role of hubris in nations' COVID-19 response. *Nature* **585**, 325–325 (2020).
16. Boyd, R. & Richerson, P. J. An evolutionary model of social learning: The effects of spatial and temporal variation. in *Social Learning: Psychological and Biological Perspectives* (eds. Zentall, T. R. & Galef, B. G.) 29–48 (Lawrence Erlbaum Associates, Inc., 1988).
17. Boyd, R. & Richerson, P. J. Why does culture increase human adaptability? *Ethology and Sociobiology* **16**, 125–143 (1995).
18. Lew-Levy, S., Reckin, R., Lavi, N., Cristóbal-Azkarate, J. & Ellis-Davies, K. How do hunter-gatherer children learn subsistence skills? A meta-ethnographic review. *Human Nature* **28**, 367–394 (2017).
19. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process*. (University of Chicago Press, 1985).
20. Cavalli-Sforza, L. L. & Feldman, M. W. *Cultural Transmission and Evolution: A Quantitative Approach*. (Princeton University Press, 1981).
21. Henrich, J. *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. (Princeton University Press, 2015).
22. Sanchez, C. & Dunning, D. Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology* **114**, 10–28 (2018).
23. Sanchez, C. & Dunning, D. Decision fluency and overconfidence among beginners. *Decision* **7**, 225–237 (2020).
24. Kruger, J. & Dunning, D. Unskilled and unaware of It: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* **77**, 1121–1134 (1999).
25. Haun, D. E., Zeringue, A., Leach, A. & Foley, A. Assessing the competence of specimen-processing personnel. *Laboratory Medicine* **31**, 633–637 (2000).
26. Burson, K. A., Larrick, R. P. & Klayman, J. Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology* **90**, 60–77 (2006).
27. Hoffman, M. & Burks, S. V. Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics* **11**, 315–348 (2020). In Figure 3 of the present paper, we have reprinted panel (a), titled "Means"; excised from Figure 1 of

Hoffman and Burks, titled "Overconfidence: comparing subjective productivity forecasts with actual worker productivity (as a function of worker tenure)". The original figure is available at https://onlinelibrary.wiley.com/action/downloadFigures?id=quan200021-fig-0001&doi=10.3982%2FQE834. The panel is reprinted under the Creative Commons Attribution Non-Commercial 4.0 International license, the details for which are available at https://creativecommons.org/licenses/by-nc/4.0/.

28. Augenblick, N. & Rabin, M. Belief movement, uncertainty reduction, and rational updating. *Quarterly Journal of Economics* **136**, 933–985 (2021).
29. Becker, G., Degroot, M. & Marschak, J. Measuring utility by a single-response sequential method. *Behavioral Science* **9**, 226–232 (1964).
30. Jansen, R. A., Rafferty, A. N. & Griffiths, T. L. A rational model of the Dunning–Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behaviour* (2021).
31. Ehrlinger, J., Johnson, K., Banner, M., Dunning, D. & Kruger, J. Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes* **105**, 98–121 (2008).
32. Bowles, S. Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* **324**, 1293–1298 (2009).
33. Lichtenstein, S. & Fischhoff, B. Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance* **20**, 159–183 (1977).
34. Moore, D. A. & Healy, P. J. The trouble with overconfidence. *Psychological Review* **115**, 502–517 (2008).
35. Kline, M. A., Boyd, R. & Henrich, J. Teaching and the life history of cultural transmission in Fijian villages. *Human Nature* **24**, 351–374 (2013).
36. Lew-Levy, S. *et al.* The life history of learning subsistence skills among Hadza and BaYaka foragers from Tanzania and the Republic of Congo. *Human Nature* (in press).
37. Boyd, R., Richerson, P. J. & Henrich, J. The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences* **108**, 10918–10925 (2011).
38. Pinker, S. The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences* **107**, 8993–8999 (2010).

39. Lew-Levy, S. & Boyette, A. H. Evidence for the adaptive learning function of work and work-themed play among Aka forager and Ngandu farmer children from the Congo Basin. *Human Nature* **29**, 157–185 (2018).
40. Salali, G. D. *et al.* Knowledge-sharing networks in hunter-gatherers and the evolution of cumulative culture. *Current Biology* **26**, 2516–2521 (2016).
41. Salali, G. D. *et al.* Development of social learning and play in BaYaka hunter-gatherers of Congo. *Scientific Reports* **9**, 11080–10 (2019).
42. Schniter, E., Gurven, M., Kaplan, H. S., Wilcox, N. T. & Hooper, P. L. Skill ontogeny among Tsimane forager-horticulturalists. *American Journal of Physical Anthropology* **158**, 3–18 (2015).
43. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences* **33**, 61–83 (2010).
44. Yarkoni, T. The generalizability crisis. *Behavioral and Brain Sciences*, 1–37 (2020).
45. de Francesco, G. *The Power of the Charlatan*. (Yale University Press, 1939).
46. Gladwell, M. *Talking to Strangers: What We Should Know about the People We Don't Know*. (Allen Lane, 2019).
47. Uscinski, J. E., Klofstad, C. & Atkinson, M. D. What drives conspiratorial beliefs? The role of informational cues and predispositions. *Political Research Quarterly* **69**, 57–71 (2016).
48. Scheirer, W. A pandemic of bad science. *Bulletin of the Atomic Scientists* **76**, 175–184 (2020).
49. Galanter, M. *Cults: Faith, Healing, and Coercion*. (Oxford University Press, 1989).
50. Boyd, R. & Richerson, P. J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology* **13**, 171–195 (1992).
51. Henrich, J. The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior* **30**, 244–260 (2009).
52. Singh, M. The cultural evolution of shamanism. *Behavioral and Brain Sciences* **41**, 1–83 (2018).
53. Lindenbaum, S. Kuru, prions, and human affairs: Thinking about epidemics. *Annual Review of Anthropology* **30**, 363–385 (2001).

54. Jones, H., Diop, N., Askew, I. & Kabore, I. Female genital cutting practices in Burkina Faso and Mali and their negative health outcomes. *Studies in Family Planning* **30**, 219–230 (1999).
55. Wagner, N. Female genital cutting and long-term health consequences – Nationally representative estimates across 13 countries. *Journal of Development Studies* **51**, 226–246 (2015).
56. Henrich, J. & Muthukrishna, M. The origins and psychology of human cooperation. *Annual Review of Psychology* **72**, 207–240 (2021).
57. Shennan, S. J., Crema, E. R. & Kerig, T. Isolation-by-distance, homophily, and "core" vs. "package" cultural evolution models in Neolithic Europe. *Evolution and Human Behavior* **36**, 103–109 (2015).
58. Yeh, D. J., Fogarty, L. & Kandler, A. Cultural linkage: The influence of package transmission on cultural dynamics. *Proceedings of the Royal Society. B, Biological Sciences* **286**, 20191951–20191951 (2019).
59. Mercier, H. & Sperber, D. *The Enigma of Reason*. (Harvard University Press, 2018).
60. Tomasello, M. The ontogenetic foundations of epistemic norms. *Episteme* **17**, 301–315 (2020).
61. Angrist, J. & Krueger, A. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* **106**, 979–1014 (1991).
62. Harmon, C. & Walker, I. Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* **85**, 1278–1286 (1995).
63. Acemoglu, D. & Angrist, J. How large are human-capital externalities? Evidence from compulsory schooling laws. in *NBER Macroeconomics Annual 2000* (eds. Bernanke, B. S. & Rogoff, K.) vol. 15, 9–59 (MIT Press, 2000).
64. Duflo, E. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* **91**, 795–813 (2001).
65. Oreopoulos, P. The compelling effects of compulsory schooling: Evidence from Canada. *Canadian Journal of Economics* **39**, 22–52 (2006).
66. Oreopoulos, P. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* **96**, 152–175 (2006).

67. Aakvik, A., Salvanes, K. G. & Vaage, K. Measuring heterogeneity in the returns to education using an education reform. *European Economic Review* **54**, 483–500 (2010).

68. Unterman, R. *Headed to College: The Effects of New York City's Small High Schools of Choice on Postsecondary Enrollment*. https://www.mdrc.org/sites/default/files/Headed_to_College_PB.pdf (2014).

69. Barrow, L., Schanzenbach, D. W. & Claessens, A. The impact of Chicago's small high school initiative. *Journal of Urban Economics* **87**, 100–113 (2015).

70. Meghir, C. & Palme, M. Educational reform, ability, and family background. *American Economic Review* **95**, 414–424 (2005).

71. Evan, A. *et al. Evaluation of the Bill & Melinda Gates Foundation's High School Grants Initiative. 2001–2005 Final Report*. https://docs.gatesfoundation.org/documents/year4evaluationairsri.pdf (2006).

72. Pischke, J.-S. & Wachter, T. Zero returns to compulsory schooling in Germany: Evidence and interpretation. *Review of Economics and Statistics* **90**, 592–598 (2008).

73. Grenet, J. Is extending compulsory schooling alone enough to raise earnings? Evidence from French and British compulsory schooling laws. *Scandinavian Journal of Economics* **115**, 176–210 (2013).

74. Stephens, M. & Yang, D.-Y. Compulsory education and the benefits of schooling. *American Economic Review* **104**, 1777–1792 (2014).

75. Hayek, F. A. The use of knowledge in society. *American Economic Review* **35**, 519–530 (1945).

76. Hornsey, M. J., Harris, E. A., Bain, P. G. & Fielding, K. S. Meta-analyses of the determinants and outcomes of belief in climate change. *Nature Climate Change* **6**, 622–626 (2016).

77. Insler, M., McQuoid, A. F., Rahman, A. & Smith, K. A. Fear and loathing in the classroom: Why does teacher quality matter? Preprint at https://www.iza.org/publications/dp/14036/fear-and-loathing-in-the-classroom-why-does-teacher-quality-matter (2021).

78. Weinberg, B. A., Hashimoto, M. & Fleisher, B. M. Evaluating teaching in higher education. *Journal of Economic Education* **40**, 227–261 (2009).

79. Boserup, E. *Population and Technological Change: A Study of Long-Term Trends*. (University of Chicago Press, 1981).

80. Kremer, M. Population growth and technological change: one million B.C. to 1990. *Quarterly Journal of Economics* **108**, 681–716 (1993).

81. Maddux, W. W., Adam, H. & Galinsky, A. D. When in Rome ... learn why the Romans do what they do: How multicultural learning experiences facilitate creativity. *Personality and Social Psychology Bulletin* **36**, 731–741 (2010).

82. Muthukrishna, M. & Henrich, J. Innovation in the collective brain. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **371**, 20150192 (2016).

83. Selten, R. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* **4**, 25–55 (1975).

84. Hooper, P. L., Demps, K., Gurven, M., Gerkey, D. & Kaplan, H. S. Skills, division of labour and economies of scale among Amazonian hunters and South Indian honey collectors. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **370**, 20150008 (2015).

85. Bird-David, N. Beyond 'The hunting and gathering mode of subsistence': Culture-sensitive observations on the Nayaka and other modern hunter-gatherers. *Man* **27**, 19–44 (1992).

