Games and Economic Behavior 144 (2024) 29–48

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

Imposing commitment to rein in overconfidence in learning [☆,☆☆]

Marcelo Ariel Fernandez [a,b], Tatiana Mayskaya [c], Arina Nikandrova [d,*]

[a] College of Business, University of Nevada, Reno, 1664 N. Virginia Street, Reno, 89557, NV, United States
[b] Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Wyman Park Building, Baltimore, 21218, MD, United States
[c] International College of Economics and Finance, Faculty of Economic Sciences, HSE University, 11 Pokrovsky Bulvar, Moscow, 109028, Russia
[d] Department of Economics, City, University of London, Northampton Square, London, EC1V 0HB, United Kingdom

ARTICLE INFO

JEL classification:
D83
D91

Keywords:
Learning
Delegation
Overconfidence
Misspecified model
Brownian motion

ABSTRACT

A rational principal delegates learning to an overconfident agent who overestimates the precision of the information he collects. The principal chooses between two contracts: commitment, in which the agent commits to the duration of learning in advance, and flexible, in which the agent decides when to stop learning in real time. When the agent is sufficiently overconfident, the principal optimally ties the agent's hands by offering him the commitment contract. When the principal can choose both the contract and the agent's level of overconfidence, selecting the rational agent is suboptimal when the cost of learning is sufficiently high.

1. Introduction

It is widely documented that individuals often display overconfidence (e.g., Moore et al. (2015)), which affects the amount of information they collect before making decisions. When these decisions are delegated, a principal may want to adjust a contract to the presence of overconfidence or avoid hiring an overconfident agent altogether. In this paper, we focus on the hiring choices of a rational principal and look at the contracts that regulate the timing of a decision delegated to an agent. We show that it may be beneficial for the principal to hire an overconfident agent who overestimates the precision of the information he collects; however, the agent's overconfidence must be reined in by imposing commitment to the decision time ex ante — that is, before the agent starts collecting information that is relevant to the decision.

We consider an environment in which the state of the world is binary and the decision involves guessing the state. A principal, referred to as she, delegates the guess to an agent, referred to as he. The principal and the agent share the same prior and believe

[☆] This paper was previously circulated under the title "Implications of Overconfidence on Information Investment."

[☆☆] We are indebted to Leeat Yariv for her numerous and thoughtful comments on our work. We are also grateful to the Editor, Nicolas Vieille, the Advisory Editor, anonymous referees, as well as Emiliano Catonini, Doruk Cetemen, Mikhail Drugov, Federico Echenique, Michael Grubb, Alexander V. Hirsch, Matias Iaryczower, Botond Kőszegi, Gilat Levy, Alexander Nesterov, Salvatore Nunnari, Romans Pancs, Luciano Pomatto, Andrei Savochkin, Kathryn Spier, Sergey Stepanov, and James Tremewan for insightful comments. We also thank seminar audiences at Caltech, the 3rd Lancaster Game Theory conference, the 18th SAET conference, the 96th WEAI conference, HSE and UL City. The study has been funded by the Russian Academic Excellence Project '5-100'.

* Corresponding author.
E-mail addresses: marceloarielf@unr.edu, fernandez@jhu.edu (M.A. Fernandez), tmayskaya@gmail.com (T. Mayskaya), a.nikandrova@gmail.com (A. Nikandrova).
URLs: https://www.tmayskaya.com (T. Mayskaya), https://sites.google.com/view/anikandrova (A. Nikandrova).

https://doi.org/10.1016/j.geb.2024.01.001
Received 16 June 2022
Available online 10 January 2024
0899-8256/

that both states are equally likely.[1] The agent has exclusive access to a source that gradually provides information about the state. He faces an optimal stopping problem: he decides when to stop listening to the source and, upon stopping, guesses the state. The agent obtains a payoff of one if his guess is correct and zero otherwise. Listening to the source is costly.

The agent is overconfident, thereby he believes that the information that the source provides is more precise than it actually is. He is unaware of his own overconfidence and does not update his beliefs about the precision of the information source in the course of learning.[2] The principal observes the agent's level of overconfidence, which we define as the ratio of the perceived precision to the actual precision of the information source.[3]

Given the agent's level of overconfidence, the principal chooses between two contracts: commitment and flexible. Under both contracts, the principal lets the agent choose the time at which he stops listening to the source and makes the guess. The contracts differ in the timing of the agent's choice. The flexible contract allows the agent to choose when to stop listening and to make the guess on the basis of the revealed information; the commitment contract requires the agent to commit to the exact duration of information collection in advance, before starting to listen to the source.

Under the commitment contract, the agent does not stop listening to the source before the time he initially chose. To implement the commitment contract, the agent expends the resources required for learning from the start. Then, the subsequent marginal cost of learning is zero and, thus, the agent never finds it optimal to terminate learning before the agreed time, irrespective of whether the principal can monitor his effort. Contracts where costs of learning are paid in advance are common in practice. For example, the agent may employ the subjects to participate in the experiment of fixed duration, or he pays a fee for a fixed-term subscription to a data stream, such as Bloomberg terminal, or he declines to participate in other projects thus paying an opportunity cost of working on the current project.

The principal cares about whether the guess is correct but not about the timing of the guess, and so, she chooses the contract that maximizes the probability that the agent's guess is correct. There are no transfers between the agent and the principal.

Under the commitment contract, overconfidence has two countervailing effects on the duration of the agent's learning and, hence, on the probability of making the correct guess. On the one hand, a higher level of overconfidence makes learning a little longer appear cheaper, thus making the agent more willing to collect information at the margin. We call this a cost effect, and it increases the probability of making the correct guess. On the other hand, a more overconfident agent overestimates the precision of the information that he collects in any given interval of time, and, thus, is less willing to collect additional information. We call this an oversaturation effect, and it decreases the probability of making the correct guess.

Under the flexible contract, in addition to the cost and oversaturation effects, overconfidence brings into play the third effect, which we refer to as an oversensitivity effect. At the heart of this effect is a discrepancy between what an overconfident agent expects to see — high-precision information flow — and what he actually observes — low-precision information flow. By assumption, the agent does not update his belief about the precision of the information flow,[4] and, thus, he wrongly interprets the observed low-precision information flow as an unlikely realization of the high-precision information flow. As a result, the agent deems the observed information flow as highly informative, which causes him to stop ongoing information collection sooner than he expected at the outset. In other words, the agent displays oversensitivity to noise — hence, the name of the effect: oversensitivity effect. Overall, the oversensitivity effect decreases the probability of making the correct guess.

If the agent is rational, all three effects of overconfidence — the cost, the oversaturation and the oversensitivity effects — are muted. In this case, it turns out that the principal unambiguously prefers the flexible contract. Intuitively, the flexible contract allows the agent to adjust the stopping time in response to the collected information. Hence, the agent's payoff — the probability of making the correct guess, as perceived by the agent, net of the cost of information collection — is higher under the flexible contract. In fact, in our model, the agent's perceived probability of making the correct guess in isolation is also higher under the flexible contract. When the agent is rational, his perceived probability of making the correct guess is equal to the principal's perceived probability of making the correct guess, and so, just like the agent, the principal prefers the flexible contract.

When the agent is overconfident, the flexible contract might no longer be optimal for the principal. The flexible contract creates the oversensitivity effect, which decreases the principal's perceived probability that the agent will guess correctly. As the overconfidence level increases, the oversensitivity effect progressively becomes stronger at first, which makes the commitment contract more attractive to the principal. At some overconfidence level, the oversensitivity effect becomes so strong that it overpowers the benefits of the flexible contract. Even though at very high overconfidence levels, the oversensitivity effect eventually weakens, we show that it remains sufficiently strong for the principal to prefer to shut it down with the commitment contract. Overall, the principal optimally chooses the flexible contract for a less overconfident agent and the commitment contract for a more overconfident agent.[5]

[1] An extension to non-uniform prior is studied in Appendix B.3.

[2] Some of our results continue to hold even if the agent updates his beliefs about the precision of the information. See the discussion at the end of Section 4.3 on page 11.

[3] In practice, employers can determine the level of overconfidence of their employees through various personality tests, such as the Myers-Briggs Type Indicator and the Occupational Personality Questionnaire, or through corporate assessment solutions, such as Aon and ghSMART. The assumptions that the principal knows the level of overconfidence and that the agent is unaware of his own overconfidence are in line with the exploitative contracting literature, reviewed in Section 6 in Kőszegi (2014).

[4] The assumption that the agent does not update his belief about the precision of his information reflects a quintessential feature of overconfidence — overconfident people do not acknowledge their overconfidence. For the discussion of this assumption, see the remark at the end of Section 2.

[5] Just like the agent does not update his perceived precision of the information source on the basis of the information flow that the source generates (see footnote 4), the agent does not make any inferences about his overconfidence after observing the type of the contract that the principal offers, and so, his level of overconfidence does not change.

In many applications, the principal can select an agent with desirable attributes from a pool of applicants. We incorporate this idea into the model by allowing the principal to choose both the contract type and the agent's level of overconfidence. At first glance, hiring the rational agent may seem appealing because the principal and the agent always agree on the decision. However, the principal does not internalize the agent's learning costs, and so there is a conflict of interest — the principal would always like the agent to learn more. Since an overconfident agent overestimates the value of additional information, and, thus, is more willing to collect it, the principal may find it optimal to hire an overconfident agent. We show that hiring an overconfident agent is optimal only if the principal offers him the commitment contract and learning is sufficiently costly. Intuitively, the commitment contract shuts down the undesirable oversensitivity effect, while an increase in the cost of learning strengthens the desired cost effect.

The contribution of the paper is twofold. First, we view the decomposition into the cost, oversaturation and oversensitivity effects as a stand-alone contribution to the understanding of the effects of overconfidence on the duration of learning. The decomposition remains valid in the learning problem faced by a single decision maker outside of any contractual environment and, therefore, does not depend on whether the principal knows the agent's level of overconfidence. Second, we study the interplay of the identified effects in a specific contractual environment. By considering just two contracts — the flexible and the commitment contracts — we bring into sharp focus the role of each of the effects. The commitment contract switches off the oversensitivity effect — the most remarkable of the three effects. Although not explicitly referred to as such, the cost and oversaturation effects feature in the existing literature which assumes one-shot information acquisition.[6] Our setting with the commitment contract also features only these two effects because it implies one-shot information acquisition. In contrast, the oversensitivity effect is completely novel and emerges because we consider a flexible contract in a setting with incremental information acquisition. The oversensitivity effect lies at the heart of both our main results: first, the commitment contract is optimal for the rational principal when agent is sufficiently overconfident, and second, the principal prefers hiring an overconfident agent only when his overconfidence can be reined in by the commitment contract.

Our model can be applied in a variety of settings. For example, in some jurisdictions, it is an established legal practice for judges to commit to a timetable for legal proceedings on a case before the hearing begins.[7] When the hearing is viewed as a learning process that aims to inform the just verdict, pre-commitment to a timetable seems to be unduly restrictive because it deprives the judges of the flexibility to react to information in real time. In this paper, we explain that the seeming inefficiency of the established practice can be rationalized when judges are overconfident.[8]

Beyond the realm of legal proceedings, the optimal contract recommendations in the presence of overconfidence are relevant for shareholders hiring a CEO,[9] investors trusting their money to fund managers,[10] a patient receiving her diagnosis from a physician,[11] or voters relying on a politician to make policy decisions.[12]

Our notion of overconfidence can be placed within the taxonomy proposed by Moore and Healy (2008). Moore and Healy (2008) classify overconfidence into three categories: overestimation, overplacement and overprecision. Believing that information is more precise than it actually is directly corresponds to the notion of *overprecision*. Notably, our notion of overconfidence could also be viewed as *overestimation*: the agent overestimates his ability to understand information, and thus treats the information stream as more precise than it actually is. Lastly, it should be noted that more accurate information leads to more extreme beliefs, and, thus, overprecision implies overreliance on one's belief. In settings where information collection happens outside the model, overprecision leads to overreliance on prior belief (see, for example, de la Rosa (2011)). In contrast, we model information collection explicitly and assume no overreliance on prior. However, because the agent overestimates the precision of the information he collects, overreliance on posterior arises endogeneously.