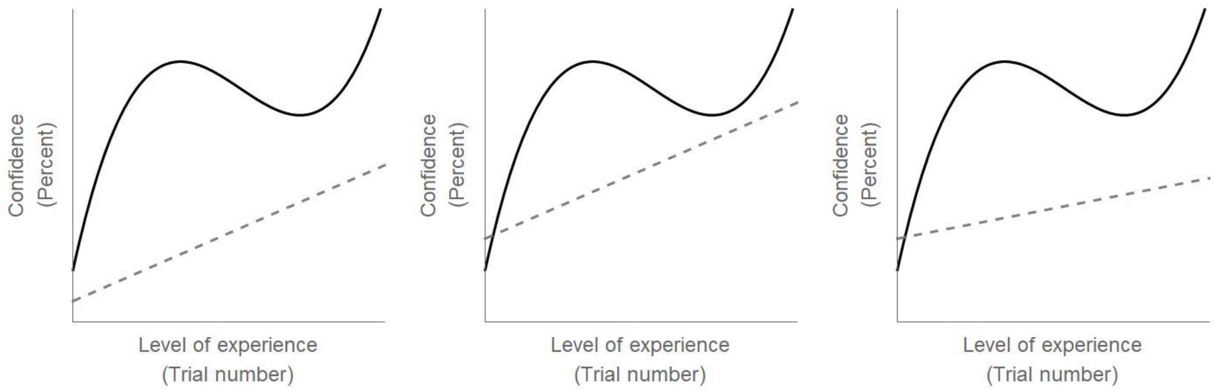


Fig. 1. Qualitative depictions of the data—the cubic fit of confidence (solid) and the linear fit of accuracy (dashed) over the number of task attempts—from different variants of the Sanchez–Dunning experiment[22,23]. The confidence graph was somehow largely independent of the subjects' actual past performance, environmental feedback that varied across the different variants of the experiment. First, the confidence graph was initially increasing in both experimental variants where it started out below the true value—Studies 1, 3, and 4 from Sanchez and Dunning's first paper and Study 1 from their second paper—and variants where it started out above the true value—Study 2 from their first paper and Study 2 from their second paper. This is depicted in the comparison between the center panel and the left panel. Second, the increase in confidence over time was comparable between experimental variants with a high increase in accuracy over time—Studies 2 and 3 from their first paper and Study 2 from their second paper—and variants with a low such increase—Studies 1 and 4 from their first paper and Study 1 from their second paper. This is depicted in the comparison between the center panel and the right panel.

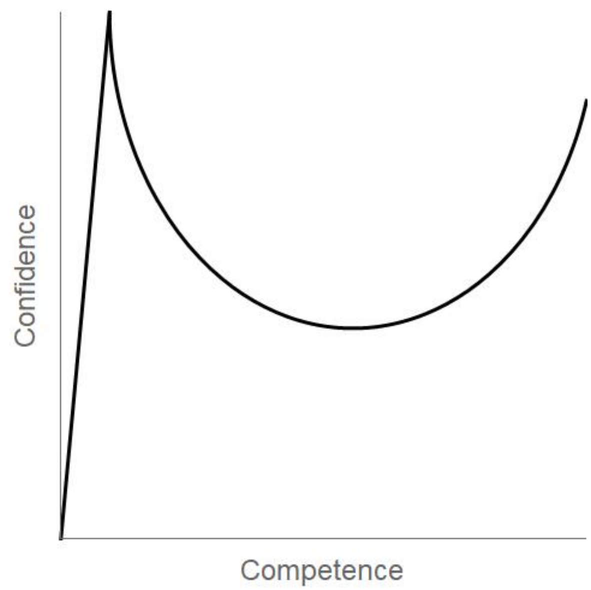


Fig. 2. Qualitative Dunning–Kruger graph depicting non-monotonic confidence. An unfortunate consequence of this non-monotonicity is that a person's confidence may not predict her actual ability; a novice may be overconfident to the point of estimating her ability to be at the same level as that of an expert.

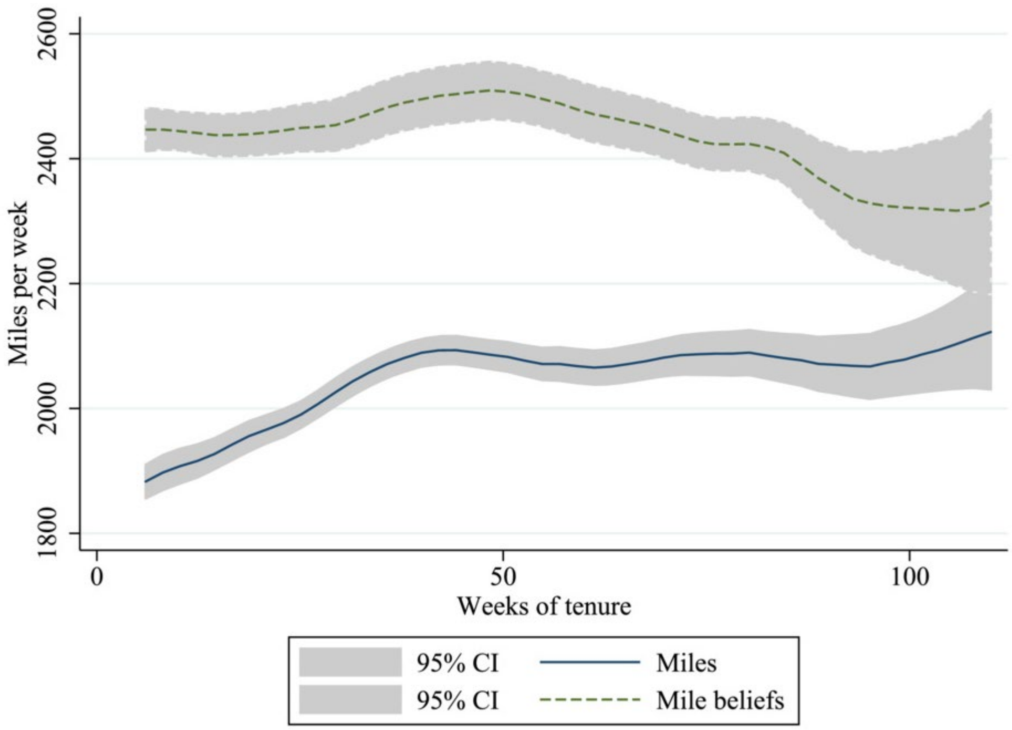


Fig. 3. A reprinting of panel (a), titled "Means"; excised from Figure 1 of Hoffman and Burks[27], titled "Overconfidence: comparing subjective productivity forecasts with actual worker productivity (as a function of worker tenure)." In their study, truckers' mean self-estimated productivity (dashed line) was non-monotonic with respect to the level of experience, with an initial phase of increase followed by a phase of decrease. Due to a lack of empirical data on higher levels of experience, the hypothesized final phase of increase is neither confirmed nor contradicted. Meanwhile, truckers' mean true productivity (solid line) was monotonically increasing in the level of experience. Bayesian learning predicts that self-estimated productivity will rapidly converge to the true productivity, and that this convergence will begin immediately. Neither prediction holds empirically.

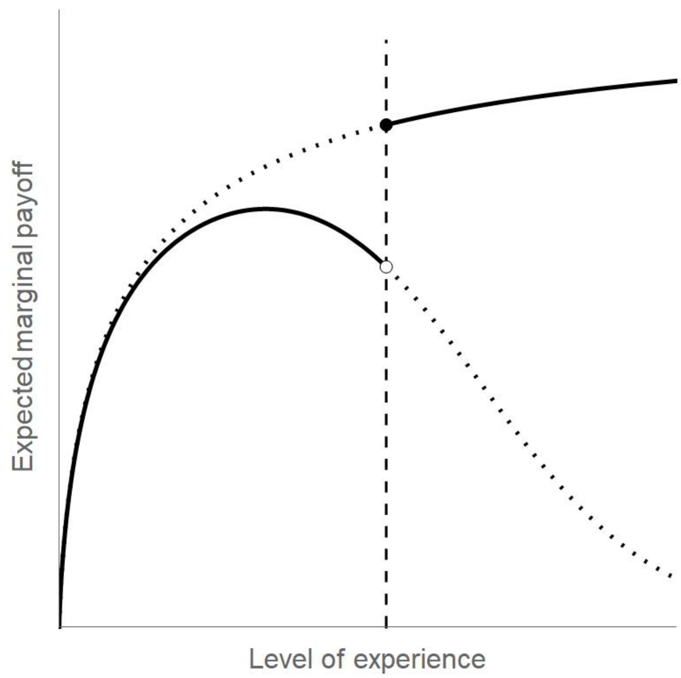


Fig. 4. The learner's optimal estimate of her expected marginal payoff—a function of her level of experience—in an illustrative parametrization of our evolutionary model of human learning (the specific model parameters can be found in Fig. S1, from which this graph is obtained). After accumulating task-specific knowledge up to a certain intermediate level, it is in her best interest to quit for an opportunity-cost task with a different teacher. At that point, the risk that her current teacher is a charlatan—and thus cannot guarantee that the current task is possible to learn how to do—is too great to ignore.

Supplementary Information for
The Evolution of Cognitive Biases in Human Learning

Peter S. Park*

*Department of Mathematics, Harvard University,

1 Oxford St., Cambridge, MA 02138, USA

Email: pspark@math.harvard.edu.

In the following text, we will develop our decision-theoretic model of cultural learning—learning from knowledge provided by one's fellow group members[19–21]—in settings of high-variance environmental feedback. Our main results, each proven under quite non-restrictive assumptions on the model parameters, can be colloquially described as follows:

- When engaged in imitation learning (in our setting, learning from a teacher who is a genuine expert in the task at hand), the student should never quit on her current task and teacher. Her expected payoff is monotonically increasing in the level of experience (Proposition 1).
- When engaged in innovation learning (in our setting, ostensibly learning from a charlatanical teacher while actually learning from individual epiphanies), the student should always quit on her current task and teacher—and begin learning an opportunity-cost task from a new teacher—at an intermediate level of experience. Her expected payoff eventually decreases to zero as a function of the level of experience (Proposition 2).
- When in one of the two aforementioned situations without knowing precisely which one, the student should continue to learn until an intermediate level of experience, perform a time-measurement experiment to determine whether she is learning by imitation or innovation (whether her teacher is an actual expert or a charlatan) at that intermediate point, quit if the latter is true, and continue to learn if the former is true. Her expected payoff

is a piecewise defined function that is eventually decreasing before the aforementioned intermediate point, but is monotonically increasing after it. Consequently, this function can be non-monotonic in the desired way: increasing at first, then decreasing, and finally reverting back to increase (Proposition 5).

These results will first be proven for a continuous approximation of the student's discrete decision-theoretic environment. Then, we will show that if this discrete environment is "sufficiently fine"—so that the continuous approximation is sufficiently good—the results proven in the continuous case carry over to the discrete case, as desired.

1 Decision-theoretic model of cultural learning

In our evolutionary model of cultural learning, the human decision-maker (DM) faces a task that requires an amount $a \in \mathbb{R}_{>0} \cup \{\infty\}$ of knowledge to master. She knows $b \leq a$ of this amount. The values of b and a determine her expected marginal payoff $f_a(b)$. By scaling, we suppose that $f_a : [0, a] \to [0, 1]$ (in general, the range of the map might be different), which we also assume is continuous with respect to both a and b. When fixing b, the payoff $f_a(b)$ is decreasing in a; when fixing a, it is increasing in b. We suppose that $f_a(a) = 1$, the maximum payoff. All payoffs are subject to time-discounting by an exponential factor $\delta \in (0, 1)$.

The amount of knowledge that the DM possesses, b, increases over time in discrete jumps, each following the acquisition of a payoff in the form of a high-variance lottery. Thus, b is a sufficient statistic for the DM's level of experience spent in the phase of learning (not including the phase of having completed learning). It increases as a *discrete learning function*

$$L_{\mathbb{S}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0},$$

a right-continuous step function for which $L_{\mathbb{S}}(0) = 0$ and $\lim_{t \to \infty} L_{\mathbb{S}}(t) = \infty$. Specifically,

$$b = \min(L_{\mathbb{S}}(t), a),$$

where t denotes the point in time. The step intervals $[t_{i-1}, t_i]$ of $L_{\mathbb{S}}$ represent the *learning periods*, each of which comprises one of the DM's attempts at the task. At the end of a learning period $[t_{i-1}, t_i]$, it yields a payoff of expected value

$$f_a\left(\min(L_{\mathbb{S}}(t_{i-1}), a)\right) \int_{t_{i-1}}^{t_i} \delta^t dt, \tag{1}$$

after which b jumps discontinuously to the next value of $L_{\mathbb{S}}$.

A Bayesian-rational agent would be able to learn from payoff observations. An obvious example of this occurs when the payoffs are realized with no uncertainty. Then, the observed values of $f_a(b)$ may comprise a sufficient statistic for the true value of a. However, humans by default may not retain observations of payoffs when they occur in the form of high-variance lotteries, a behavior that likely evolved in response to unfavorable fitness tradeoffs from overcommitting attention. Thus, only the information in the complement of payoff observations—which we aggregate into an abstract statistic, the level b of knowledge—is available to the human DM in our setting.

Let $t_1 < t_2 < \cdots$ denote the jump discontinuities of $L_{\mathbb{S}}$, which bound the learning periods. Let $b_i = L_{\mathbb{S}}(t)$ for $t_i \leq t < t_{i+1}$ denote the level of knowledge that the DM reaches after the ith learning period, and $\Delta_i = t_i - t_{i-1}$ denote the length of time of the ith learning period, where we have used the notation $t_0 = 0$ and $b_0 = 0$. Note that pairs of sequences $\{\Delta_i\}_{i>0}$ and $\{b_i\}_{i>0}$, where all $\Delta_i > 0$ and $0 < b_1 < b_2 < \cdots$, bijectively correspond to discrete learning functions. The DM's total payoff from time 0 to $t^* \in [0, \infty) \cup \{\infty\}$ is given by

$$\Pi_{L_{\mathbb{S}}}(t^*) = \sum_{i>0: t_i \leq t^*} f_a\left(\min(L_{\mathbb{S}}(t_{i-1}), a)\right) \int_{t_{i-1}}^{t_i} \delta^t dt.$$

Since the DM cannot retain payoff observations by assumption, we can employ a continuous approximation of nature, assuming its discrete learning function is sufficiently fine. Specifically, we consider a sequence of discrete learning functions $L_{\mathbb{S}_n}$ that monotonically converges to a *continuous learning function* $L : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, a monotonically increasing bijective function. The DM's payoff from time 0 to t^* in this continuous approximation is given by

$$\Pi_L(t^*) = \int_0^{t^*} \delta^t f_a(\min(L(t), a)) dt,$$

to which $\Pi_{L_{\mathbb{S}_n}}(t^*)$ converges as $n \to \infty$, by the dominated convergence theorem. We will work with a continuous learning model (Table S2) as a tractable approximation of a sufficiently fine discrete learning model (Table S1), which we propose as a veridical model of human learning in the evolutionary past.

2 The continuous learning model

The DM is faced with two types of tasks: those learned from innovation while under the wings of a charlatanical teacher (*tasks of the first type*), which occur with probability $1-q$; and those learned from imitating an actual expert teacher (*tasks of the second type*), which occur with probability q. In every continuous learning model and discrete learning model, each of the two task types has its own learning function and probability distribution of difficulty values. Formally, a *continuous learning model* (L_1, L_2) is defined by the data of the continuous learning functions L_j of tasks of the jth type, the probability distributions μ_j of difficulty values a of jth-type tasks, the aforementioned probability value $q \in [0,1]$, and the payoff functions $f_a(\cdot)$. A *discrete learning model* $(L_{\mathbb{S}}, L_{\mathbb{S}'})$ is defined by the same data, except the continuous learning functions L_j are replaced with discrete learning functions: $L_{\mathbb{S}}$ for tasks of the first type and $L_{\mathbb{S}'}$ for tasks of the second type. Additionally, we require that the values b_i, to which the DM's level of knowledge jumps at the end of the ith learning period, are shared between $L_{\mathbb{S}}$ and $L_{\mathbb{S}'}$.

Consider a sequence of discrete learning models $\{(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})\}_{n>0}$. From this point on, we denote the points of time bounding the learning periods of $L_{\mathbb{S}_n}$ by $\{t_{n,i}\}_{i\geq 0}$, those of $L_{\mathbb{S}'_n}$ by $\{t'_{n,i}\}_{i\geq 0}$, the lengths of the learning periods of $L_{\mathbb{S}_n}$ by $\{\Delta_{n,i}\}_{i>0}$, those of $L_{\mathbb{S}'_n}$ by $\{\Delta'_{n,i}\}_{i>0}$, and the knowledge values—which are shared by $L_{\mathbb{S}_n}$ and $L_{\mathbb{S}'_n}$, as we have assumed—by $\{b_{n,i}\}_{i\geq 0}$. Then, we say that the sequence of discrete learning models $\{(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})\}_{n>0}$ *converges to* a continuous learning model (L_1, L_2) if the following conditions hold.

- The sequence of functions $\{L_{\mathbb{S}_n}\}_{n>0}$ monotonically converges (increasing with respect to n) to L_1 in a way such that $L_1(t_{n,i}) = b_{n,i}$ for all n and i.
- The difficulty values of tasks of the first type are given by the same distribution μ_1 for L_1 and all $L_{\mathbb{S}_n}$.
- The sequence of functions $\{L_{\mathbb{S}'_n}\}_{n>0}$ monotonically converges (increasing with respect to n) to L_2 in a way such that $L_2(t'_{n,i}) = b_{n,i}$ for all n and i.
- The difficulty values of tasks of the second type are given by the same distribution μ_2 for L_2 and all $L_{\mathbb{S}'_n}$.

- The probability of tasks of the second type q and the payoff functions $f_a(\cdot)$ are shared by (L_1, L_2) and all $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$.

The two types of tasks differ in the following ways. First, tasks of the first type have positive probability p on the event that $a = \infty$: in other words, that the task—due to its impossibility—can never be learned to completion and always give a marginal payoff of $f_\infty(b) = 0$. Tasks of the second type do not. Second, imitation learning is faster than innovation learning. In the continuous learning model (L_1, L_2), the latter difference corresponds to the following condition.

Assumption C1 We have $L_1(t) \leq L_2(t)$ for all $t \in \mathbb{R}_{\geq 0}$.

When a sequence of discrete learning models $\{(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})\}_{n>0}$ converges to the continuous learning model (L_1, L_2), the above condition on (L_1, L_2) follows from an analogous one applied to $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$.

Assumption D1 We have $\Delta_{n,i} \geq \Delta'_{n,i}$ for all n and i.

Recall our assumption that in the discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$, the knowledge jumps from $b_{n,i-1}$ to $b_{n,i}$ that occur during innovation learning are indistinguishable from those that occur during imitation learning. Thus, the only way to tell the two types of tasks apart is by observing whether the amount of time spent in the ith learning period is $\Delta_{n,i}$ or the possibly smaller value $\Delta'_{n,i}$. Furthermore, this can only be done when these two values are distinct.

In the approximating continuous learning model (L_1, L_2), the DM's information set in an infinitesimal period of time is comprised of her level of knowledge b on her current task, whether b has reached a and thus stopped increasing, and whether she has ruled out the possibility that $j = 1$ or $j = 2$. The latter will be subject to a constraint detailed in Assumption C2.

It remains to define the DM's action space in an infinitesimal period of time within the continuous learning model (L_1, L_2). She maximizes her expected total payoff within the continuous learning game's overall action space—which only pertains to quitting—and this payoff determines the main term (as $n \to \infty$) of her optimal expected total payoff in the approximated discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$. Later, we will define the discrete learning model's finer action space—containing strategic considerations other than quitting—which determines the error term. Furthermore, it explains the evolutionary benefit of both having an accurate innate estimate of the expected marginal payoff and ruling out the possibility that the current task is of the first type at the latest possible moment: right before quitting.

As we have alluded to above, the DM's action space in an infinitesimal period of time within the continuous learning model (L_1, L_2) is comprised of two options: continuing to learn the current task (the default option, which in particular is always taken when b has reached a) and quitting for an opportunity-cost task (which resets b to zero). Whenever the latter action is taken, as well as in the very beginning, a task (j, a) is drawn from the distribution μ of tasks, each of which is comprised of a type $j \in \{1, 2\}$ and a difficulty value $a \in \mathbb{R}_{>0} \cup \{\infty\}$. When a task is drawn from μ, the value of j is first determined: $j = 1$ with probability $1 - q$ and $j = 2$ with probability q. Then, the value of a is drawn from the distribution μ_j of tasks of the jth type. The DM begins learning every task without knowing its type j. We introduce the following constraint to identifying j and quitting that is exogenous to the continuous learning model, but will be—in the following section—endogenously justified for the discrete learning model that it approximates.

Assumption C2 To identify j, the DM must commit to quitting the task immediately afterwards if she finds out that $j = 1$. She can only perform this identification when her level of knowledge b is $\geq \beta$ for a fixed $\beta > 0$.

We qualitatively describe the DM's optimal quitting strategy when all tasks are of the second type ($q = 1$), of the first type ($q = 0$), or are nontrivially divided between the two types ($0 < q < 1$). When $q = 1$, the DM never quits. When $q = 0$, the DM always quits at a finite level of knowledge. Finally, when $0 < q < 1$, the DM never quits tasks of the second type, and quits tasks of the first type at a finite level of knowledge. The constraint imposed by Assumption C2 plays a vital role in determining the optimal quitting strategy in the latter case.

2.1 The case of $q = 1$ (no tasks of infinite difficulty)

Suppose that all tasks are of the second type, or functionally speaking, that there are no tasks of infinite difficulty, $a = \infty$. Then, consider the following assumption on the payoff functions $f_a(\cdot)$.

Assumption C3 The payoff functions $f_a(\cdot)$ satisfy

$$f_{b+m}(b) < f_{b'+m}(b')$$

for all $b < b'$ and $m > 0$.

This axiom is plausible because a fixed amount of knowledge m constitutes a larger fraction of the total knowledge of an easy task than of a difficult task; consequently, not knowing it would cause a harsher penalty in the former case.

When the distribution $\mu = \mu_2$ of task difficulties is exponential, Assumption C3 implies the first-order stochastic dominance of being at a level of knowledge $\overline{b}$ that has not yet caught up to the true task difficulty a, over being at a level of knowledge $\underline{b} < \overline{b}$ that has also not yet caught up to a. From this, we deduce two conclusions. First, the *expected payoff function*

$$g(b) = \frac{\int_{a>b} f_a(b) d\mu(a)}{\int_{a>b} d\mu(a)},$$

defined by the expected value of the marginal payoff at level of knowledge b that has not yet caught up to a, is monotonically increasing. Second, define the *quitting value function* $V(b)$ by the expected total payoff from quitting at level of knowledge $b \in \mathbb{R}_{\geq \beta} \cup \{\infty\}$ that has not yet caught up to a. It is given by the solution to

$$\begin{aligned} V = &\int_0^b \left(\int_0^{L_2^{-1}(a)} \delta^t f_a(L_2(t)) dt + \int_{L_2^{-1}(a)}^{\infty} \delta^t dt \right) d\mu(a) \\ &+ \int_b^{\infty} \left(\int_0^{L_2^{-1}(b)} \delta^t f_a(L_2(t)) dt + \delta^{L_2^{-1}(b)} V \right) d\mu(a). \end{aligned}$$

Then, $V(b)$ is uniquely maximized at $b = \infty$.

Proposition 1. *Suppose $q = 1$ and $\mu = \mu_2$ is an exponential distribution of decay factor $\eta \in (0, 1)$. Under Assumption C3, the following are true.*

a) The quitting value function $V(b)$ is uniquely maximized at $b = \infty$.

b) The expected payoff function $g(b)$ is strictly increasing.

In other words, if all tasks were learnable in finite time, the DM's expected marginal payoff would be monotonically increasing in the level of knowledge.

2.2 The case of $q = 0$ (with tasks of infinite difficulty)

Suppose that all tasks are of the first type, or functionally speaking, that tasks with infinite difficulty $a = \infty$ occur with positive probability. Specifically, we assume that $\mu = \mu_1$ is the distribution with probability $p \in (0,1)$ on the event $a = \infty$ and the remaining probability distributed exponentially on $\mathbb{R}_{>0}$ with decay factor η. The quitting value function $V(b)$ is then given by the solution to

$$V = \int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^{\infty} \delta^t dt \right) d\mu(a) + \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t))dt + \delta^{L_1^{-1}(b)} V \right) d\mu(a),$$

Consider the following mild differential condition.