Our paper belongs to the literature on misspecified models — that is, models in which the support of the prior does not contain the actual state of the world. We model overconfidence as the misspecified degenerate prior over the precision of one's information; other papers that model overconfidence in a similar manner include Kyle and Wang (1997), Benos (1998), Daniel et al. (1998), Odean (1998), and Bernardo and Welch (2001). Overconfidence has also been modeled as the overestimation of one's ability (Heidhues et al. (2018)), correlation neglect (Ortoleva and Snowberg (2015)), and spurious correlation (Scheinkman and Xiong (2003)). In all these papers, due to the degenerate prior, agents do not update their beliefs about the variable underlying their overconfidence.

Our paper contributes to the behavioral contracting literature with moral hazard (see Kőszegi (2014) for a review). In particular, Santos-Pinto (2008) and de la Rosa (2011) model an overconfident agent who overestimates his ability to increase the value of a

[6] For example, in Gervais et al. (2011) the cost effect drives overinvestment in information (see Prediction 4 in Section III), while in Goel and Thakor (2008) the oversaturation effect drives underinvestment in information by an overconfident decision maker (see Lemma 3 in Section IV.E).

[7] For example, the United Kingdom Competition Appeal Tribunal undertakes active case management which includes "fixing a target date for the main hearing [...] together with a timetable for the proceedings up to the main hearing, [...] planning the structure of the main hearing in advance [...] and ensuring that the main hearing is conducted within defined time-limits" (see para 4.(5)(c),(e),(f) in The UK Competition Appeal Tribunal Rules). In Hong Kong, general civil proceedings at the High Court and the District Court require that "a court-determined timetable will set 'milestone dates' for the major steps in the proceedings and these dates must be adhered to" (see Q18 in FAQs on Civil Justice Reform).

[8] Empirical evidence indicates that legal decision-making is not immune to overconfidence. Loftus et al. (1988), Goodman-Delahunty et al. (2010) and Eigen and Listokin (2012) found that lawyers tend to be overconfident in estimating their chances of obtaining a self-set goal in cases that would be going to trial soon. Using data on bail decisions made by judges in New York City between 2008 and 2013, Kleinberg et al. (2018) show that judges tend to treat high-risk cases as if they are low-risk.

[9] For example, Ben-David et al. (2013) provide empirical evidence of overconfidence among executives.

[10] The overconfidence among fund managers is widely documented (e.g., Gort et al. (2008), Puetz and Ruenzi (2011), Chow et al. (2011)).

[11] Berner and Graber (2008) review evidence of diagnostic error due to overconfidence.

[12] Malmendier and Tate (2009) argue that public attention to leaders exacerbates their overconfidence.

project. Similar to our paper, these papers find that hiring an overconfident agent might be beneficial to the rational principal. The paper that is closest to ours in this literature is Gervais et al. (2011), in which an overconfident agent overestimates the precision of the acquired information, but, in contrast to our model, the information collection decision is binary. In Gervais et al. (2011), the principal unambiguously prefers hiring a more overconfident agent. Indeed, when the information collection decision is binary, the agent's overconfidence only displays the cost effect. In contrast, we show that when information collection is incremental, due to the oversaturation and oversensitivity effects, overconfidence no longer unambiguously benefits the principal.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 characterizes the optimal contract for the principal, taking the agent's level of overconfidence as given. Section 4 jointly optimizes over the type of the contract and the agent's level of overconfidence. Section 5 discusses extensions and concludes.

2. The model

Time is continuous and indexed by $t \geq 0$. The time horizon is infinite. At time $t = 0$, a principal delegates to an agent the decision of guessing an unknown binary state $z \in \{-1, 1\}$. Initially, the principal and the agent share a common prior belief according to which $z = -1$ and $z = 1$ are equally likely.

PAYOFFS. The agent has exclusive access to an information source that gradually reveals the state. Listening to the source is associated with a flow cost $c > 0$. The agent listens to the source for some time, denoted by T, and then announces his guess $v \in \{-1, 1\}$. Then, the agent's payoff is

$$\frac{vz+1}{2} - cT, \tag{1}$$

which is the payoff from the guess — that is, a payoff of 1 if the guess matches the state, and 0 otherwise — minus the cumulative cost of information collection, cT. The principal's payoff is the payoff from the guess,

$$\frac{vz+1}{2}, \tag{2}$$

so that she does not internalize the cost of information collection. There is no discounting.

LEARNING. The source reveals the state through a Brownian motion process X_t, which starts at $X_0 = 0$ and has variance normalized to 1 and state-dependent drift z[13]:

$$\mathrm{d}\,X_t = z\,\mathrm{dt} + \mathrm{d}\,W_t, \tag{3}$$

where W_t is a standard Brownian motion.[14] We reserve notation $\{x_t \mid t \geq 0\}$ for a sample path of process X_t.

The agent observes $\{x_t \mid t \geq 0\}$ but believes that it is generated by process X_t^η with variance $1/\eta$:

$$\mathrm{d}\,X_t^\eta = z\,\mathrm{dt} + \frac{1}{\sqrt{\eta}}\,\mathrm{d}\,W_t. \tag{4}$$

We do not allow the agent to learn about his misspecification (see a remark at the end of this section). Parameter $\eta > 0$ measures the level of overconfidence: the agent is **rational** if $\eta = 1$ and is **overconfident** if $\eta > 1$. We also allow η to be less than 1 — that is, the agent may suffer from underconfidence — but focus on the case of overconfidence when discussing intuition throughout the paper.[15]

The principal cannot listen to the source herself, but she has the correct belief about its variance; that is, she cannot observe $\{x_t \mid t \geq 0\}$ but believes that this sample path is generated by process X_t. The principal knows the overconfidence level η.

STRATEGIES. At $t = 0$, when delegating the guess to the agent, the principal chooses between two contracts: **commitment** and **flexible**. Under both contracts, the agent selects any stopping time $T \geq 0$ at which to stop listening to the source and, at time T, makes his guess v.[16] Under the commitment contract, the agent commits to a fixed T at $t = 0$.[17] Under the flexible contract, the agent decides when to stop in the course of collecting information — that is, the stopping time T is a random variable because it depends on the sample path $\{x_t \mid t \geq 0\}$, with the natural restriction that, at any instant t, event $\{T \leq t\}$ must be measurable with respect to the information available at time t.[18],[19]

---

[13] The normalization is without loss of generality because changing the variance of X_t is equivalent to changing the cost c of listening to the source.

[14] We discuss the robustness of our results to other information technologies in Appendices B.2.1 and B.2.2.

[15] The focus on $\eta \geq 1$ is justified on empirical grounds. Moore et al. (2015) point out that it is exceedingly rare for people to display underconfidence, in the sense of being less sure that they are right than they deserve to be.

[16] Allowing the principal to choose the duration of learning makes the problem trivial: because the principal does not internalize the agent's learning costs, she optimally chooses the longest duration that would be acceptable to the agent.

[17] Allowing the agent to stop *before* the pre-committed time effectively removes the commitment constraint. If required to commit to the *maximum* duration of learning, the agent would choose $+\infty$ as the pre-committed time, effectively making the partial commitment contract equivalent to the flexible contract. Allowing the agent to continue learning *after* the pre-committed time has a similar effect: if required to commit to the *minimum* duration of learning, the agent would choose 0 as the pre-committed time, thus again making this partial commitment contract equivalent to the flexible contract.

[18] We assume that the agent never rejects a contract. Accepting any contract is optimal if, for example, the agent's outside option is 0.

[19] The environment is not renegotiation-proof because the agent prefers the flexible contract over the commitment contract. If the agent could freely renegotiate the agreed duration of learning under the commitment contract, he would continuously do so, thus effectively mimicking the flexible contract.

A remark on variance misperception

The agent in our model is sophisticated insofar as he is able to use Bayesian updating to revise his belief about the drift of the Brownian motion. However, he does not attempt to deduce the variance of the Brownian motion, even though it can be learned with certainty in an arbitrarily short time interval.[20]

We view the lack of variance updating as a virtue of our model because it captures the idea that overconfident people are oblivious to their overconfidence and refuse to acknowledge it even in the presence of abundant evidence that their beliefs are wrong.[21] In economics, the stubborn refusal to unfavorably update beliefs about own traits has been rationalized by self-image concerns.[22]

Moreover, the empirical evidence suggests that humans find it difficult to understand variability in data. Tversky and Kahneman (1971), Kahneman and Tversky (1972) and the subsequent literature on the "law of small numbers" and "sample size neglect" document that people tend to view a small random sample from a population as highly representative. Furthermore, experimental evidence suggests that in inference problems, people tend to report overly narrow confidence intervals (see López-Pérez et al. (2021) and references therein).

We conjecture that the main forces that we have identified would remain present in a model, where the agent does update his beliefs about the precision of the information source, but the variance can no longer be deduced with certainty in an arbitrary short interval of time. A slow variance updating can be achieved, for example, through discretization of the model. Even if more realistic, a model with some variance updating would be considerably less tractable and the results would be less clear-cut.[23]

3. Analysis

3.1. Optimal guess

The agent's expected payoff from guess v is equal to his posterior belief that $z = v$. Lemma 1 derives the agent's posterior belief.

Lemma 1. *Suppose that a given sample path $\{x_t \mid t \geq 0\}$ is generated by process X_t^η, and z is equally likely to be -1 or 1. Then, at time t, the posterior probability that $z = 1$ is equal to $p(\eta x_t)$, where function $p(x)$ is defined as*

$$p(x) = \frac{1}{1 + \exp(-2x)}. \tag{5}$$

Proof. See Appendix A.1. □

The agent optimally guesses the state that is more likely according to his posterior belief — that is, he optimally guesses $v = 1$ if $p(\eta x_T) > 0.5$ and $v = -1$ if $p(\eta x_T) < 0.5$. By (5), condition $p(\eta x_T) > 0.5$ is equivalent to condition $x_T > 0$. Thus, the agent's optimal guess is

$$v = \mathrm{sign}(x_T). \tag{6}$$

3.2. Commitment contract

Under the commitment contract, at $t = 0$, the agent chooses the duration of learning T that maximizes the expected payoff from the optimal guess, derived in Lemma 2, minus the cost of information collection, cT.

Lemma 2. *Suppose that a given sample path $\{x_t \mid t \geq 0\}$ is generated by process X_t^η, and z is equally likely to be -1 or 1. Then, at $t = 0$, the probability that guess (6) matches the state z is $f(\eta T)$, where function $f(t)$ is defined as*

$$f(t) = \frac{1}{2}\left(1 + erf\left(\sqrt{\frac{t}{2}}\right)\right), \tag{7}$$

[20] Scheinkman and Xiong (2003) circumvent the problem of variance updating by making the variance of the information process independent of the agent's level of overconfidence. In particular, in Scheinkman and Xiong (2003), an overconfident agent overestimates the correlation between the state process and the noisy process he observes. We cannot use the same modeling technique because our state is not stochastic.