Assumption C4 The learning function $L_1(\cdot)$ is continuously differentiable and satisfies

$$L_1'(t) \ll \eta^{-L_1(t)}$$

as $t \to \infty$.

This asymptotic upper bound on the derivative $L_1'(t)$ guarantees that the introduction of infinite-difficulty tasks into the continuous learning model changes $b = \infty$, which previously was the unique optimal quitting point, to be suboptimal. Instead, every optimal quitting point of $V(\cdot)$ is finite. Assumption C4 is reasonable; it is immediately satisfied, for example, if the continuously differentiable learning function L_1 is Lipschitz.

Another mild differential condition—which constrains the derivatives of the payoff functions $f_a(b)$ from being uniformly too large—guarantees that the expected payoff function $g(b)$ is monotonically decreasing for all sufficiently large b.

Assumption C5 The payoff functions $f_a(\cdot)$ are continuously differentiable and satisfy

$$\int_{a>b} f_a'(b)\eta^a da \ll \eta^b$$

as $b \to \infty$.

Overall, we have the following non-monotonicity.

Proposition 2. *Suppose $q = 0$ and $\mu = \mu_1$ is the distribution with probability $p > 0$ on the event $a = \infty$ and the remaining probability distributed exponentially on $\mathbb{R}_{>0}$ with decay factor $\eta \in (0,1)$.*

a) *Under Assumption C4, $V(b)$ is strictly decreasing for all sufficiently large b. In particular, the one or more quitting points b^* that maximize $V(\cdot)$ are finite.*

b) *The expected payoff function $g(b)$ converges to zero as $b \to \infty$. In particular, $g(\cdot)$ attains its maximum at one or more finite points b.*

c) *Under Assumption C5, $g(b)$ is strictly decreasing for all sufficiently large b.*

In other words, allowing impossible tasks to occur with positive probability prevents the expected payoff function from being monotonically increasing. Also, the optimal quitting point(s) b^* must be finite. In fact, we can show that they can be made arbitrarily large in our continuous learning model by suitably shrinking the probability mass p on the event $a = \infty$. Colloquially, it is feasible that the DM does not quit soon, for arbitrarily generous notions of "soon."

Corollary 3. *Suppose Assumptions C2 and C3 hold, retain the setting of Proposition 2, and fix all choices except that of p. For every $\gamma \geq \beta > 0$, there exists p such that any optimal quitting point b^* of Proposition 2's game for which the event $a = \infty$ has probability p satisfies*

$$b^* > \gamma.$$

To prove this, we will use the following lemma, a comparative-statics result which is also of independent interest.

Lemma 4. *For any fixed $b \in [\beta, \infty) \cup \{\infty\}$, the value of $V(b)$ is strictly decreasing with respect to p.*

2.3 The mixed case of $0 < q < 1$ (with both innovation and imitation learning)

We now consider the mixed case, which we offer as an approximation of a veridical model of ancestral humans' learning environment. The task distribution is defined by

$$\mu(1, a) = (1 - q)\mu_1(a) \quad \text{and} \quad \mu(2, a) = q\mu_2(a), \tag{2}$$

where μ_1 and μ_2 are of the form defined in Propositions 2 and 1, respectively.

The proof of Proposition 1 (provided in Subsection 4.1) shows that the DM always prefers being at a positive level of knowledge b (that has not yet caught up to a) for a task of the second type, to being at zero knowledge for a task of the second type, which is—by Assumption C1—always preferred to being at zero knowledge for a task of the first type. Thus, any optimal quitting strategy must be of the following form: never quit tasks of the second type, and quit tasks of the first type only when the level of knowledge b reaches some value $b^* \in [\beta, \infty) \cup \{\infty\}$ before catching up to a. Denote this by the *quitting strategy* b^*. Recall the following content of Assumption C2: when employing quitting strategy b^*, the DM has not ruled out either $j = 1$ or $j = 2$ when $b < b^*$; but when $b = b^*$, she finds out the type of the task, quits if $j = 1$, and does not quit if $j = 2$.

Define the *conditional quitting value functions* $V_1(b)$ *and* $V_2(b)$ by the expected total payoff from quitting at level of knowledge b that has not yet caught up to a, conditional on the task being of type $j = 1$ or $j = 2$, respectively. Also, define the *unconditional quitting value function* $V_u(b) = (1-q)V_1(b) + qV_2(\infty)$ by the expected total payoff from employing quitting strategy b: that of quitting tasks of the first type at level of knowledge b that has not yet caught up to a.

Similarly, the *conditional expected payoff functions* $g_1(b)$ and $g_2(b)$ are defined by the expected marginal payoff at level of knowledge b that has not yet caught up to a, conditional on the task being of type $j = 1$ or $j = 2$, respectively. Also, define the *unconditional expected payoff function*

$$g_u(b) = \frac{\int_{a>b} f_a(b) d\mu(j,a)}{\int_{a>b} d\mu(j,a)}$$

by the expected marginal payoff at level of knowledge b that has not yet caught up to a, unconditional on the task type. Distinguish the latter from the *true expected payoff function* $g_{b^*}(b)$ *associated to quitting strategy* b^*, defined by the piecewise function given by $g_u(b)$ if $b < b^*$ and $g_2(b)$ if $b \geq b^*$.

Proposition 5. *Suppose $0 < q < 1$ and μ is given by (2) such that μ_1 is the distribution with probability $p > 0$ on the event $a = \infty$ and the remaining probability distributed exponentially on $\mathbb{R}_{>0}$ with decay factor $\eta \in (0,1)$, and μ_2 is the exponential distribution of decay factor η. Under Assumptions C1 and C2, the following are true.*

a) The conditional quitting value function $V_2(b)$ is uniquely maximized at $b = \infty$ under Assumption C3, while under Assumption C4, the unconditional quitting value function $V_u(b)$ is strictly decreasing for all sufficiently large b. In particular, the one or more *quitting strategies b^* maximizing*

$$\max_{b \in \mathbb{R}_{\geq \beta} \cup \{\infty\}} V_u(b)$$

are finite under Assumption C4.

b) *The conditional expected payoff function $g_2(b)$ is strictly increasing under Assumption C3, while the unconditional expected payoff function $g_u(b)$ converges to zero as $b \to \infty$.*

c) *Under Assumption C5, $g_u(b)$ is strictly decreasing for all sufficiently large b.*

Note that if $\beta = 0$, then the optimal quitting strategy would always be $b^* = 0$. In other words, only tasks of the second type would ever be learned, and the associated true expected payoff function $g_{b^*}(b)$ would be monotonically increasing. Consequently, a non-monotonic true expected payoff function $g_{b^*}(b)$ requires that $\beta > 0$.

Recall our feasibility result, Corollary 3, that making the probability p of the event $a = \infty$ sufficiently small guarantees that the DM does not quit soon, for arbitrarily generous notions of "soon." We show an analogous feasibility result. By making both p and the probability q of tasks of the second type sufficiently small in our continuous learning model, we can guarantee that the DM does not quit soon in our current setting.

Corollary 6. *Suppose Assumptions C1–C3 hold, retain the setting of Proposition 5, and fix all choices except those of p and q. For every $\gamma \geq \beta > 0$, there exist p and q such that any optimal quitting strategy b^* of Proposition 5's game for which tasks of the first type have difficulty $a = \infty$ with probability p and tasks of the second type occur with probability q satisfies*

$$b^* > \gamma.$$

To prove this, we will use the following lemma, a comparative-statics result which is also of independent interest.

Lemma 7. *For any fixed $b \in [\beta, \infty) \cup \{\infty\}$, the value of $V_u(b)$ is strictly decreasing with respect to p and strictly increasing with respect to q.*

In summary, for any optimal quitting strategy b^*, which can be made arbitrarily late by a suitable choice of model parameters, the associated true expected payoff function $g_{b^*}(b)$ is a piecewise-defined function that is given by an increasing function, $g_2(b)$, for $b \geq b^*$; and by a function that is eventually decreasing, $g_u(b)$, for $b < b^*$ (see Fig. S1).

3 Limit of discrete learning models

Regarding the approximation of a fine discrete learning model with a continuous learning model, two issues remain. First, we have not yet justified the exogenous Assumption C2. Second, the continuous learning model does not explain the evolutionary benefit of innately possessing a then-accurate estimate of the expected marginal payoff. We address both of these issues by introducing *side opportunities* that provide positive, but negligible payoffs; and a less negligible, endogenous cost to identifying the type j of the current task by a time-measurement experiment. These two negligible factors can be thought of as mathematically analogous to trembles—small risks that agents in a strategic game may play unintended strategies by accident—in the notion of trembling-hand perfect equilibria[83]. Our overall story then holds for a sufficiently fine discrete learning model in a sequence of discrete learning models converging to a continuous learning model.

Ethnographic evidence shows that even in hunter-gatherer societies like those of the evolutionary past, humans take on highly specialized roles, which require copious amounts of experience and knowledge[84]. These specializations can be thought of as analogous to modern workers' full-time jobs. However, hunter-gatherers also face incentives to be opportunistic: to accurately appraise—and based on the result of said appraisal, possibly procure—additional foraging opportunities as they arise[85]. These opportunities can be thought of as analogous to the side gigs that modern workers do in addition to their jobs. It is thus realistic for our learning model, which in its continuous form only considered the tasks analogous to full-time jobs, to also consider unrelated opportunities analogous to side gigs.

To the discrete form of our learning model, we proceed to add side opportunities that require an accurate estimation of payoffs from the default foraging task to optimally exploit—taking such an opportunity only when its marginal payoff exceeds the expected marginal payoff of the default foraging method—but whose payoff values decay to zero in the continuous limit. In other words, the continuous approximation (L_1, L_2) of nature's discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$ creates a lexicographic evolutionary objective. Payoffs from the default foraging method comprise the $\Theta(1)$ main term, and payoffs from side opportunities comprise the $o(1)$ error term. The quitting strategy affects the main term, and is thus of primary importance. The strategy of which side opportunities should be taken in lieu of the default option does not affect the main term; it only affects the error term, and is thus of secondary importance.

Specifically, we suppose that for every discrete learning model in our sequence of $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$

converging to the continuous learning model (L_1, L_2), the ith learning period contains a side opportunity whose procurement requires an expected fraction $R_{n,i} \in (0,1)$ of the learning period's time (taking into account the time-discounting). The DM can choose whether or not to devote an expected fraction $R_{n,i}$ of the learning period's time to obtaining the side opportunity's marginal payoff P, drawn from some random distribution whose support is $[0,1]$. In other words, the DM chooses between the default expected payoff (1) and its altered form,

$$\left(R_{n,i}P + (1 - R_{n,i})\, f_a\left(\min(b_{n,i-1}, a)\right)\right) \int_{t_{n,i-1}}^{t_{n,i}} \delta^t dt,$$

that results from taking the side opportunity. Then, an accurate mental estimate $g(\cdot)$ of the expected marginal payoff from the default foraging method allows for the maximization of additional payoffs from side opportunities. Indeed, the DM should procure the side opportunity if $g(b_{n,i-1}) < P$ and forgo it if $g(b_{n,i-1}) > P$.

The following assumption endogenizes the above discussion that side opportunities provide positive, but negligible additional payoffs. In the continuous limit, the total additional payoff from side opportunities decays to zero and thus ceases to factor into the optimization problem

$$\max_{b \in \mathbb{R}_{\geq \beta} \cup \{\infty\}} V_u(b)$$

of the quitting strategy.

Assumption D2 As $n \to \infty$, the expected fraction $R_{n,i}$ of the ith learning period's time that can be used for a side opportunity is uniformly $o(1)$.

It remains to endogenize Assumption C2 with an explanation of why the DM switches priors—from $j \in \{1,2\}$ to either $j=1$ or $j=2$—only just before quitting at the optimal level of knowledge b^*. Consider our sequence of discrete learning models $\{(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})\}_{n>0}$ satisfying Assumption D1 that converges to the continuous learning model (L_1, L_2). The lengths of time $\Delta_{n,i} \geq \Delta'_{n,i}$ of the associated learning periods are either equal or distinct. In the former case, $\Delta_{n,i} = \Delta'_{n,i}$, the DM has no way of distinguishing between the possibilities $j=1$ and $j=2$ within that learning period; this can only be done in the latter case, $\Delta_{n,i} > \Delta'_{n,i}$. Thus, the following assumption, a revision of Assumption D1, endogenizes the part of Assumption C2 that constrains quitting to only occur after passing a fixed level of knowledge $\beta > 0$.

Assumption D1* For all i such that $b_{n,i-1} < \beta$, we have $\Delta_{n,i} = \Delta'_{n,i}$. For all other i, we have $\Delta_{n,i} > \Delta'_{n,i}$.

This assumption is consistent with the finding of Sanchez and Dunning[23] that each progressive trial in their experiment decreased in length of time. We hypothesize that this decrease occurs faster for imitation learning than it does for innovation learning, so that while the lengths of time $\Delta_{n,i}$ and $\Delta'_{n,i}$ are indistinguishable in the beginning, they decrease and eventually branch off from one another as $i \to \infty$ (see Fig. S2). As a result, the DM can keep track of the amount of time it takes her to traverse the ith learning period, and depending on whether that amount is $\Delta_{n,i}$ or $\Delta'_{n,i}$, she can conclude that $j = 1$ or $j = 2$, respectively.

If the DM incurs no cost in performing this time-measurement experiment, then she would do so at the earliest possible time: the ith learning period for the smallest i such that $\Delta_{n,i} > \Delta'_{n,i}$. This is consistent with our story if $b^* = \beta$: in other words, if the DM quits at the earliest possible quitting point. However, in the more feasible case that $b^* > \beta$, the DM would costlessly identify j at the earliest possible level of knowledge, β, so that the resulting change in the confidence estimate from $g_u(b)$ to $g_1(b)$ or $g_2(b)$ would increase the additional payoffs from side opportunities. Then, the remaining part of Assumption C2—which states that after any identification of j, the DM immediately quits if j turns out to be 1—would not hold true. Thus, we need an additional assumption like the following to endogenize this condition.

Assumption D3 The undiscounted expected cost of a time-measurement experiment $C_{n,i}$ during the ith learning period is uniformly $o(1)$ as $n \to \infty$. However, it is greater than the expected additional payoff the DM would obtain from side opportunities by improving her confidence estimate from $g_u(b)$ to $g_1(b)$ or $g_2(b)$, over the remainder of the learning game (accounting for time discounting).

This $o(1)$ cost can be explained by the increased risk of evolutionary threats—such as ambushes by other humans—resulting from overcommitting attention to the time-measurement experiment. The DM quits if $j = 1$, and otherwise, she changes priors to $j = 2$ and continues learning the current task. The time-measurement experiment provides a $\Theta(1)$ benefit for an $o(1)$ cost, so the DM must perform this experiment before reaching the level of knowledge $b_{n,i}$ at which she quits tasks of the first type. Assumption D3 guarantees that the DM does so at the latest possible moment: during the learning period that precedes the jump to $b_{n,i}$. Indeed, we have assumed the $o(1)$ cost of the time-measurement experiment to always be greater than the $o(1)$ expected increase in additional payoffs from identifying j: information that allows the

DM to more optimally take side-opportunity payoffs in lieu of default payoffs, and thus would be taken as early as possible if it were free.

In the absence of infinite-difficulty tasks, the introduction of side opportunities and time-measurement costs is not necessary to prove for discrete learning models the monotonicity of the expected payoff function. Indeed, the proof of Proposition 1 (provided in Subsection 4.1) also demonstrates that a discrete learning model with only tasks of the second type has both a monotonically increasing expected marginal payoff function and the property that quitting is always suboptimal. Specifically, the proof shows that the DM would always prefer being at a positive level of knowledge b (that has not yet caught up to a) for a task of the second type, to being at zero knowledge for a task of the second type, which is preferred to being at zero knowledge for a task of the first type. Thus, any optimal quitting strategy of a discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$ must be to only quit in the following situation: when the level of knowledge b reaches some $b_{n,i-1}$—which is allowed to be infinite—before catching up to a, perform a time-measurement experiment; if the type is $j = 2$, never quit; and if the type is $j = 1$, quit at level of knowledge $b_{n,i}$—the start of the ith learning period—if learning has not completed by then. Denote this by the *discrete quitting strategy* $b_{n,i}$.

Define the *discrete quitting value function* $V_{u,n} : \{b_{n,i} : \beta \leq b_{n,i}\} \to \mathbb{R}_{\geq 0}$ by the expected total payoff from employing the discrete quitting strategy $b_{n,i}$. Recall that we have denoted its counterpart for the continuous learning model (L_1, L_2) by V_u; and the expected payoff functions (conditional on b not having caught up to a), g_1, g_2, and g_u. The latter are also the expected marginal payoff functions of the discrete learning models $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$.

To recap our hypothesis, side opportunities incentivized ancestral humans to estimate the expected marginal payoff from their primary foraging method as a function of their level of knowledge on it. Furthermore, when the speeds of innovation learning and imitation learning branch off, a time-measurement experiment can differentiate between the two types of tasks. However, the evolutionary cost of performing this experiment made it suboptimal to do so until the last possible moment: the intermediate learning period after which tasks of innovation learning should be quit. Due to the negligibility of these factors, the discrete learning model is well-approximated by a continuous learning model.

Proposition 8. *Suppose we have a sequence of discrete learning models* $\{(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})\}_{n>0}$*— all having task distribution* μ *of the form in Proposition 5; satisfying Assumptions D1*, D2, and D3; and sharing payoff functions* $f_a(\cdot)$ *satisfying Assumption C3—that converges to the continuous learning model* (L_1, L_2).

a) *Given a side opportunity of marginal value P, the DM procures it if P is greater than*

$$\begin{cases} 1 & \text{if her level of knowledge } b_{n,i} \text{ has caught up to } a, \\ g_u(b_{n,i}) & \text{if } b_{n,i} \text{ has not caught up to } a, \text{ and the task type } j \text{ is unknown,} \\ g_2(b_{n,i}) & \text{if } b_{n,i} \text{ has not caught up to } a, \text{ and the task type is known to be } j = 2; \end{cases}$$

leaves it if the reverse inequality holds; and is indifferent between the two options if equality holds.

b) *For every $\varepsilon > 0$, the discrete quitting value function $V_{u,n}$ is ε-close in sup norm to V_u for all sufficiently large n. In particular, if (L_1, L_2) satisfies Assumption C4, then for all sufficiently large n, the one or more optimal discrete quitting strategies $b^*_{n,i}$ maximizing $V_{u,n}(\cdot)$ are finite.*

We hypothesize that human learners, when informationally lacking a Bayesian estimate of future payoffs, default to an innate estimate

$$b_{n,i} \mapsto \begin{cases} g_u(b_{n,i}) & \text{if } b_{n,i} < b^*_{n,i}, \\ g_2(b_{n,i}) & \text{if } b_{n,i} \geq b^*_{n,i}, \end{cases} \tag{3}$$

associated to an optimal discrete quitting strategy $b^*_{n,i}$ of the evolutionary past's learning environment $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$. Recall from Proposition 5 that g_2 is monotonically increasing with respect to the level of knowledge $b_{n,i}$, while g_u eventually decays to zero; in fact, the latter is eventually decreasing under Assumption C5. The aforementioned piecewise-defined function (3) can thus exhibit the desired non-monotonicity—initially increasing, then decreasing, and finally reverting to increase—depending on the choice of model parameters.

3.1 Directions for generalization

Assumption D3, which endogenously justifies the part of Assumption C2 pertaining to quitting, guarantees that the DM performs the time-measurement experiment at the latest possible moment before the optimal quitting point $b^*_{n,i}$: during the learning period directly preceding $b^*_{n,i}$. However, while this is sufficient for our main hypothesis—that the piecewise definition of the expected marginal payoff can explain its non-monotonicity—it is not necessary. In general, the evolutionary cost of performing the time-measurement experiment (due to risks from over-comitting attention) and the expected benefit it would provide (due to better exploitation of side opportunities) can be such that the optimal moment to perform the time-measurement experiment is any learning period between β and $b^*_{n,i}$. If this learning period, say the Ith one, is not the latest learning period between β and $b^*_{n,i}$, and the time-measurement experiment reveals that the task type is $j = 1$, then the DM replaces her confidence estimate g_u with g_1 in the remaining learning periods before quitting at level of knowledge $b^*_{n,i}$. The DM's confidence estimate at level of knowledge $b_{n,i}$ (that has not yet caught up to a) and task type j is thus given by

$$(b_{n,i}, j) \mapsto \begin{cases} g_u(b_{n,i}) & \text{if } b_{n,i} < b_{n,I}, \\ g_1(b_{n,i}) & \text{if } b_{n,I} \leq b_{n,i} < b^*_{n,i} \text{ and } j = 1, \\ g_2(b_{n,i}) & \text{if } b_{n,I} \leq b_{n,i} \text{ and } j = 2, \end{cases} \tag{4}$$

which can still be non-monotonic with respect to the level of knowledge in the desired way: general increase except for an intermediate period of decrease. The empirical data—humans' confidence estimates and quitting behavior when they learn tasks without meaningfully retaining environmental feedback—may be consistent with (3), or with its more general form (4).

Also, the DM's optimal estimate of her future payoff—her confidence—can in general vary with additional contextual information (in the complement of high-variance payoff observations) that is available to her, not just the one-dimensional level of experience. Our learning model, when suitably generalized in this way, can be applied to myriad settings of knowledge-based learning in the absence of meaningfully retained payoff observations. For example, it can model the decision-making of basic-science researchers who are tasked with choosing between research programs to work on, just as it can model the decision-making of ancestral humans who chose between foraging tasks learned from fellow group members. While the latter is the focus of this paper, there is a diverse array of potential applications for our novel decision-theoretic model and its generalizations.

4 Proofs of results

We provide proofs of the aforementioned results.

4.1 Proof of Proposition 1

The conditional probability distributions $\mu_2|_{a>b}$ are naturally isomorphic for all b; all are exponential distributions of the same decay factor η. Assumption C3 implies that for a DM whose learning has not completed yet, the conditional distribution of payoffs at level of knowledge $\underline{b}$ is first-order stochastic dominated by that at level of knowledge $\overline{b} > \underline{b}$, which proves part (b). It follows that when b has not yet reached a, the DM would always prefer to be at a greater level of knowledge b, which proves part (a).

4.2 Proof of Proposition 2

To prove part (a), it suffices to show that the derivative of

$$\begin{aligned} V(b) = & \frac{1}{1 - \int_{a>b} \delta^{L_1^{-1}(b)} d\mu(a)} \\ & \cdot \left(\int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t)) dt + \int_{L_1^{-1}(a)}^{\infty} \delta^t dt \right) d\mu(a) \right. \\ & \left. + \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t)) dt \right) d\mu(a) \right), \end{aligned}$$

which is given by

$$\begin{aligned} V'(b) = & \frac{1}{\left(1 - \int_{a>b} \delta^{L_1^{-1}(b)} d\mu(a)\right)^2} \\ & \cdot \left(\left(1 - \int_{a>b} \delta^{L_1^{-1}(b)} d\mu(a)\right) \left(\mu(b) \int_{L_1^{-1}(b)}^{\infty} \delta^t dt + \int_{a>b} \frac{\delta^{L_1^{-1}(b)} f_a(b)}{L_1'(L_1^{-1}(b))} d\mu(a) \right) \right. \\ & - \left(\mu(b) \delta^{L_1^{-1}(b)} + \int_{a>b} \frac{\delta^{L_1^{-1}(b)} \log \frac{1}{\delta}}{L_1'(L_1^{-1}(b))} d\mu(a) \right) \\ & \cdot \left(\int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t)) dt + \int_{L_1^{-1}(a)}^{\infty} \delta^t dt \right) d\mu(a) \right. \\ & \left. \left. + \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t)) dt \right) d\mu(a) \right) \right), \end{aligned}$$

is negative for all sufficiently large b (excluding any isolated points b at which $L_1'(L_1^{-1}(b)) = 0$, where the derivative may be undefined). Indeed, the distribution of

$$\int_{a>b} d\mu(a)$$

becomes dominated by the point mass on the event $a = \infty$ as $b \to \infty$. Using Assumption C4 to bound the error term, it follows that

$$V'(b) < -\kappa \frac{\delta^{L_1^{-1}(b)}}{L_1'(L_1^{-1}(b))}$$

for all sufficiently large b, where $\kappa > 0$ is some constant.