[21] In psychology, Baumeister et al. (1996) reviews evidence suggesting that people with unreasonably high self-esteem choose to respond to unfavorable feedback with violence, instead of revising their self-views downwards. In neuroscience, there is a well-known syndrome called anosognosia, which refers to the tendency in some sane, intelligent and articulate paralysis patients to ignore or even deny the fact that a part of their own body is paralyzed. In the book Ramachandran and Blakeslee (1998), Dr Ramachandran claims that "[w]atching [anosognosia] patients is like observing human nature through a magnifying lens" and that these clinical cases are "a comically exaggerated version of all those psychological defense mechanisms [...] used by you, me and everyone else when we are confronted with disturbing facts about ourselves.".

[22] Benabou and Tirole (2002) provide micro-foundations for awareness management and self-deception to maintain positive self-image. Although we do not model it explicitly, our agent's refusal to update the precision of the information source downwards can be viewed as an example of such self-serving deception. Had the agent updated the precision downwards, he would have had to admit that his information was less precise, which might indicate his low ability to understand information.

[23] In Appendix B.1, we consider a model specification in which we allow an overconfident agent to learn the variance in an arbitrarily short time interval. In this specification, we show that some of our results, in particular, Proposition 4 and Theorem 2, continue to hold.

and $erf(y) = \frac{2}{\sqrt{\pi}} \int_0^y \exp(-\tilde{y}^2)\, d\tilde{y}$ *is the Gauss error function.*

Proof. See Appendix A.2. □

Hence, the agent chooses T, which solves

$$\max_{T \in [0,\infty)} f(\eta T) - cT. \tag{8}$$

Since f is concave, the first-order condition characterizes the unique maximum:

$$f'(\eta T) = \frac{c}{\eta}. \tag{9}$$

The solution to (9) exists and is unique because $f'(t)$ is decreasing from $+\infty$ at $t = 0$ to 0 at $t = +\infty$.

By (2), the principal maximizes her estimate of the probability that the agent's guess is correct — that is, it matches the state. The principal believes that $\{x_t \mid t \geq 0\}$ is generated by process $X_t = X_t^1$. Thus, by Lemma 2, she estimates the probability that the agent's guess is correct as $f(T)$.

To determine the effect of overconfidence η on the principal's payoff $f(T)$, we note that $f(T)$ increases in T and analyze the effect of η on the agent's choice of T. First, the entry of η on the right-hand side of (9) captures the **cost effect**. Higher η effectively lowers the flow cost of learning, thus incentivizing the agent to collect additional information — that is, the optimal T increases through the cost effect. Second, the entry of η on the left-hand side of (9) captures the **oversaturation effect**. Higher η increases the perceived precision of information already collected by T, thus disincentivizing the agent from collecting additional information — that is, the optimal T decreases through the oversaturation effect. The oversaturation effect emerges because the agent can choose any positive T. Had the agent been restricted to a binary choice between $T = 0$ and an exogeneously given $T = \bar{T} > 0$, there would have been only the cost effect of η on the optimal T.[24]

3.3. Flexible contract

Under the flexible contract, the agent chooses, at every t, whether to continue listening to the source or to stop and make a guess. It is well known (e.g., Shiryaev (1978), Chapter 4, Theorem 5, p.185) that the solution to optimal stopping problems of this kind is characterized by two thresholds on the agent's current belief: the agent listens to the source when his belief that $z = 1$ is between thresholds $\underline{p}$ and $\bar{p}$, where $\underline{p} < \bar{p}$; otherwise, he stops and makes the optimal guess. Due to the symmetry of the payoff function — the agent gets the same payoff when $v = z = 1$ and $v = z = -1$, and when $v \neq z = 1$ and $v \neq z = -1$ — the belief thresholds are symmetric around 0.5; that is, $\underline{p} = 1 - \bar{p}$. As noted in a remark on page 186 in Shiryaev (1978), the optimal $\bar{p} \in (0.5, 1)$ is defined as the unique root of the equation

$$\frac{2\eta}{c} = \frac{\bar{p}}{1-\bar{p}} - \frac{1-\bar{p}}{\bar{p}} + 2\ln\frac{\bar{p}}{1-\bar{p}}. \tag{10}$$

By Lemma 1, at time t, the agent's belief that $z = 1$ is equal to $p(\eta x_t)$. Then, according to formula (5) for function $p(x)$, thresholds $1 - \bar{p}$ and $\bar{p}$ on the belief process uniquely correspond to thresholds $-\chi$ and χ on the process X_t^η:

$$p(\eta\chi) = \bar{p}. \tag{11}$$

Hence, the agent optimally stops listening to the source when the absolute value of the process, $|X_t^\eta|$, reaches threshold χ for the first time.

The principal maximizes her estimate of the probability that the agent's guess is correct. Given the agent's optimal guess (6) and his stopping rule $|x_T| = \chi$, this probability is equal to

$$\Pr\left(z = 1 \mid X_T = \chi\right) \cdot \Pr\left(X_T = \chi\right) + \Pr\left(z = -1 \mid X_T = -\chi\right) \cdot \Pr\left(X_T = -\chi\right). \tag{12}$$

By symmetry, ex ante, the process X_t is equally likely to hit either of the stopping thresholds — that is, $\Pr\left(X_T = \chi\right) = \Pr\left(X_T = -\chi\right) = 0.5$ — and the guess is equally likely to be correct at each threshold — that is, $\Pr\left(z = 1 \mid X_T = \chi\right) = \Pr\left(z = -1 \mid X_T = -\chi\right)$. Hence, expression (12) is equal to $\Pr\left(z = 1 \mid X_T = \chi\right)$. Since $X_t = X_t^1$, by Lemma 1, $\Pr\left(z = 1 \mid X_T = \chi\right)$ is equal to $p(\chi)$.

To determine the effect of overconfidence η on the principal's payoff $p(\chi)$, we note that $p(\chi)$ increases in χ and analyze the effect of η on the agent's choice of χ. The optimal χ is defined by equations (10) and (11), which contain two entries of η. First, the entry of η on the left-hand side of (10) captures the cost effect from increasing η because it is equivalent to the effect from decreasing c. We have already encountered this effect in the commitment contract. Since the right-hand side of (10) is increasing $\bar{p}$, according to the cost effect, a higher level of overconfidence η increases the belief threshold $\bar{p}$ at which the agent stops learning, which, in turn, by (11), increases the optimal χ. Second, the entry of η on the left-hand side of (11) captures the change in χ due to the distortion of

[24] The cost effect of overconfidence on a binary decision whether to collect information is presented in Section III in Gervais et al. (2011).

the belief-updating process. Formally, by (11), for a given belief threshold $\bar{p}$, the product of η and χ must be constant, and so, higher η lowers χ. Intuitively, higher η increases the informativeness of process X_t^η, thus lowering threshold χ needed to achieve a given belief threshold. In the next section, we show that the entry of η in equation (11) corresponds to a combination of the oversaturation effect, familiar from the commitment contract, and a new effect, which we refer to as the oversensitivity effect.

3.4. Oversensitivity effect

The principal chooses the contract by comparing her expected payoffs under the commitment contract and the flexible contract. Both payoffs are affected by the cost and the oversaturation effects created by the agent's overconfidence. However, the dynamic nature of the agent's optimization problem under the flexible contract introduces an additional effect, referred to as the oversensitivity effect, which lowers the principal's payoff. The presence of the oversensitivity effect under the flexible contract is the key driving force that makes the commitment contract more appealing to the principal.

Under the flexible contract, both the oversaturation and oversensitivity effects operate through the distortion of the belief-updating process, and so can be analyzed by looking at the entry of η in equation (11). Under the commitment contract, the oversaturation effect works through the time at which the agent stops learning. Therefore, to extract the oversaturation effect from equation (11), we map χ to the stopping time. More specifically, we look at the principal's estimate of the expected stopping time — that is, the expected time needed for process $|X_t|$ to reach threshold χ for the first time. We denote this time by $T_1^e(\chi)$. Clearly, $T_1^e(\chi)$ increases with χ because process X_t spends a longer time within the strip $[-\chi, \chi]$ when the strip is wider. Hence, there is a one-to-one correspondence between χ and $T_1^e(\chi)$. We analyze the effect of η on $T_1^e(\chi)$, keeping the product $\eta\chi \equiv \bar{\chi}$ fixed because equation (11) disciplines this product through a fixed belief threshold $\bar{p}$ — that is, we look at the effect of η on $T_1^e(\bar{\chi}/\eta)$ keeping $\bar{\chi}$ fixed.

One way to compare the contracts is to look at the total distortion caused by the agent's overconfidence relative to the rational agent. To this end, we look at the difference between the expected stopping times $T_1^e(\bar{\chi}/\eta)$, calculated for some level of overconfidence $\eta > 1$, and $T_1^e(\bar{\chi})$, calculated for the rational agent.

The oversaturation effect can be extracted from $T_1^e(\bar{\chi}/\eta) - T_1^e(\bar{\chi})$ by adding and subtracting the expected stopping time, as estimated by the agent — that is, the expected time needed for process $|X_t^\eta|$ to reach threshold $\chi = \bar{\chi}/\eta$ for the first time. We denote this time by $T_\eta^e(\bar{\chi}/\eta)$.

$$T_1^e(\bar{\chi}/\eta) - T_1^e(\bar{\chi}) = \underbrace{T_\eta^e(\bar{\chi}/\eta) - T_1^e(\bar{\chi})}_{\text{oversaturation effect}} + \underbrace{T_1^e(\bar{\chi}/\eta) - T_\eta^e(\bar{\chi}/\eta)}_{\text{oversensitivity effect}}. \tag{13}$$

Proposition 1 implies that the oversaturation effect is negative because $T_\eta^e(\bar{\chi}/\eta)$ decreases in η.

Proposition 1. *For any $\bar{\chi} > 0$, the expected stopping time $T_\eta^e(\bar{\chi}/\eta)$ is decreasing in $\eta > 0$.*

Proof. See Appendix A.3. □

The intuition behind the oversaturation effect under the flexible contract is the same as under the commitment contract: an overconfident agent overestimates the precision of any given stock of information and, therefore, stops listening to the source sooner than the rational agent would.

The second effect in (13) is new and captures the discrepancy between $T_\eta^e(\bar{\chi}/\eta)$ and $T_1^e(\bar{\chi}/\eta)$. Relative to the rational agent, an overconfident agent overestimates the expected stopping time because, by Proposition 2 stated below, $T_\eta^e(\chi)$ is increasing in η, and so, if $\eta > 1$, then $T_\eta^e(\chi) > T_1^e(\chi)$ for any fixed threshold χ, including $\chi = \bar{\chi}/\eta$.

Proposition 2. *For any $\chi > 0$, the expected stopping time $T_\eta^e(\chi)$ is increasing in $\eta > 0$.*

Proof. See Appendix A.3. □

We refer to the effect of η on $T_1^e(\bar{\chi}/\eta) - T_\eta^e(\bar{\chi}/\eta)$ as the **oversensitivity effect** because it emerges due to the agent's oversensitivity to noise in the observed process. Fig. 1 illustrates the oversensitivity effect. Let's fix state $z = 1$ and a sample path of the standard Brownian motion W_t. Then, the right-hand panel of Fig. 1 depicts the corresponding sample path of process X_t defined in (3). This sample path crosses threshold χ at time $T_1(\chi)$. The agent, however, expects to observe the process X_t^η defined in (4) and depicted in the left-hand panel of Fig. 1. Had the agent been rational — that is, $\eta = 1$ — the processes X_t and X_t^η would have been identical. Fig. 1, however, is drawn for $\eta = 4$, and, thus, process X_t^η is a compressed version of X_t because the overconfidence level η suppresses the variance. According to Fig. 1, the sample path of the expected process X_t^η crosses threshold χ at time $T_\eta(\chi) > T_1(\chi)$. Proposition 2 proves that the depicted sample paths are not exceptional, and inequality $T_\eta(\chi) > T_1(\chi)$ holds in expectation, which captures the oversensitivity effect.