Part (b) follows from the fact that the conditional distribution $\mu_1|_{a>b}$ is dominated by the event $a = \infty$ for large b.

To prove part (c), we need to show that the derivative

$$\begin{aligned} g'(b) =& \frac{d}{db} \frac{(1-p)\left(\log \frac{1}{\eta}\right) \int_b^\infty f_a(b)\eta^a da}{p + (1-p)\left(\log \frac{1}{\eta}\right) \int_b^\infty \eta^a da} \\ =& \frac{1}{\left(p + (1-p)\left(\log \frac{1}{\eta}\right) \int_b^\infty \eta^a da\right)^2} \\ & \cdot \left(\left(p + (1-p)\left(\log \frac{1}{\eta}\right) \int_b^\infty \eta^a da\right) \left((1-p)\left(\log \frac{1}{\eta}\right) \left(-\eta^b + \int_b^\infty f_a'(b)\eta^a da\right)\right) \right. \\ & \left. + (1-p)\left(\log \frac{1}{\eta}\right) \eta^b \left((1-p)\left(\log \frac{1}{\eta}\right) \int_b^\infty f_a(b)\eta^a da\right) \right) \end{aligned}$$

is negative for all sufficiently large b. Using Assumption C5 to bound the error term, we find that

$$g'(b) < -\kappa\eta^b$$

for all sufficiently large b, where $\kappa > 0$ is some constant.

4.3 Proof of Corollary 3

Let μ_p be the distribution that has probability mass p on $a = \infty$ and probability mass $1 - p$ distributed exponentially on $\mathbb{R}_{>0}$ with a fixed decay factor $\eta \in (0, 1)$. Let $V_{\mu_p}(\cdot)$ denote the quitting value function of Proposition 2's game when the task distribution is μ_p. By applying Lemma 4, we will show a stronger statement than that of Corollary 3; we will show that for every decreasing sequence of probability values $\{p_n\}_{n\in\mathbb{Z}_{>0}}$ converging to zero, there exists N such that any optimal quitting point b^* of $V_{\mu_{p_n}}$ is greater than γ for all $n \geq N$.

By Lemma 4, the continuous functions $\{V_{\mu_{p_n}}\}_{n\in\mathbb{Z}_{>0}}$ monotonically converge (increasing with respect to n) to V_{μ_0}, which is also continuous. An application of Dini's theorem thus shows that the convergence of $V_{\mu_{p_n}}$ to V_{μ_0} on the compactified space $[\beta, \infty) \cup \{\infty\}$ is uniform. Let

$$\varepsilon = V_{\mu_0}(\infty) - \max_{b\in[\beta,\gamma]} V_{\mu_0}(b),$$

which is positive by Proposition 1. By uniform convergence, there exists N such that for all $n \geq N$, we simultaneously have

$$|V_{\mu_{p_n}}(\infty) - V_{\mu_0}(\infty)| < \frac{\varepsilon}{2}$$

and

$$|V_{\mu_{p_n}}(b) - V_{\mu_0}(b)| < \frac{\varepsilon}{2}$$

for all $b \in [\beta, \gamma]$. Since

$$V_{\mu_0}(b) < V_{\mu_0}(\infty)$$

for every $b \in [\beta, \gamma]$ with a difference of at least ε, it follows from the triangle inequality that for any $n \geq N$, we have

$$V_{\mu_{p_n}}(b) < V_{\mu_{p_n}}(\infty)$$

for all $b \in [\beta, \gamma]$. In particular, no $b \in [\beta, \gamma]$ maximizes $V_{\mu_{p_n}}$.

4.4 Proof of Lemma 4

Recall that

$$V_{\mu_p}(b) = \frac{1}{1 - \delta^{L_1^{-1}(b)}\left(p + (1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)}$$

$$
\cdot (1-p)\left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt+\int_{L_1^{-1}(a)}^{\infty}\delta^t dt\right)\eta^a da \right.
$$
$$
\left. +\int_b^{\infty}\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\right)\Bigg). \qquad (5)
$$

Its partial derivative with respect to p is

$$
\frac{\partial}{\partial p}V_{\mu_p}(b) = \frac{1}{\left(1-\delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^{\infty}\eta^a da\right)\right)^2}
$$
$$
\cdot\Bigg(\left(1-\delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^{\infty}\eta^a da\right)\right)
$$
$$
\cdot\Bigg(-\left(\log\frac{1}{\eta}\right)\Bigg(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt+\int_{L_1^{-1}(a)}^{\infty}\delta^t dt\right)\eta^a da
$$
$$
+\int_b^{\infty}\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\Bigg)\Bigg)
$$
$$
+\delta^{L_1^{-1}(b)}\left(1-\left(\log\frac{1}{\eta}\right)\int_b^{\infty}\eta^a da\right)
$$
$$
\cdot (1-p)\left(\log\frac{1}{\eta}\right)\Bigg(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt+\int_{L_1^{-1}(a)}^{\infty}\delta^t dt\right)\eta^a da
$$
$$
+\int_b^{\infty}\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\Bigg)\Bigg),
$$

which is equal to a positive term multiplied by

$$
\delta^{L_1^{-1}(b)}\left(1-p-(1-p)\left(\log\frac{1}{\eta}\right)\int_b^{\infty}\eta^a da\right)
$$
$$
-\left(1-\delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^{\infty}\eta^a da\right)\right) = -1+\delta^{L_1^{-1}(b)} < 0.
$$

Thus, $V_{\mu_p}(b)$ is strictly decreasing with respect to p.

4.5 Proof of Proposition 5

The first part of (a) follows from the argument in the proof of Proposition 1. To prove the second part, it suffices to show that the derivative

$$V_u'(b) = (1-q)V_1'(b)$$

is negative for all sufficiently large b (excluding any isolated points b at which $L_1'(L_1^{-1}(b)) = 0$, where the derivative $V_1'(b)$ may be undefined). The conditional quitting value function $V_1(b)$ is the solution to

$$\begin{aligned} V = &\int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_1(a) \\ &+ \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t))dt + \delta^{L_1^{-1}(b)} \left((1-q)V + qV_2(\infty)\right) \right) d\mu_1(a), \end{aligned}$$

where

$$V_2(\infty) = \int_0^{\infty} \left(\int_0^{L_2^{-1}(a)} \delta^t f_a(L_2(t))dt + \int_{L_2^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_2(a).$$

Thus, we have

$$\begin{aligned} V_1(b) = &\frac{1}{1-(1-q)\int_{a>b} \delta^{L_1^{-1}(b)} d\mu_1(a)} \\ &\cdot \left(\int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_1(a) \right. \\ &\left. + \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t))dt + \delta^{L_1^{-1}(b)} qV_2(\infty) \right) d\mu_1(a) \right), \end{aligned}$$

whose derivative is given by

$$\begin{aligned} V_1'(b) = &\frac{1}{\left(1-(1-q)\int_{a>b} \delta^{L_1^{-1}(b)} d\mu_1(a)\right)^2} \\ &\cdot \left(\left(1-(1-q)\int_{a>b} \delta^{L_1^{-1}(b)} d\mu_1(a) \right) \right. \end{aligned}$$

$$
\cdot \left(\mu_1(b) \int_{L_1^{-1}(b)}^{\infty} \delta^t dt + \int_{a>b} \frac{\delta^{L_1^{-1}(b)} f_a(b)}{L_1'(L_1^{-1}(b))} d\mu_1(a) \right.
$$

$$
\left. - qV_2(\infty) \left(\mu_1(b)\delta^{L_1^{-1}(b)} + \int_{a>b} \frac{\delta^{L_1^{-1}(b)} \log \frac{1}{\delta}}{L_1'(L_1^{-1}(b))} d\mu_1(a) \right) \right)
$$

$$
- (1-q) \left(\mu_1(b)\delta^{L_1^{-1}(b)} + \int_{a>b} \frac{\delta^{L_1^{-1}(b)} \log \frac{1}{\delta}}{L_1'(L_1^{-1}(b))} d\mu_1(a) \right)
$$

$$
\cdot \left(\int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t)) dt + \int_{L_1^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_1(a) \right.
$$

$$
\left. + \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t)) dt + \delta^{L_1^{-1}(b)} qV_2(\infty) \right) d\mu_1(a) \right) \Bigg).
$$

Just as in the proof of Proposition 2(a), we can use Assumption C4 to show that the above expression is

$$
< -\kappa \frac{\delta^{L_1^{-1}(b)}}{L_1'(L_1^{-1}(b))}
$$

for all sufficiently large b, where $\kappa > 0$ is some constant.

The first part of (b) follows from the argument in the proof of Proposition 1, while the second part follows from the fact that the conditional distribution $\mu|_{a>b}$ is dominated by the event $(j, a) = (1, \infty)$ for large b.

To prove part (c), we show that the derivative

$$
g_u'(b) = \frac{d}{db} \frac{(1-(1-q)p) \left(\log \frac{1}{\eta} \right) \int_b^{\infty} f_a(b) \eta^a da}{(1-q)p + (1-(1-q)p) \left(\log \frac{1}{\eta} \right) \int_b^{\infty} \eta^a da}
$$

is negative for all sufficiently large b, under Assumption C5. This has already been done in the proof of Proposition 2(c).

4.6 Proof of Corollary 6

Let $V_{p,q}(b)$ denote the unconditional quitting value function of Proposition 5's game for which tasks of the first type have difficulty $a = \infty$ with probability p and tasks of the second type occur with probability q. By applying Lemma 7, we will show a stronger statement than that of Corollary 6; we will show that for any pair of decreasing sequences $\{p_n\}_{n \in \mathbb{Z}_{>0}}$ and

$\{q_m\}_{m\in\mathbb{Z}_{>0}}$ converging to zero, there exists N such that for any $n \geq N$, we can find M_n such that any optimal quitting strategy b^* of V_{p_n,q_m} is greater than γ for all $m \geq M_n$.

The proof of Corollary 3 shows that there exists N such that for any $n \geq N$, we have

$$V_{p_n,0}(b) < V_{p_n,0}(\infty)$$

for all $b \in [\beta, \gamma]$. Fix $n \geq N$. By Lemma 7, the continuous functions $\{V_{p_n,q_m}\}_{m\in\mathbb{Z}_{>0}}$ monotonically converge (decreasing with respect to m) to $V_{p_n,0}$, which is also continuous. It thus follows from Dini's theorem that the convergence of V_{p_n,q_m} to $V_{p_n,0}$ on the compactified space $[\beta, \infty) \cup \{\infty\}$ is uniform. Let

$$\varepsilon = V_{p_n,0}(\infty) - \max_{b\in[\beta,\gamma]} V_{p_n,0}(b),$$

which is positive by our choice of n. By uniform convergence, there exists M_n such that for all $m \geq M_n$, we simultaneously have

$$|V_{p_n,q_m}(\infty) - V_{p_n,0}(\infty)| < \frac{\varepsilon}{2}$$

and

$$|V_{p_n,q_m}(b) - V_{p_n,0}(b)| < \frac{\varepsilon}{2}$$

for all $b \in [\beta, \gamma]$. Since

$$V_{p_n,0}(b) < V_{p_n,0}(\infty)$$

for all $b \in [\beta, \gamma]$ with a difference of at least ε, it follows from the triangle inequality that for any $m \geq M_n$, we have

$$V_{p_n,q_m}(b) < V_{p_n,q_m}(\infty)$$

for all $b \in [\beta, \gamma]$. In particular, no $b \in [\beta, \gamma]$ maximizes V_{p_n,q_m}.

4.7 Proof of Lemma 7

We first show that $V_u(b)$ is strictly decreasing with respect to p. This has been done in Lemma 4 for the special case that $q = 0$; we now show the general statement. Recall that

$$V_u(b) = (1 - q)V_1(b) + qV_2(\infty), \tag{6}$$

from which we obtain

$$\frac{\partial}{\partial p}V_u(b) = (1-q)\frac{\partial}{\partial p}V_1(b).$$

To demonstrate that $\frac{\partial}{\partial p}V_u(b) < 0$, we need to show that the partial derivative of

$$\begin{aligned} V_1(b) =& \frac{1}{1-\delta^{L_1^{-1}(b)}(1-q)\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)} \\ &\cdot\left(\delta^{L_1^{-1}(b)}q\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)V_2(\infty)\right. \\ &\quad + (1-p)\left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt\right)\eta^a da\right. \\ &\quad\quad \left.\left. + \int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\right)\right) \end{aligned} \tag{7}$$

with respect to p,

$$\begin{aligned} \frac{\partial}{\partial p}V_1(b) =& \frac{1}{\left(1-\delta^{L_1^{-1}(b)}(1-q)\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)\right)^2} \\ &\cdot\left(\left(1-\delta^{L_1^{-1}(b)}(1-q)\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)\right)\right. \\ &\quad\cdot\left(\delta^{L_1^{-1}(b)}q\left(1-\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)V_2(\infty)\right. \\ &\quad\quad - \left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt\right)\eta^a da\right. \\ &\quad\quad\quad \left.\left. + \int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\right)\right) \\ &\quad + \delta^{L_1^{-1}(b)}(1-q)\left(1-\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right) \\ &\quad\cdot\left(\delta^{L_1^{-1}(b)}q\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)V_2(\infty)\right. \\ &\quad\quad + (1-p)\left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt\right)\eta^a da\right. \end{aligned}$$

$$+\int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\Bigg)\Bigg),$$

is negative. Observe that this is a positive term multiplied by

$$\begin{aligned}
&\delta^{L_1^{-1}(b)}q\left(1-\left(\log\frac{1}{\eta}\right)\int_b^\infty\eta^a da\right)V_2(\infty)\\
&\quad-\left(1-\delta^{L_1^{-1}(b)}(1-q)\right)\\
&\qquad\cdot\left(\log\frac{1}{\eta}\right)\Bigg(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt+\int_{L_1^{-1}(a)}^\infty\delta^t dt\right)\eta^a da\\
&\qquad\qquad+\int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\Bigg)\\
&=-\left(\log\frac{1}{\eta}\right)\Bigg(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt+\int_{L_1^{-1}(a)}^{L_1^{-1}(b)}\delta^t dt\right)\eta^a da\\
&\qquad+\int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\Bigg)\\
&\quad-\left(\log\frac{1}{\eta}\right)\int_0^b\left(\int_{L_1^{-1}(b)}^\infty\delta^t dt\right)\eta^a da\\
&\quad+\delta^{L_1^{-1}(b)}(1-q)\left(\left(\log\frac{1}{\eta}\right)\int_b^\infty\eta^a da\right)\\
&\qquad\cdot\left(\log\frac{1}{\eta}\right)\Bigg(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt+\int_{L_1^{-1}(a)}^{L_1^{-1}(b)}\delta^t dt\right)\eta^a da\\
&\qquad\qquad+\int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\Bigg)\\
&\quad+\delta^{L_1^{-1}(b)}(1-q)\left(\left(\log\frac{1}{\eta}\right)\int_b^\infty\eta^a da\right)\left(\log\frac{1}{\eta}\right)\int_0^b\left(\int_{L_1^{-1}(b)}^\infty\delta^t dt\right)\eta^a da\\
&\quad+\delta^{L_1^{-1}(b)}(1-q)\left(1-\left(\log\frac{1}{\eta}\right)\int_b^\infty\eta^a da\right)\\
&\qquad\cdot\left(\log\frac{1}{\eta}\right)\Bigg(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt+\int_{L_1^{-1}(a)}^\infty\delta^t dt\right)\eta^a da\\
&\qquad\qquad+\int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\Bigg)
\end{aligned}$$

$$+\,\delta^{L_1^{-1}(b)} q \left(1 - \left(\log \frac{1}{\eta}\right) \int_b^\infty \eta^a da\right) V_2(\infty).$$

The above expression is negative, since it is the sum of the negative term

$$-\left(1 - \delta^{L_1^{-1}(b)}(1-q)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right) \cdot \left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^{L_1^{-1}(b)} \delta^t dt\right)\eta^a da + \int_b^\infty\left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t))dt\right)\eta^a da\right)$$

and the negative term

$$\begin{aligned}
&-\delta^{L_1^{-1}(b)}\left(\log\frac{1}{\eta}\right)\int_0^b\left(\int_0^\infty \delta^t dt\right)\eta^a da \\
&\quad + \delta^{L_1^{-1}(b)}\left(1 - \left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)(1-q) \\
&\qquad \cdot \left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt\right)\eta^a da\right. \\
&\qquad\qquad \left. + \int_b^\infty\left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(b)}^\infty \delta^t dt\right)\eta^a da\right) \\
&\quad + \delta^{L_1^{-1}(b)}\left(1 - \left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right) qV_2(\infty);
\end{aligned}$$

the negativity of the latter term follows from the fact that

$$\int_0^\infty \delta^t dt$$

is the maximum possible total payoff, and in particular is greater than

$$qV_2(\infty) + (1-q)\left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt\right)\eta^a da + \int_b^\infty\left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(b)}^\infty \delta^t dt\right)\eta^a da\right).$$

This overall shows that $V_1(b)$, and thus, $V_u(b)$, is strictly decreasing with respect to p.

Next, we show that $V_u(b)$ is strictly increasing with respect to q. From (6), we obtain

$$\frac{\partial}{\partial q}V_u(b) = V_2(\infty) - V_1(b) + (1-q)\frac{\partial}{\partial q}V_1(b).$$

Since $V_2(\infty) > V_1(b)$, it suffices to show that $\frac{\partial}{\partial q}V_1(b) > 0$. We take the partial derivative of (7) with respect to q, which yields

$$\begin{aligned}\frac{\partial}{\partial q}V_1(b) =& \frac{1}{\left(1-\delta^{L_1^{-1}(b)}(1-q)\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)\right)^2} \\ &\cdot\left(\left(1-\delta^{L_1^{-1}(b)}(1-q)\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)\right)\right. \\ &\cdot \delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)V_2(\infty) \\ &- \delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right) \\ &\cdot\left(\delta^{L_1^{-1}(b)}q\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)V_2(\infty)\right. \\ &+ (1-p)\left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt\right)\eta^a da\right. \\ &\left.\left.\left.+\int_b^\infty\left(\int_0^{L_1^{-1}(b)}\delta^t f_a(L_1(t))dt\right)\eta^a da\right)\right)\right).\end{aligned}$$

As desired, this is a positive term multiplied by

$$\begin{aligned}&\left(\delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)\right. \\ &\left.-\left(\delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right)\right)^2\right)V_2(\infty) \\ &- \delta^{L_1^{-1}(b)}\left(p+(1-p)\left(\log\frac{1}{\eta}\right)\int_b^\infty \eta^a da\right) \\ &\cdot (1-p)\left(\log\frac{1}{\eta}\right)\left(\int_0^b\left(\int_0^{L_1^{-1}(a)}\delta^t f_a(L_1(t))dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt\right)\eta^a da\right.\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad + \int_b^\infty \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_1(t))dt \right) \eta^a da \Bigg) \Bigg) \\
&= \left(\delta^{L_1^{-1}(b)} \left(p + (1-p) \left(\log \frac{1}{\eta} \right) \int_b^\infty \eta^a da \right) \right. \\
&\qquad \left. - \left(\delta^{L_1^{-1}(b)} \left(p + (1-p) \left(\log \frac{1}{\eta} \right) \int_b^\infty \eta^a da \right) \right)^2 \right) V_2(\infty) \\
&\quad - \delta^{L_1^{-1}(b)} \left(p + (1-p) \left(\log \frac{1}{\eta} \right) \int_b^\infty \eta^a da \right) \\
&\qquad \cdot \left(1 - \delta^{L_1^{-1}(b)} \left(p + (1-p) \left(\log \frac{1}{\eta} \right) \int_b^\infty \eta^a da \right) \right) V_{\mu_p}(b) \\
&= \delta^{L_1^{-1}(b)} \left(p + (1-p) \left(\log \frac{1}{\eta} \right) \int_b^\infty \eta^a da \right) \\
&\qquad \cdot \left(1 - \delta^{L_1^{-1}(b)} \left(p + (1-p) \left(\log \frac{1}{\eta} \right) \int_b^\infty \eta^a da \right) \right) \big(V_2(\infty) - V_{\mu_p}(b) \big) > 0,
\end{aligned}
$$

where we have used the formula (5) for $V_{\mu_p}(b)$ and the fact that it is less than $V_2(\infty)$. We have shown that $V_1(b)$, and thus, $V_u(b)$, is strictly increasing with respect to q.

4.8 Proof of Proposition 8

Only part (b) requires proof. Choose N large enough that the expected payoff deviation due to the procurement of side opportunities in the discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$ is less than $\varepsilon/3$ for all $n \geq N$, which is possible due to Assumption D2. By possibly making N larger, the expected payoff deviation due to time-measurement experiments in the discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$ is less than $\varepsilon/3$ for all $n \geq N$, where we have used the hypothesis of uniform $o(1)$ decay in Assumption D3 and the quitting constraint $b \geq \beta$ implied by Assumption D1*.

Furthermore, by possibly making N even larger, the difference between the expected total payoff $V_{u,n}$ of the discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$—henceforward excluding deviations due to side opportunities and time-measurement costs—and that of its approximating continuous learning model (L_1, L_2), given by V_u, is less than $\varepsilon/3$ for all $n \geq N$. To show this, we may as well assume that the payoff of each learning period $[T + t_{n,i-1}, T + t_{n,i}]$ of the discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$, given by

$$f_a\left(\min(b_{n,i-1}, a)\right) \int_{T+t_{n,i-1}}^{T+t_{n,i}} \delta^t dt,$$

is obtained as a flow payoff of $\delta^t f_a\left(\min(b_{n,i-1}, a)\right) dt$. Using this equivalence, we can "connect the dots" of $V_{u,n}$ in the following sense. Define the unconditional quitting value function $\tilde{V}_{u,n}(b)$ of this flow-payoff game analogously to $V_u(b)$ of the continuous learning model's flow-payoff game, by the expected total payoff of quitting tasks of the first type at time $L_1^{-1}(b)$ unless learning has completed by then. In other words,

$$\tilde{V}_{u,n}(b) = (1-q)V_{1,n}(b) + qV_{2,n}(\infty),$$

where

$$V_{2,n}(\infty) = \int_0^\infty \left(\int_0^{L_2^{-1}(a)} \delta^t f_a(L_{\mathbb{S}'_n}(t)) dt + \int_{L_2^{-1}(a)}^\infty \delta^t dt \right) d\mu_2(a)$$

is the expected total payoff conditional on $j = 2$; and $V_{1,n}(b)$, the expected total payoff conditional on $j = 1$, is defined by the solution to

$$\begin{aligned} V = &\int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_{\mathbb{S}_n}(t)) dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt \right) d\mu_1(a) \\ &+ \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_{\mathbb{S}_n}(t)) dt + \delta^{L_1^{-1}(b)} \left((1-q)V + qV_{2,n}(\infty)\right) \right) d\mu_1(a), \end{aligned}$$

which is

$$\begin{aligned} V_{1,n}(b) = &\frac{1}{1-(1-q)\int_{a>b} \delta^{L_1^{-1}(b)} d\mu_1(a)} \\ &\cdot \left(\int_0^b \left(\int_0^{L_1^{-1}(a)} \delta^t f_a(L_{\mathbb{S}_n}(t)) dt + \int_{L_1^{-1}(a)}^\infty \delta^t dt \right) d\mu_1(a) \right. \\ &\left. + \int_{a>b} \left(\int_0^{L_1^{-1}(b)} \delta^t f_a(L_{\mathbb{S}_n}(t)) dt + \delta^{L_1^{-1}(b)} qV_{2,n}(\infty) \right) d\mu_1(a) \right). \end{aligned}$$