The oversensitivity effect — that is, the difference $T_1^e(\bar{\chi}/\eta) - T_\eta^e(\bar{\chi}/\eta)$ — is non-monotone in the level of overconfidence. Since the difference $T_1^e(\bar{\chi}/\eta) - T_\eta^e(\bar{\chi}/\eta)$ is negative, its decrease implies strengthening, while its increase implies weakening, of the oversensitivity effect. For a given threshold χ, by Proposition 2, the difference $T_1^e(\chi) - T_\eta^e(\chi)$ is decreasing in η, which indicates that the

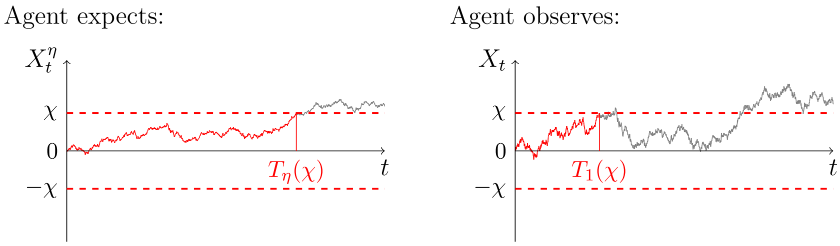


Fig. 1. Sample paths of processes $X_t = t + W_t$ (right panel) and $X_t^\eta = t + \frac{1}{\sqrt{\eta}} W_t$ for $\eta = 4$ (left panel).

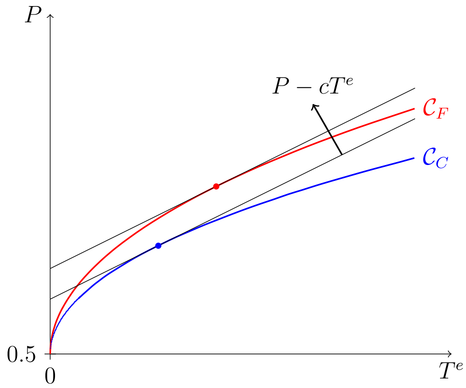


Fig. 2. The rational agent's choice. The expected stopping time, T^e, is on the x-axis; the highest achievable probability of making the correct guess, P, is on the y-axis. Curve C_F : $T = T_1^e(\chi)$, $P = p(\chi)$ corresponds to the flexible contract; curve C_C : $P = f(T^e)$ corresponds to the commitment contract. The tangency points of C_F and C_C with line $P - cT^e =$ const correspond to the rational agent's choice for $c = 1$.

oversensitivity effect strengthens with η. However, threshold $\chi = \bar{\chi}/\eta$ itself depends on η, and functions $T_\eta^e(\chi)$ and $T_1^e(\chi)$ change in χ at different rates, which may result in weakening of the oversensitivity effect with η. Proposition 3 shows that the oversensitivity effect strengthens for low η and weakens for high η.

Proposition 3. *For any $\bar{\chi} > 0$, there exists $\eta_h > 1$ such that the difference $T_1^e(\bar{\chi}/\eta) - T_\eta^e(\bar{\chi}/\eta)$ is decreasing in $\eta \in (0, \eta_h)$ and increasing in $\eta \in (\eta_h, +\infty)$.*

Proof. See Appendix A.4. □

Proposition 3 is intuitive. It is expected that the difference $T_1^e(\bar{\chi}/\eta) - T_\eta^e(\bar{\chi}/\eta)$ is decreasing in η for η just above 1. Indeed, this difference is 0 at $\eta = 1$ and, by Proposition 2, negative for all $\eta > 1$. Eventually, this difference must start increasing because it is close to 0 at high η. Indeed, for high η, threshold $\bar{\chi}/\eta$ is low, which means that both $T_1^e(\bar{\chi}/\eta)$ and $T_\eta^e(\bar{\chi}/\eta)$ are close to 0, and so, the difference $T_1^e(\bar{\chi}/\eta) - T_\eta^e(\bar{\chi}/\eta)$ is close to 0. In words, as η increases from 1, the oversensitivity effect emerges and starts to strengthen; as η becomes very high, both the time at which the agent expects to stop and the actual time at which the agent stops become low in expectation, and so, their difference, which characterizes the oversensitivity effect, is low.

3.5. Optimal contract

When choosing the contract, the principal faces the trade-off between the cost of the oversensitivity effect, described in the previous section, and the benefit of flexible information acquisition. The optimal resolution of this trade-off depends on the level of overconfidence.

If the agent is rational, the principal optimally chooses the flexible contract. For any given expected time, denoted by T^e, Fig. 2 depicts P, which is the highest probability of making the correct guess that the rational agent can achieve under each contract. The flexible contract curve C_F always lies above the commitment contract curve C_C because under the flexible contract, the agent can always replicate his behavior under the commitment contract and can do strictly better. Under each contract, the agent's learning strategy corresponds to a point on the relevant curve, C_F or C_C. The agent's choice trades off the probability of making the correct guess and the expected time in the same way under both contracts: in the (T^e, P) space, the agent's objective function corresponds to a straight line $P - cT^e$, with weight c on T^e common to both contracts. According to Fig. 2, the optimal point on C_F has a higher probability of making the correct guess than the optimal point on C_C — that is, the probability of making the correct guess is higher under the flexible contract. Hence, the principal chooses the flexible contract.

If the agent is overconfident, there is a wedge between the agent's and the principal's estimates of the probability that the agent guesses correctly. From the agent's perspective, the flexible contract always leads to a higher probability of making the correct guess. However, this might not be the case from the principal's perspective because of the oversensitivity effect.

Overall, the commitment plays a double role. On the one hand, it prevents an overconfident agent from reacting to noise and stopping too soon — that is, it turns off the oversensitivity effect. On the other hand, the agent's commitment also prevents him from adjusting the stopping time in response to useful information, which is sub-optimal from the principal's perspective. At low levels of overconfidence η, by Proposition 3, the oversensitivity effect becomes stronger with η, and, thus, intuitively, the relative benefit of the commitment contract increases with η. Theorem 1 states that there exists a threshold for the level of overconfidence such that the commitment contract is optimal above this threshold and the flexible contract is optimal below this threshold. The uniqueness of the threshold is somewhat surprising because, according to Proposition 3, the oversensitivity effect is non-monotone. Since the oversensitivity effect weakens with η at high η, one might expect two thresholds to delineate the region of η where the commitment contract is optimal. However, Theorem 1 proves that the oversensitivity effect remains sufficiently strong for the commitment contract to be optimal for large η.

Theorem 1. *For any marginal cost $c > 0$, there exists a threshold $\eta^*(c) > 1$ such that the principal optimally chooses the commitment contract if $\eta > \eta^*(c)$ and the flexible contract if $0 < \eta < \eta^*(c)$. Threshold $\eta^*(c)$ is increasing in c.*

Proof. See Appendix A.5. □

The threshold for the overconfidence level, at which the flexible contract ceases to be optimal for the principal, increases with the cost of information collection. Intuitively, as the cost increases, the advantage of the flexible contract becomes stronger because flexibility allows the agent to expend costly resources more efficiently. At the same time, the oversensitivity effect, which underlies the advantage of the commitment contract, does not depend on the cost because it arises through the distortion of the belief updating captured in equation (11). Thus, the region of overconfidence levels, in which the flexible contract is optimal for the principal, expands with the cost.

4. Selecting an agent

In the model description in Section 2 and the analysis in Section 3.5, we take the agent's overconfidence level as exogenously given. In reality, employers routinely screen job applicants through various personality tests that reveal each applicant's overconfidence level. In this section, we incorporate into the model the reduced form of the overconfidence screening by allowing the principal to choose the overconfidence level η. Sections 4.1 and 4.2 investigate the behavior of the principal's payoff as a function of the agent's level of overconfidence under the commitment and the flexible contract, respectively. Section 4.3 maximizes the principal's payoff over both the contract's type and the level of overconfidence.

4.1. Commitment contract

According to Section 3.2, the level of overconfidence η affects the principal's payoff through the cost and oversaturation effects, which work in opposite directions. For low η, as η increases, the cost effect takes an upper hand over the oversaturation effect, increasing the agent's learning time and, thus, the principal's payoff. For high η, the oversaturation effect is stronger and the principal's payoff decreases in η. Hence, the principal's payoff as a function of η has a hump-shaped form, as stated in Proposition 4 and illustrated in Fig. 3.

Intuitively, there is an asymmetry between the effects. The combination of the cost and the oversaturation effects determines the learning time, but the learning time affects only the oversaturation effect and not the cost effect. If the cost effect prevails, the learning time increases. An increase in the learning time strengthens the oversaturation effect because it increases the amount of collected information through which the oversaturation effect operates. This channel is stronger for higher η because the perceived precision of information collected during the additional learning time increases in η. Hence, whenever the cost effect takes an upper hand, the oversaturation effect immediately strengthens in proportion to η; thus, for high η, the oversaturation effect prevails.

Proposition 4. *Under the commitment contract, the principal's payoff increases in $\eta \in (0, \eta_C^*)$ and decreases in $\eta \in (\eta_C^*, +\infty)$, where*

$$\eta_C^* = 2c\sqrt{2e\pi}. \tag{14}$$

Proof. See Appendix A.6. □

It is intuitive that the optimal level of overconfidence η_C^* in (14) is increasing in cost c. A higher cost decreases the amount of collected information and, therefore, weakens the oversaturation effect. Hence, the range of η where the cost effect overpowers

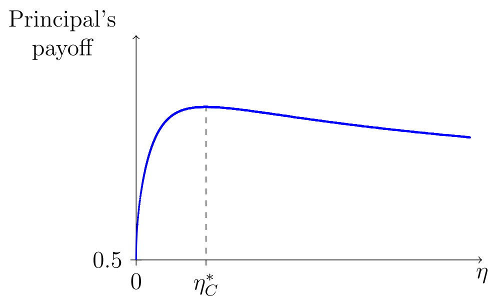


Fig. 3. The principal's payoff under the commitment contract as a function of the level of overconfidence η. The marginal cost of learning is $c = 1$. Point η_C^* indicates the level of overconfidence that maximizes the principal's payoff.

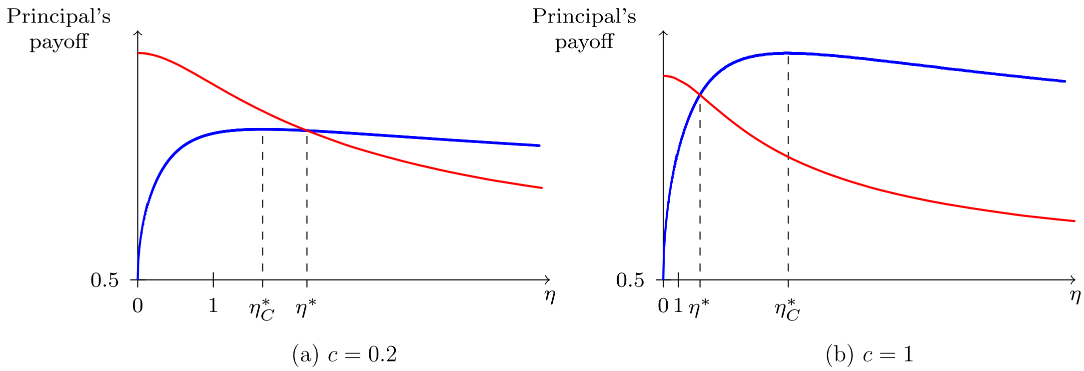


Fig. 4. The principal's payoff under the commitment contract (blue curve) and under the flexible contract (red curve) as a function of the level of overconfidence η for different learning costs.

the oversaturation effect expands, moving to the right the level of overconfidence η_C^* that balances the cost and the oversaturation effects.[25]

4.2. Flexible contract

According to Sections 3.3 and 3.4, the level of overconfidence η affects the principal's payoff through three effects: the cost effect increases the principal's payoff as η increases, while the oversaturation and oversensitivity effects work in the opposite direction. At first glance, the overall effect of η seems to be ambiguous. However, Proposition 5 shows that there is no ambiguity, and overconfidence always has a detrimental effect on the principal's payoff under the flexible contract.

Proposition 5. *Under the flexible contract, the principal's payoff is decreasing in $\eta > 0$.*

Proof. See Appendix A.7. □

Intuitively, the oversaturation effect is strong when the agent has already collected a lot of information. In contrast, the oversensitivity effect is strong when the agent has little information and, therefore, his beliefs are especially sensitive to noise in the observed process. Overall, irrespective of the amount of information already collected, working together, the oversaturation and oversensitivity effects overpower the cost effect.[26]

4.3. Agent selection and optimal contract

Fig. 4 illustrates the principal's payoff under the commitment contract and the flexible contract. In accordance with Proposition 4, the principal's payoff from the commitment contract is hump-shaped with the peak η_C^*. In accordance with Proposition 5, the payoff from the flexible contract is decreasing throughout the whole range $\eta > 0$.

Theorem 2 maximizes the principal's payoff over both the contract's type and the level of overconfidence. In Fig. 4, the maximum corresponds to the highest point on the upper envelope of the two payoff functions.

[25] Noteworthy, in Proposition 4, the peak of the principal's payoff under the commitment contract may occur at $\eta_C^* < 1$.

[26] In Appendix B.3, we show that if the agent has a non-uniform prior, then the effect of overconfidence on the principal's payoff is no longer unambiguous.

Theorem 2. *Suppose that the principal can select any $\eta \geq \underline{\eta}$ for some fixed $\underline{\eta} \in (0, 1]$. Then, there exists $c^*(\underline{\eta}) > 0$ such that if $c < c^*(\underline{\eta})$, the principal selects the agent with $\eta = \underline{\eta}$ and offers him the flexible contract; and if $c > c^*(\underline{\eta})$, she selects the overconfident agent with $\eta = \overline{\eta}_C^* > 1$ defined in* (14) *and offers him the commitment contract.*

Proof. See Appendix A.8. □

Theorem 2 assumes that the principal can choose any overconfidence level η greater than or equal to $\underline{\eta} \in (0, 1]$.[27] Remarkably, she always hires an overconfident agent with $\eta = \eta_C^* > 1$, whenever she finds it optimal to offer the commitment contract.

Although the theorem is formulated for any $\underline{\eta} \in (0, 1]$, case $\underline{\eta} = 1$ is especially interesting because it corresponds to the rational agent. In this case, the theorem implies that whenever the principal finds it optimal to offer the flexible contract, she hires the rational agent. For the remainder of the section, we assume that $\underline{\eta} = 1$.

As illustrated in Fig. 4, the principal's payoffs under the commitment and the flexible contracts intersect at a unique point $\eta^* > 1$ introduced in Theorem 1. Hence, if point η_C^*, at which the payoff under the commitment contract achieves its maximum, is lower than the intersection point η^*, then the principal optimally hires the rational agent and offers him the flexible contract. This case is illustrated in the left panel of Fig. 4. If η_C^* is greater than η^*, then the principal's choice is narrowed down to two options: the commitment contract with $\eta = \eta_C^*$ and the flexible contract with $\eta = 1$. The principal chooses the option that offers her the highest payoff. The right panel of Fig. 4 illustrates the case when the payoff under the commitment contract at $\eta = \eta_C^*$ is higher than the payoff under the flexible contract at $\eta = 1$; thus, the principal optimally hires the overconfident agent with $\eta = \eta_C^*$ and offers him the commitment contract.

Fig. 4 contrasts the principal's optimal choice at different costs of learning: the left panel is drawn for $c = 0.2$ and the right panel is drawn for $c = 1$. In accordance with Theorem 2, the flexible contract with $\eta = 1$ is optimal for low cost in the left panel, and the commitment contract with $\eta = \eta_C^*$ is optimal for high cost in the right panel. Intuitively, the principal does not internalize the agent's costs, and so, always prefers more prolonged learning by the agent. She may be able to exploit the agent's overconfidence to increase his learning time. The advantage of overconfidence is in the cost effect, which strengthens with the cost of learning. However, hiring an overconfident agent may backfire and decrease the learning time through the oversaturation and oversensitivity effects, which are independent of the learning cost. If the principal had only the flexible contract at her disposal, the disadvantage of the combination of the oversaturation and oversensitivity effects would outweigh the benefit of overconfidence brought about by the cost effect. However, in combination with the commitment contract, for a high learning cost, the cost effect overpowers the downside of overconfidence because the commitment contract shuts down the oversensitivity effect. Hence, the principal optimally chooses the rational agent for low c and an overconfident agent for high c.

Theorem 2 for $\underline{\eta} = 1$ continues to hold even if, in the course of learning, the agent updates his beliefs about the precision of the information. Indeed, as mentioned in the remark on variance misperception on page 5, if the agent is allowed to revise his belief about the variance of the Brownian motion, he learns it in an arbitrarily short interval of time and, thus, under the flexible contract, behaves as the rational agent. Then, in effect, the principal's choice set is smaller than in the original model without variance updating — the principal cannot choose any $\eta > 1$ together with the flexible contract. According to Theorem 2, however, the principal will never choose the flexible contract with some $\eta > 1$ even if she can. Hence, Theorem 2 may continue to hold even if the agent updates his belief about the variance of the Brownian motion.[28]

5. Discussion and conclusion

We show that the principal may find it optimal to hire an overconfident agent to make decisions that require costly learning. This seems reassuring in light of empirical evidence suggesting that people in positions of power, such as CEOs or high-ranking politicians, display higher levels of overconfidence than the average person. However, we also conclude that the discretion of overconfident CEOs and politicians must be restrained through commitment to a particular timing of the decision.

Throughout the paper, we focus on the case of overconfidence, that is, $\eta > 1$. However, all formal results also apply to the case of underconfidence, that is, $\eta < 1$. Underconfidence gives rise to the same three effects — the cost, oversasuration, and oversensitivity effects. However, in comparison to the case of overconfidence, underconfidence makes each effect work in the opposite direction: relative to the rational agent with $\eta = 1$, an underconfidence level $\eta < 1$ causes the cost effect to decrease and the other two effects

[27] The lower bound $\underline{\eta}$ is greater than zero to ensure that the agent's perceived information process always has finite variance.

[28] In general, it is not immediately clear how to extend Theorem 2 to any prior belief about the variance, because it is not obvious how to define the level of overconfidence for the agent with a non-degenerate belief about the variance of the information process. However, we identified two cases in which Theorem 2 is easy to extend.

First, if the agent places a positive but arbitrary small prior probability on all variances other than $1/\eta$, we conjecture that the duration of the agent's learning under the commitment contract does not materially depart from his duration of learning in the benchmark model with the degenerate misspecified prior. Thus, Theorem 2 remains unchanged.

Second, in Appendix B.1, we formulate an alternative model in which the agent with the level of overconfidence η places a positive prior probability α on the correct variance 1 and the complementary probability $1 - \alpha$ on variance $1/\eta$. In this model, we show that under the commitment contract, the principal's payoff as a function of η still has a hump-shaped form with the peak at some $\eta_C^*(\alpha, c)$, that is, Proposition 4 continues to hold. Moreover, the commitment contract offered to the agent with the level of overconfidence $\eta = \eta_C^*(\alpha, c)$ is better for the principal than the flexible contract offered to the rational agent if and only if c is above some threshold $c^*(\alpha)$, that is, Theorem 2 continues to hold. We thank an anonymous referee for suggesting this extension.

to increase the amount of collected information. By Theorem 1, the principal always offers the flexible contract to an underconfident agent. By Theorem 2, whenever the principal offers the flexible contract, she hires the most underconfident agent she has access to.

Throughout the paper, we work with the uniform prior to shut down any effects due to the ex post disagreement between the principal and the agent. With the uniform prior, after any amount of learning, the principal and the agent agree on the optimal guess. When the prior is non-uniform, a little bit of learning may be enough to sway an overconfident agent, but not enough to sway the principal in favor of the a priori less likely state. When the learning cost is sufficiently high, the agent stops learning quickly. As a result, anticipating the ex post disagreement about the best guess, the principal does not want to hire the agent at all. Another consequence of a non-uniform prior is that under the flexible contract, the principal's payoff has a hump shape as a function of the overconfidence level. A non-uniform prior weakens the oversensitivity effect for low levels of overconfidence, where the oversaturation effect is also weak. Hence, for low levels of overconfidence, the cost effect overpowers the combination of the oversaturation and oversensitivity effects, and the principal's payoff increases in the overconfidence level. Our main result from Theorem 1 carries over to the setting with a non-uniform prior: whenever hiring an agent is worthwhile, the principal prefers the flexible contract for low levels of overconfidence and the commitment contract for high levels of overconfidence. Theorem 2, however, holds only partially. As with the uniform prior, the principal may find it optimal to hire an overconfident agent. Yet, since the principal's payoff under the flexible contract takes the hump-shaped form, the principal may find it optimal to offer the hired overconfident agent the flexible contract. See Appendix B.3 for details.

In our analysis, we rely on a specific information technology — namely, a Brownian motion, which in the case of the commitment contract, results in a learning cost that is linear in time t and in the probability of the correct guess given by $f(\eta t)$ in (7). Under the commitment contract, the agent chooses time t, and with that, the learning cost $\zeta := ct$, which gives him the probability of the correct guess equal to $\tilde{f}(\eta\zeta) := f(\eta\zeta/c)$. Appendix B.2.1 shows that the explicit form of function $\tilde{f}(\zeta)$ does not matter for Proposition 4, and, under quite general assumptions on $\tilde{f}(\zeta)$, the principal's payoff has a hump-shaped form with the unique maximum. Consequently, our insight about two opposing forces — the cost and oversaturation effects — that are at work under the commitment contract generalizes to alternative information technologies.[29] Under the flexible contract, generalizing information technology is mathematically more challenging. In Appendix B.2.2, we discuss the robustness of the oversensitivity effect to other information technologies. We conclude that to avoid introducing additional effects and changing the definition of overconfidence, the information process should be based on a normal distribution. Furthermore, overconfidence should only distort the variance of the information process, without affecting its expected increment in any given time interval.

Declaration of competing interest

The authors declare that have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the paper "Imposing Commitment to Rein in Overconfidence in Learning".

Data availability

No data was used for the research described in the article.