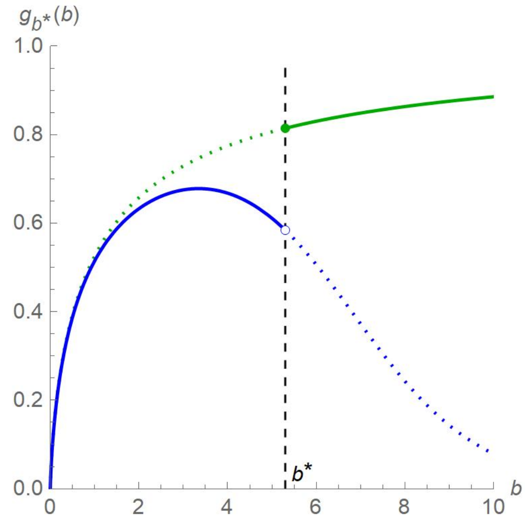
The restriction of $\tilde{V}_{u,n} : [\beta, \infty) \cup \{\infty\} \to \mathbb{R}_{\geq 0}$ to the subdomain $\{b_{n,i} : \beta \leq b_{n,i}\}$ is precisely

$V_{u,n}$ (after excluding the error term due to side opportunities and time-measurement costs). Also, $\{\tilde{V}_{u,n}\}_{n>0}$ is a sequence of continuous functions on the compactified space $[\beta, \infty) \cup \{\infty\}$ monotonically converging to V_u, which is also continuous. Thus, this convergence is uniform by Dini's theorem. In particular, we have

$$\sup_{b_{n,i} \geq \beta} |V_u(b_{n,i}) - V_{u,n}(b_{n,i})| \leq \sup_{b \geq \beta} \left| V_u(b) - \tilde{V}_{u,n}(b) \right| < \frac{\varepsilon}{3}$$

for all sufficiently large n, as desired.

Our overall result then follows from the triangle inequality.



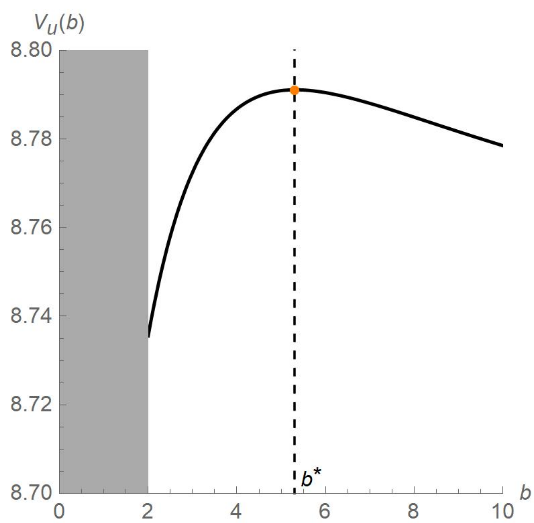


Fig. S1. The continuous learning model (L_1, L_2) whose parameters are given by $L_1(t) = t, L_2(t) = \left\{ \begin{array}{ll} t & \text{for } t < 2 \\ 2t & \text{for } t \geq 2 \end{array} \right\}, f_a(b) = b/a, \delta = 0.9, \eta = 0.5, p = 0.01, q = 0.01$, and $\beta = 2$. When quitting is restricted to $b \geq \beta$ by Assumption C2, the unconditional quitting value function $V_u(b)$ (bottom) has a unique maximum at $b^* \approx 5.32071$. The true expected payoff function $g_{b^*}(b)$ associated to this unique optimal quitting strategy b^* (top) is non-monotonic.

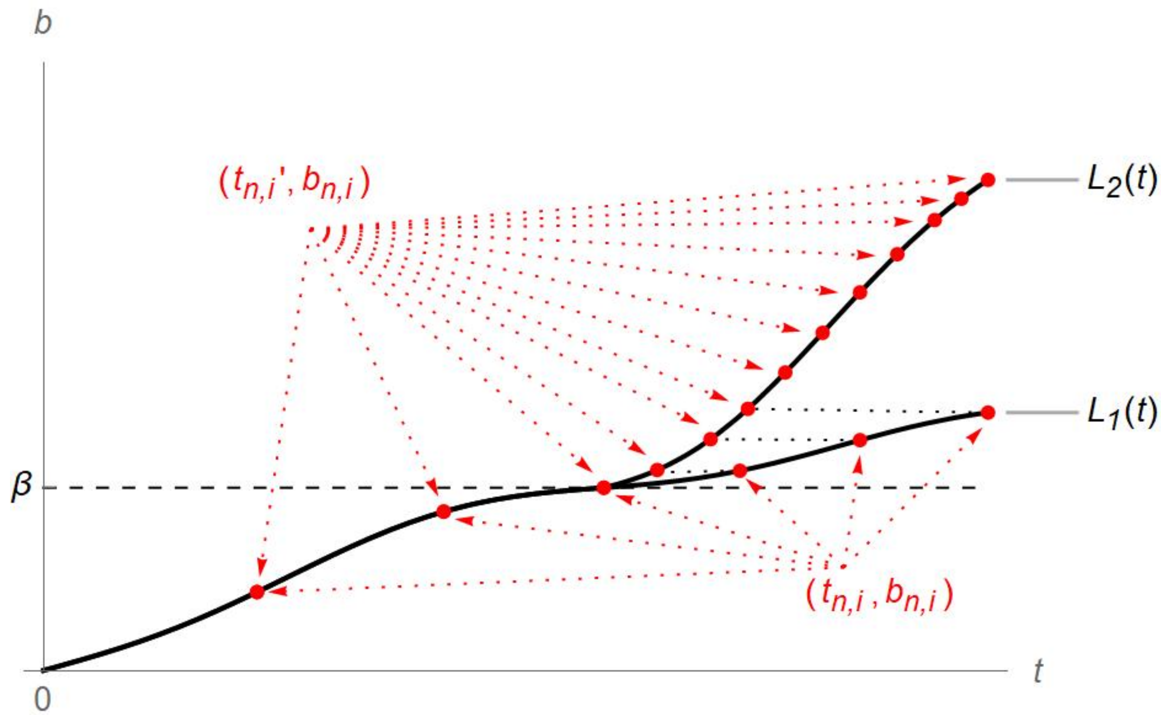


Fig. S2. Generic continuous learning functions $L_1(t)$ and $L_2(t)$ approximating generic discrete learning functions $L_{\mathbb{S}_n}(t)$ and $L_{\mathbb{S}'_n}(t)$, whose knowledge jumps satisfy Assumption D1*. Before the DM's level of knowledge $b_{n,i}$ reaches β, the ith learning period of $L_{\mathbb{S}_n}(t)$ and the corresponding learning period of $L_{\mathbb{S}'_n}(t)$ have equal length. Afterwards, the two branch off; the former becomes larger in length than the latter. Since these corresponding learning periods' knowledge jumps share the same b-value, the DM requires a time-measurement experiment to identify the type j of her current task. This hypothesis is consistent with the finding of Sanchez and Dunning[23] that their experimental trials progressively decreased in length of time.

Table S1. Discrete learning model $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$ as a strategic game against nature.

Information space	The DM knows b, as well as whether $b = b_{n,i} < a$ or $b = a$. She may also have come to possess the information $j = 1$ or $j = 2$, without which she operates under the starting belief, $j \in \{1, 2\}$.
Action space	Full payoffs are guaranteed from the main task if $b = a$. Otherwise, the DM makes two choices before each learning period: first, whether to perform a time-measurement experiment; and second, whether to devote part of the learning period to a side opportunity—whose marginal payoff P, which is observed, is drawn from a distribution with support equal to $[0, 1]$—in lieu of part of the main task's payoff. At the end of the learning period, she obtains the payoff, the result of the time-measurement experiment if any, and the opportunity to quit.
Payoff	We only write the payoff (possibly after having quit previous tasks) of the ith learning period $[T + t_{i-1}, T + t_i]$ when $j = 1$, since that of $j = 2$ is analogous: $$\left(\mathbb{I}R_{n,i}P + (1 - \mathbb{I}R_{n,i})\, f_a\left(\min(b_{n,i-1}, a)\right)\right) \int_{T+t_{n,i-1}}^{T+t_{n,i}} \delta^t dt,$$ where $\mathbb{I}$ is equal to 1 if the DM procures the side opportunity and to 0 otherwise. The cost $-\delta^{T+t_{n,i}} C_{n,i}$ is added if she has performed the time-measurement experiment.
Assumption C3	The payoff functions $f_a(\cdot)$ satisfy $f_{b+m}(b) < f_{b'+m}(b')$ for all $b < b'$ and $m > 0$. Colloquially, a fixed amount of knowledge m constitutes a larger fraction of total knowledge of an easy task than of a difficult task, so ignorance causes a harsher penalty in the former case.
Assumption D1*	For all i such that $b_{n,i-1} < \beta$, we have $\Delta_{n,i} = \Delta'_{n,i}$. For all other i, we have $\Delta_{n,i} > \Delta'_{n,i}$. Colloquially, the speeds of innovation and imitation learning are indistinguishable at first, but eventually branch off.
Assumption D2	As $n \to \infty$, the expected fraction $R_{n,i}$ of the ith learning period's time that can be used for a side opportunity is uniformly $o(1)$. Colloquially, side opportunities comprise nontrivial, but negligible opportunity costs.
Assumption D3	For every n, the undiscounted expected cost of a time-measurement experiment $C_{n,i}$ during the ith learning period is uniformly $o(1)$ as $n \to \infty$. However, it is greater than the expected additional payoff the DM would obtain from side opportunities by improving her confidence estimate from $g_u(b)$ to $g_1(b)$ or $g_2(b)$, over the remainder of the learning game (accounting for time discounting). Colloquially, the only reason to perform the time-measurement experiment is to immediately quit tasks of innovation learning.

Table S2. Continuous learning model (L_1, L_2), a limit of discrete learning models $(L_{\mathbb{S}_n}, L_{\mathbb{S}'_n})$, as an optimal stopping game against nature.

Information space	The DM knows b, as well as whether $b < a$ or $b = a$.
Action space	The DM's sole action is quitting, which can be done at any level of knowledge $b \geq \beta$.
Payoff	The DM obtains a flow payoff of $\delta^t f_a(b)dt$.
Assumption C1	We have $L_1(t) \leq L_2(t)$ for all $t \in \mathbb{R}_{\geq 0}$. Follows from Assumption D1*.
Assumption C2[a]	To identify j, the DM must commit to quitting the task immediately afterwards if she finds out that $j = 1$; this follows from Assumptions D2 and D3. She can only perform this identification when her level of knowledge b is $\geq \beta$; this follows from Assumption D1*.
Assumption C3	See Table S1.
Assumption C4	The learning function $L_1(\cdot)$ is continuously differentiable and satisfies $$L_1'(t) \ll \eta^{-L_1(t)}$$ as $t \to \infty$. Colloquially, the rate of learning is not too high.
Assumption C5	The payoff functions $f_a(\cdot)$ are continuously differentiable and satisfy $$\int_{a>b} f_a'(b)\eta^a da \ll \eta^b$$ as $b \to \infty$. Colloquially, the first derivatives of the payoff functions are well-behaved.

[a]Other than by providing the constraint $b \geq \beta$ on quitting, Assumption C2 does not factor into the currently described game.