Appendix A. Proofs

A.1. Proof of Lemma 1

The motion rule (4) implies that at time t, X_t^η is distributed according to a normal distribution with mean zt and variance t/η. Then, by Bayes' rule,

$$\begin{aligned}\Pr\left(z=1 \mid X_t^\eta = x_t\right) \\ &= \frac{\Pr(z=1)\Pr\left(X_t^\eta = x_t \mid X_t^\eta \sim \mathcal{N}(t,t/\eta)\right)}{\Pr(z=1)\Pr\left(X_t^\eta = x_t \mid X_t^\eta \sim \mathcal{N}(t,t/\eta)\right)+\Pr(z=-1)\Pr\left(X_t^\eta = x_t \mid X_t^\eta \sim \mathcal{N}(-t,t/\eta)\right)} \\ &= \frac{\Pr(z=1)}{\Pr(z=1)+\Pr(z=-1)\exp\left(-2\eta x_t\right)}. \qquad \text{(A.1)}\end{aligned}$$

By assumption, the prior is uniform; that is, $\Pr(z=1)=\Pr(z=-1)$. Then, (A.1) is equal to $p(\eta x_t)$, where function $p(x)$ is defined in (5).

A.2. Proof of Lemma 2

Given $X_T^\eta \sim \mathcal{N}(zT, T/\eta)$ and the uniform prior for $z \in \{-1,1\}$,

[29] An information technology that generates a given function $\tilde{f}(\zeta)$ is a binary signal that matches the state with probability $\tilde{f}(\zeta)$.

$$\Pr(\text{sign}(x_T)=z)=\frac{1}{2}\frac{1}{\sqrt{2\pi T/\eta}}\int_0^{+\infty}\exp\left(-\frac{(x-T)^2}{2T/\eta}\right)\mathrm{d}x + \frac{1}{2}\frac{1}{\sqrt{2\pi T/\eta}}\int_{-\infty}^{0}\exp\left(-\frac{(x+T)^2}{2T/\eta}\right)\mathrm{d}x=\frac{1}{2}+\frac{1}{\sqrt{\pi}}\int_0^{\sqrt{\frac{\eta T}{2}}}\exp\left(-x^2\right)\mathrm{d}x, \quad \text{(A.2)}$$

which is equal to $f(\eta T)$, where function $f(t)$ is defined in (7).

A.3. Proof of Propositions 1 and 2

Denote

$$T_\eta^e(\chi)=\mathbb{E}\left[\inf\left\{t\geq 0 \mid |X_t^\eta|=\chi\right\}\right]$$

the expected time needed for process $|X_t^\eta|$ to reach threshold χ for the first time. Lemma 3 gives the expression for this time for any threshold $\chi>0$ and overconfidence level $\eta>0$.

Lemma 3.

$$T_\eta^e(\chi)=\frac{1-\exp(-2\eta\chi)}{1+\exp(-2\eta\chi)}\chi.$$

Proof. By Darling and Siegert (1953) (see the formula in the middle of p.633 and formula (5.7)), the expected time of the first exit from the strip $[-\chi,\chi]$ for a Brownian motion with drift μ and variance $1/\eta$ is

$$\mathcal{T}_\eta^e(\mu,\chi)=\frac{1-\exp(-2\eta\chi)}{1+\exp(-2\eta\chi)}\chi. \quad \text{(A.3)}$$

The drift of process X_t^η is either 1 or -1 with equal probability. Then, the statement of the lemma follows by substituting (A.3) into

$$T_\eta^e(\chi)=\frac{1}{2}\mathcal{T}_\eta^e(1,\chi)+\frac{1}{2}\mathcal{T}_\eta^e(-1,\chi). \quad \square \quad \text{(A.4)}$$

Lemma 3 implies that

$$T_\eta^e\left(\frac{\bar\chi}{\eta}\right)=\frac{1-\exp(-2\bar\chi)}{1+\exp(-2\bar\chi)}\frac{\bar\chi}{\eta}, \quad \text{(A.5)}$$

which is clearly decreasing in η. Hence, Proposition 1 follows.

Lemma 3 gives the expression for $T_\eta^e(\chi)$. Proposition 2 follows from

$$\frac{\partial}{\partial\eta}T_\eta^e(\chi)=\frac{4\exp(2\eta\chi)\chi^2}{(1+\exp(2\eta\chi))^2}>0. \quad \text{(A.6)}$$

A.4. Proof of Proposition 3

The first derivative of the difference $T_1^e(\bar\chi/\eta)-T_\eta^e(\bar\chi/\eta)$ can cross zero only from below because, whenever the first derivative is zero, the second derivative is positive:

$$\frac{\partial^2}{\partial\eta^2}\left(T_1^e\left(\frac{\bar\chi}{\eta}\right)-T_\eta^e\left(\frac{\bar\chi}{\eta}\right)\right)=\frac{8\bar\chi^2/\eta^4}{1+\exp(-2\bar\chi/\eta)}\left(\frac{\bar\chi/\eta}{1+\exp(2\bar\chi/\eta)}+\frac{1}{1+\exp(2\bar\chi)}\right) + \frac{2}{\eta}\left(\frac{2\bar\chi/\eta}{1+\exp(-2\bar\chi/\eta)}-1\right)\frac{\partial}{\partial\eta}\left(T_1^e\left(\frac{\bar\chi}{\eta}\right)-T_\eta^e\left(\frac{\bar\chi}{\eta}\right)\right). \quad \text{(A.7)}$$

As η approaches 1, the first derivative is negative:

$$\lim_{\eta\to 1}\frac{\partial}{\partial\eta}\left(T_1^e\left(\frac{\bar\chi}{\eta}\right)-T_\eta^e\left(\frac{\bar\chi}{\eta}\right)\right)=-\frac{4\bar\chi^2\exp(2\bar\chi)}{(1+\exp(2\bar\chi))^2}<0. \quad \text{(A.8)}$$

As η approaches infinity, the first derivative converges to 0. To determine whether it converges to 0 from below or from above, we scale the derivative by η^2:

$$\lim_{\eta\to+\infty}\eta^2\frac{\partial}{\partial\eta}\left(T_1^e\left(\frac{\bar\chi}{\eta}\right)-T_\eta^e\left(\frac{\bar\chi}{\eta}\right)\right)=\frac{\bar\chi(\exp(2\bar\chi)-1)}{1+\exp(2\bar\chi)}>0, \quad \text{(A.9)}$$

so that the first derivative is positive for sufficiently large η.

A.5. Proof of Theorem 1

Denote the principal's payoff under the commitment contract by $\Pi^C(\eta,c)$ and under the flexible contract by $\Pi^F(\eta,c)$.

The proof proceeds in two steps. First, we study the payoff difference $\Pi^C(\eta,c)-\Pi^F(\eta,c)$ as a function of η, keeping ratio

$$q := \frac{\eta}{c} \tag{A.10}$$

fixed. We show that there exists a unique $\hat{\eta}(q)$ at which the payoff difference crosses 0. Second, we show that function $\hat{\eta}(q)$ is decreasing and argue that it implies the theorem statement.

STEP 1. Under the commitment contract, the principal's payoff $\Pi^C(\eta,c)$ is equal to $f(T)$ from (7), with T defined as the unique solution to (9). Equation (9) can be rewritten as

$$2\exp\left(\frac{T\eta}{2}\right)\sqrt{2T\eta\pi} = \frac{\eta}{c}, \tag{A.11}$$

which, after a change of variables, implies that

$$\rho := \frac{T\eta}{2} \tag{A.12}$$

is a function of q defined in (A.10):

$$4\exp(\rho)\sqrt{\pi\rho} = q. \tag{A.13}$$

Using the new notation (A.12), we rewrite the principal's payoff as

$$\Pi^C(\eta,c) = \frac{1}{2}+\frac{1}{2}\operatorname{erf}\left(\sqrt{\frac{1}{\eta}\rho\left(\frac{\eta}{c}\right)}\right), \tag{A.14}$$

where $\rho(q)>0$ is the unique solution to (A.13).

Under the flexible contract, the principal's payoff $\Pi^F(\eta,c)$ is equal to $p(\chi)$ from (5), where χ solves (10)-(11):

$$\frac{2\eta}{c} = \frac{p(\eta\chi)}{1-p(\eta\chi)} - \frac{1-p(\eta\chi)}{p(\eta\chi)} + 2\ln\frac{p(\eta\chi)}{1-p(\eta\chi)}. \tag{A.15}$$

The expression for $p(\eta\chi)$ is given in (5), and so (A.15) can be rewritten as[30]

$$\frac{\exp(2\eta\chi)-\exp(-2\eta\chi)}{2}+2\eta\chi = \frac{\eta}{c}, \tag{A.16}$$

which, after a change of variables, implies that

$$x := 2\eta\chi \tag{A.17}$$

is a function of q defined in (A.10):

$$\frac{\exp(x)-\exp(-x)}{2}+x = q. \tag{A.18}$$

Using the new notation (A.17), we rewrite the principal's payoff as

$$\Pi^F(\eta,c) = \frac{1}{1+\exp\left(-\frac{1}{\eta}x\left(\frac{\eta}{c}\right)\right)}, \tag{A.19}$$

where $x(q)>0$ is the unique solution to (A.18).

By (A.14) and (A.19), the difference $\Pi^C(\eta,c)-\Pi^F(\eta,c)$ is a function of η, $\rho(\eta/c)$ and $x(\eta/c)$:

$$\frac{1}{2}+\frac{1}{2}\operatorname{erf}\left(\sqrt{\frac{\rho}{\eta}}\right) - \frac{1}{1+\exp\left(-\frac{x}{\eta}\right)}. \tag{A.20}$$

Function (A.20) is decreasing in x and equal to 0 at some $\hat{x}(\eta,\rho)$. Hence, the difference $\Pi^C(\eta,c)-\Pi^F(\eta,c)$ has the same sign as function $\Upsilon(\eta,\eta/c)$ defined as $\Upsilon(\eta,q) := \hat{x}(\eta,\rho(q))-x(q)$. Solving for $\hat{x}(\eta,\rho)$, we derive

$$\Upsilon(\eta,q) = \eta\ln\left(\frac{1+\operatorname{erf}\left(\sqrt{\frac{\rho(q)}{\eta}}\right)}{1-\operatorname{erf}\left(\sqrt{\frac{\rho(q)}{\eta}}\right)}\right) - x(q), \tag{A.21}$$

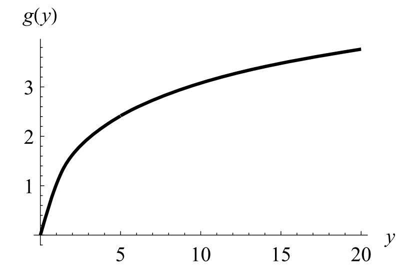
[30] Equation (A.16) is the first-order condition for the agent's optimization problem $\max_{\chi>0} p(\eta\chi)-cT^e_\eta(\chi)$, where the expression for the expected stopping time $T^e_\eta(\chi)$ is derived in Lemma 3 in Appendix A.3.

where $\rho(q)$ and $x(q)$ were defined in (A.13) and (A.18), respectively.

The partial derivative of $\Upsilon(\eta, q)$ with respect to η turns out to be a function of one variable:

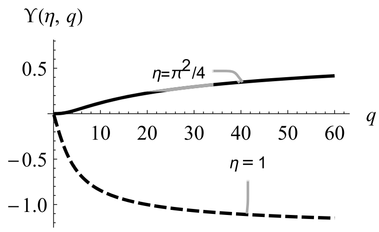
$$\frac{\partial \Upsilon(\eta, q)}{\partial \eta} = g\left(\sqrt{\frac{\rho(q)}{\eta}}\right),$$
$$g(y) = \ln\left(\frac{1+\operatorname{erf}(y)}{1-\operatorname{erf}(y)}\right) - \frac{2y\exp(-y^2)}{\sqrt{\pi}\left(1-\operatorname{erf}(y)^2\right)}.$$

Function $g(y)$ is positive for all $y > 0$, as illustrated below and formally proven in Lemma 13 in Appendix C.2.



Hence, $\Upsilon(\eta, q)$ increases in $\eta > 0$.

The graphs below illustrate that $\Upsilon(1, q)$ is negative and $\Upsilon\left(\pi^2/4, q\right)$ is positive for all $q > 0$; the formal proof is in Lemmas 16 and 17 in Appendix C.2. Hence, function $\Upsilon(\eta, q)$ crosses 0 at the unique $\eta = \hat{\eta}(q)$, which lies between 1 and $\pi^2/4$.

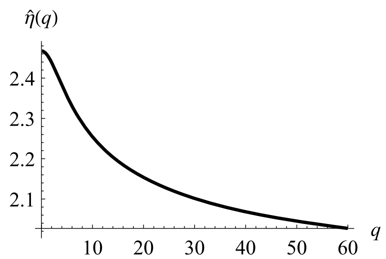


Since function $\Upsilon(\eta, \eta/c)$ crosses 0 from below, it has the same sign as function

$$\xi(\eta, c) := \eta - \hat{\eta}\left(\frac{\eta}{c}\right). \tag{A.22}$$

<u>*STEP 2.*</u> The graph below illustrates that function $\hat{\eta}(q)$ is decreasing; the formal proof is in Lemma 20 in Appendix C.2. Hence, function $\xi(\eta, c)$ is increasing in $\eta > 0$:

$$\frac{\partial \xi(\eta, c)}{\partial \eta} = 1 - \frac{1}{c}\hat{\eta}'\left(\frac{\eta}{c}\right) > 0.$$



Since $1 < \hat{\eta}(q) < \frac{\pi^2}{4}$, function $\xi(\eta, c)$ is negative at $\eta = 1$ and positive at $\eta = \frac{\pi^2}{4}$. Thus, there exists a threshold $1 < \eta^*(c) < \frac{\pi^2}{4}$ such that $\xi(\eta, c)$ is negative for all $\eta < \eta^*(c)$ and positive for all $\eta > \eta^*(c)$.

Threshold $\eta^*(c)$ is increasing in c because, by Lemma 20, function $\hat{\eta}(q)$ is decreasing:

$$\frac{\mathrm{d}\,\eta^*(c)}{\mathrm{d}\,c} = -\left.\frac{\frac{\partial \xi(\eta,c)}{\partial(c)}}{\frac{\partial \xi(\eta,c)}{\partial \eta}}\right|_{\eta=\eta^*(c)} = \frac{-\eta^*(c)\,\hat{\eta}'\left(\frac{\eta^*(c)}{c}\right)}{c^2 - c\hat{\eta}'\left(\frac{\eta^*(c)}{c}\right)} > 0.$$

A.6. Proof of Proposition 4

The principal's payoff is equal to $f(T)$, where function $f(t)$ is defined in (7) and $T > 0$ uniquely solves (9). Since function $f(t)$ is increasing, the principal's payoff increases if and only if T increases.

Treating T, the solution to (9), as a function of the level of overconfidence η, we apply the implicit function theorem to (9):

$$T'(\eta) = -\frac{f'(\eta T) + \eta T f''(\eta T)}{\eta^2 f''(\eta T)}. \tag{A.23}$$

Given the definition (7) of $f(t)$, the denominator of (A.23) is negative because f is concave, and the numerator of (A.23) becomes

$$f'(\eta T) + \eta T f''(\eta T) = \frac{1 - \eta T}{4\sqrt{2\pi \eta T}} \exp\left(-\frac{\eta T}{2}\right). \tag{A.24}$$

Thus, $T'(\eta)$ has the same sign as $g(\eta) := 1 - \eta T(\eta)$.

Function $g(\eta) = 0$ whenever $\eta T(\eta) = 1$. Substituting $\eta T(\eta) = 1$ into the argument of function f' in equation (9) yields that $g\left(\eta_C^*\right) = 0$ at point $\eta_C^* = \frac{c}{f'(1)}$. Point $\eta_C^* = \frac{c}{f'(1)}$ becomes (14) after we substitute $f'(1)$ obtained from the definition (7) of $f(t)$. Because $T'(\eta) = 0$ whenever $g(\eta) = 0$ and

$$g'(\eta)\big|_{g(\eta)=0} = -T(\eta) - \eta T'(\eta)\big|_{T'(\eta)=0} < 0, \tag{A.25}$$

function $g(\eta)$ crosses 0 only once, and from above.

A.7. Proof of Proposition 5

The principal's payoff is derived in (A.19), where $x(q) > 0$ is the unique solution to (A.18). Taking the first derivative with respect to η, we get

$$\frac{\partial \Pi^F(\eta, c)}{\partial \eta} = \frac{\exp\left(\frac{1}{\eta} x\left(\frac{\eta}{c}\right)\right)}{\left(1 + \exp\left(\frac{1}{\eta} x\left(\frac{\eta}{c}\right)\right)\right)^2 \eta^2} \left(\frac{\eta}{c} x'\left(\frac{\eta}{c}\right) - x\left(\frac{\eta}{c}\right)\right), \tag{A.26}$$

and so, its sign coincides with the sign of $h(\eta/c)$, where

$$h(q) := q x'(q) - x(q). \tag{A.27}$$

Applying the implicit function theorem to (A.18), we get

$$x'(q) = \frac{2\exp(x(q))}{(1 + \exp(x(q)))^2}. \tag{A.28}$$

Substituting $x'(q)$ from (A.28) and q from (A.18) into (A.27), we get

$$h(q) = -\frac{g(x(q))}{(1 + \exp(x(q)))^2}, \tag{A.29}$$

where function $g(x)$ is defined as

$$g(x) := 1 + x + \exp(2x)(x - 1). \tag{A.30}$$

Function $g(x)$ is convex, as $g''(x) = 4x\exp(2x) > 0$. Hence, the derivative of g is increasing. Since $g'(0) = 0$, $g'(x)$ is positive for $x > 0$. Consequently, g is increasing from $g(0) = 0$. Hence, $g(x) > 0$ for all $x > 0$. It follows that $h(\eta/c) < 0$ for all $\eta > 0$ and all $c > 0$, and so, the principal's payoff $\Pi^F(\eta, c)$ is decreasing in $\eta > 0$.

A.8. Proof of Theorem 2

By Proposition 5, under the flexible contract, the principal's payoff is decreasing in $\eta \geq \underline{\eta}$, and, thus, it achieves its maximum at $\eta = \underline{\eta}$. By Proposition 4, under the commitment contract, the principal's payoff increases for all $\eta < \eta_C^*$ and decreases for all $\eta > \eta_C^*$, and so, on $\eta \geq \underline{\eta}$, it achieves its maximum at $\eta = \max\{\underline{\eta}, \eta_C^*\}$.

<u>*REGION* $\eta_C^* \leq 1$.</u> If $\eta_C^* \leq 1$, then, under each contract, the principal's payoff is decreasing for all $\eta \geq 1$. Thus, the principal optimally selects an agent with some $\eta \leq 1$. By Theorem 1, when $\eta \leq 1$, then the principal's payoff is higher under the flexible contract, which means that this contract is optimal for the principal. Hence, if $\eta_C^* \leq 1$, then the principal selects the agent with $\eta = \underline{\eta}$ and offers him the flexible contract.

By (14), condition $\eta_C^* \leq 1$ is equivalent to $c \leq \bar{c}$, where

$$\bar{c} := \frac{1}{2\sqrt{2e\pi}}. \tag{A.31}$$

REGION $\eta_C^* > 1$. If $\eta_C^* > 1$, then the principal's payoff is maximized either at $\eta = \underline{\eta}$ with the flexible contract, in which case she gets $\Pi^F(\underline{\eta}, c)$, or at $\eta = \eta_C^*$ with the commitment contract, in which case she gets $\Pi^C\left(\eta_C^*, c\right)$. Payoffs $\Pi^F(\eta, c)$ and $\Pi^C(\eta, c)$ are calculated in (A.19) and in (A.14), respectively.

Condition $\eta_C^* > 1$ is equivalent to $c > \bar{c}$, and so, to prove the theorem, it is sufficient to show that there exists $c^* > \bar{c}$ such that $\Pi^F(\underline{\eta}, c) > \Pi^C\left(\eta_C^*, c\right)$ for all $\bar{c} \le c < c^*$ and $\Pi^F(\underline{\eta}, c) < \Pi^C\left(\eta_C^*, c\right)$ for all $c > c^*$.

By (A.14),

$$\Pi^C\left(\eta_C^*, c\right) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\sqrt{\frac{1}{\eta_C^*}\rho\left(\frac{\eta_C^*}{c}\right)}\right), \tag{A.32}$$

where $\rho(q) > 0$ is the unique solution to (A.13). The solution to equation (A.13) is explicit at $q = \eta_C^*/c$ because $q = 2\sqrt{2e\pi}$ by (14); thus, $\rho\left(\eta_C^*/c\right) = 1/2$. Substituting $\rho\left(\eta_C^*/c\right) = 1/2$ and η_C^* from (14) into (A.32), we get

$$\Pi^C\left(\eta_C^*, c\right) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\sqrt{\frac{1}{4c\sqrt{2e\pi}}}\right). \tag{A.33}$$

By (A.19),

$$\Pi^F\left(\underline{\eta}, c\right) = \frac{1}{1 + \exp\left(-\frac{1}{\underline{\eta}}x\left(\frac{\underline{\eta}}{c}\right)\right)}, \tag{A.34}$$

where $x(q) > 0$ is the unique solution to (A.18); that is, $x\left(\underline{\eta}/c\right) > 0$ solves

$$\frac{\exp(x) - \exp(-x)}{2} + x = \frac{\underline{\eta}}{c}. \tag{A.35}$$

Expressing $x\left(\underline{\eta}/c\right)$ from (A.34) and then substituting x in (A.35) with this expression, we get c as a function of $\Pi^F\left(\underline{\eta}, c\right)$ and $\underline{\eta}$:

$$c = \frac{1}{s\left(\Pi^F\left(\underline{\eta}, c\right), \underline{\eta}\right)}, \tag{A.36}$$

where

$$s(y, \underline{\eta}) := \frac{1}{2\underline{\eta}}\left(\left(\frac{y}{1-y}\right)^{\underline{\eta}} - \left(\frac{1-y}{y}\right)^{\underline{\eta}}\right) + \ln\frac{y}{1-y}. \tag{A.37}$$

Substituting c in (A.33) with (A.36), we express the payoff difference as a function of $\Pi^F\left(\underline{\eta}, c\right)$ and $\underline{\eta}$:

$$\Pi^C\left(\eta_C^*, c\right) - \Pi^F\left(\underline{\eta}, c\right) = g\left(\Pi^F\left(\underline{\eta}, c\right), \underline{\eta}\right), \tag{A.38}$$

where

$$g\left(y, \underline{\eta}\right) := \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\sqrt{\frac{s(y, \underline{\eta})}{4\sqrt{2e\pi}}}\right) - y. \tag{A.39}$$

In light of (A.38), we conclude that to show that there exists $c^* > \bar{c}$ such that $\Pi^F(\underline{\eta}, c) > \Pi^C\left(\eta_C^*, c\right)$ for all $\bar{c} \le c < c^*$ and $\Pi^F(\underline{\eta}, c) < \Pi^C\left(\eta_C^*, c\right)$ for all $c > c^*$, it is sufficient to show that there exists $c^* > \bar{c}$ such that $g\left(\Pi^F\left(\underline{\eta}, c\right), \underline{\eta}\right) < 0$ for all $\bar{c} \le c < c^*$ and $g\left(\Pi^F\left(\underline{\eta}, c\right), \underline{\eta}\right) > 0$ for all $c > c^*$.

Argument $y = \Pi^F\left(\underline{\eta}, c\right)$ of function $g(y, \underline{\eta})$ belongs to the region $(0.5, 1)$ because expression (A.34) is increasing in $x(\underline{\eta}/c) > 0$ from 0.5 to 1. Define

$$\bar{y}(\underline{\eta}) = \Pi^F\left(\underline{\eta}, \bar{c}\right), \tag{A.40}$$

or equivalently, by (A.36), $\bar{y}(\underline{\eta})$ solves

$$s\left(\bar{y}, \underline{\eta}\right) = \frac{1}{\bar{c}}. \tag{A.41}$$

Since $s(y, \underline{\eta})$ is increasing in $y \in (0.5, 1)$ from 0 to $+\infty$, equation (A.41) uniquely defines $\bar{y} \in (0.5, 1)$. Moreover, by direct computations,

$$s\left(\frac{1}{1+0.05^{1/\underline{\eta}}},\underline{\eta}\right)\overset{(A.37)}{=}\frac{1}{2\underline{\eta}}\left(\frac{1}{0.05}-0.05-2\ln 0.05\right)>\frac{2\sqrt{2e\pi}}{\underline{\eta}}\overset{(A.31)}{=}\frac{1}{\bar{c}\underline{\eta}}\overset{\underline{\eta}\leq 1}{\geq}\frac{1}{\bar{c}},\tag{A.42}$$

which implies that

$$\bar{y}(\underline{\eta})<\frac{1}{1+0.05^{1/\underline{\eta}}}\tag{A.43}$$

for all $\underline{\eta}\in(0,1]$. Expression (A.34) for $\Pi^F\left(\underline{\eta},c\right)$ is decreasing in c because $x\left(\underline{\eta}/c\right)$, defined as a solution to (A.35), is decreasing in c. Hence, condition $c>\bar{c}$ is equivalent to $\Pi^F\left(\underline{\eta},c\right)<\bar{y}$, and so, the region to which the argument $y=\Pi^F\left(\underline{\eta},c\right)$ of function $g(y,\underline{\eta})$ belongs can be narrowed down to $(0.5,\bar{y})$. Moreover, since $\Pi^F\left(\underline{\eta},c\right)$ is decreasing in c, condition $c<c^*$ translates to $y>y^*$ for $y^*=\Pi^F\left(\underline{\eta},c^*\right)$. Thus, to show that there exists $c^*>\bar{c}$ such that $g\left(\Pi^F\left(\underline{\eta},c\right),\underline{\eta}\right)<0$ for all $\bar{c}\leq c<c^*$ and $g\left(\Pi^F\left(\underline{\eta},c\right),\underline{\eta}\right)>0$ for all $c>c^*$, it is sufficient to show that there exists $y^*\in(0.5,\bar{y})$ such that $g(y,\underline{\eta})>0$ for all $y\in(0.5,y^*)$ and $g(y,\underline{\eta})<0$ for all $y\in(y^*,\bar{y})$.

The sign of the second derivative of $g(y,\underline{\eta})$ with respect to y coincides with the sign of function $h(y,\underline{\eta})$ defined as

$$h(y,\underline{\eta}):=\left(\frac{16\sqrt{e\pi}l\left((1+l)(2y-1)+\underline{\eta}(1-l)\right)}{(1+l)^3}-\sqrt{2}\right)s(y,\underline{\eta})-4\sqrt{e\pi},\quad\text{where}\quad l\equiv\left(\frac{1-y}{y}\right)^{\underline{\eta}},\tag{A.44}$$

because

$$\frac{\partial^2 g(y,\underline{\eta})}{\partial y^2}=\frac{\exp\left(-\frac{3}{4}-\frac{s(y,\underline{\eta})}{4\sqrt{2e\pi}}\right)\left(\left(\frac{y}{1-y}\right)^{\underline{\eta}}+1\right)^4\left(\frac{1-y}{y}\right)^{2\underline{\eta}}h(y,\underline{\eta})}{2^{29/4}\pi^{5/4}(1-y)^2y^2s(y,\underline{\eta})^{3/2}}.\tag{A.45}$$

Consider the expression

$$\frac{\partial^3 h(y,\underline{\eta})}{\partial y^3}-\frac{2(2y-1)}{(1-y)y}\frac{\partial^2 h(y,\underline{\eta})}{\partial y^2}=-\frac{1}{(1-y)^3y^3}\left\{\frac{16\sqrt{e\pi}\underline{\eta}l}{(1+l)^4}h_1(l)+\frac{\underline{\eta}^3}{l(1+l)^6\sqrt{2}}h_2(l)\right.$$
$$\left.+\frac{1}{\sqrt{2}l}\left(2(1-y)y(1+l)^2+\underline{\eta}(1-l^2)(2y-1)+(1+l^2)\underline{\eta}^2(1-\underline{\eta})\right)\right\},\tag{A.46}$$

where

$$h_1(l)=3(1-l^2)+\left(1-4l+l^2\right)\ln l,\tag{A.47}$$

$$h_2(l)=\left(1+l^2\right)(1+l)^6-16\sqrt{2e\pi}l^2\left(5\left(1-l^2\right)\left(1-10l+l^2\right)+\left(1-26l+66l^2-26l^3+l^4\right)\ln l\right)\tag{A.48}$$

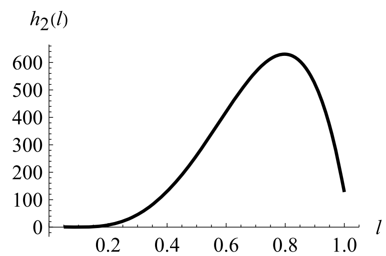
and notation l is reserved for the ratio

$$l\equiv\left(\frac{1-y}{y}\right)^{\underline{\eta}}.\tag{A.49}$$

We argue that expression (A.46) is negative for all $0.5<y<\frac{1}{1+0.05^{1/\underline{\eta}}}$, $0<\underline{\eta}\leq 1$. First, given the restrictions $0.5<y<\frac{1}{1+0.05^{1/\underline{\eta}}}$ and $\underline{\eta}>0$, the ratio l belongs to the interval $(0.05,1)$. Second, function $h_1(l)$ is positive for all $l\in(0.05,1)$ because it's concave:

$$h_1''(l)=-\frac{(1+l)(1+3l)}{l^2}+2\ln l<0\tag{A.50}$$

for all $l\in(0,1)$, takes a positive value at $l=0.05$, $h_1(0.05)>0$, and equal to 0 at $l=1$, $h_1(1)=0$. Finally, function $h_2(l)$ is positive for all $l\in(0.05,1)$, which is illustrated on the graph below and formally proved in Appendix C.3.



For $0.5 < y < \frac{1}{1+0.05^{1/\underline{\eta}}}$, since expression (A.46) is negative, we conclude that $\frac{\partial^2 h(y,\underline{\eta})}{\partial y^2}$ may cross 0 only from above; thus, $\frac{\partial h(y,\underline{\eta})}{\partial y}$ is increasing and then decreasing; thus, $h(y,\underline{\eta})$ is decreasing, increasing and then decreasing. More precisely, there exist $0.5 \le \hat{y}_1 \le \hat{y}_2 \le \frac{1}{1+0.05^{1/\underline{\eta}}}$ such that $h(y,\underline{\eta})$ is decreasing on $y \in (0.5, \hat{y}_1)$, increasing on $y \in (\hat{y}_1, \hat{y}_2)$, and then decreasing on $y \in \left(\hat{y}_2, \frac{1}{1+0.05^{1/\underline{\eta}}}\right)$. By direct computations, $h(0.5,\underline{\eta}) = -4\sqrt{e\pi} < 0$ and

$$h\left(\frac{1}{1+0.05^{1/\underline{\eta}}}, \underline{\eta}\right) = C_1 + \frac{C_2}{\underline{\eta}}\left(\underbrace{\frac{1}{1+0.05^{1/\underline{\eta}}} - \frac{1}{1+0.05}}_{\ge 0} + C_3\right) > 0, \tag{A.51}$$

where C_1, C_2 and C_3 are some positive constants ($C_1 \approx 13$, $C_2 \approx 55$, $C_3 \approx 0.12$). Hence, there exists $0.5 < \hat{y}_3 < \frac{1}{1+0.05^{1/\underline{\eta}}}$ such that $h(y,\underline{\eta}) < 0$ for all $y \in (0.5, \hat{y}_3)$ and $h(y,\underline{\eta}) > 0$ for all $y \in \left(\hat{y}_3, \frac{1}{1+0.05^{1/\underline{\eta}}}\right)$.

Given the connection (A.45) between functions $h(y,\underline{\eta})$ and $g(y,\underline{\eta})$, we conclude that $\frac{\partial g(y,\underline{\eta})}{\partial y}$ is decreasing for all $y \in (0.5, \hat{y}_3)$ and increasing for all $y \in \left(\hat{y}_3, \frac{1}{1+0.05^{1/\underline{\eta}}}\right)$. By direct computations, $\frac{\partial g(y,\underline{\eta})}{\partial y}$ goes to $+\infty$ as $y \to 0.5$. Thus, there exist $0.5 < \hat{y}_4 \le \hat{y}_5 \le \frac{1}{1+0.05^{1/\underline{\eta}}}$ such that $g(y,\underline{\eta})$ is increasing on $(0.5, \hat{y}_4)$, decreasing on $(\hat{y}_4, \hat{y}_5)$, and then increasing on $\left(\hat{y}_5, \frac{1}{1+0.05^{1/\underline{\eta}}}\right)$. By direct computations, $g(0.5,\underline{\eta}) = 0$. By (A.43), $\bar{y}$ is less than $\frac{1}{1+0.05^{1/\underline{\eta}}}$. At $y = \bar{y}$, the value of $g(y,\underline{\eta})$ is negative because by (A.38) and (A.40), $g(\bar{y},\underline{\eta})$ is equal to the difference $\Pi^C(\eta_C^*, \bar{c}) - \Pi^F(\underline{\eta}, \bar{c})$, and given our analysis in region $\eta_C^* \le 1$, $\Pi^F(\underline{\eta}, c) > \Pi^C(\eta_C^*, c)$ for $c = \bar{c}$. Thus, we conclude that there exists $y^* \in (0.5, \bar{y})$ such that $g(y,\underline{\eta}) > 0$ for all $y \in (0.5, y^*)$ and $g(y,\underline{\eta}) < 0$ for all $y \in (y^*, \bar{y})$.

Appendices B and C. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2024.01.001.

References

Baumeister, R.F., Smart, L., Boden, J.M., 1996. Relation of threatened egotism to violence and aggression: the dark side of high self-esteem. Psychol. Rev. 103, 5–33. https://doi.org/10.1037/0033-295X.103.1.5.

Ben-David, I., Graham, J.R., Harvey, C.R., 2013. Managerial miscalibration. Q. J. Econ. 128, 1547–1584. https://doi.org/10.1093/qje/qjt023.

Benabou, R., Tirole, J., 2002. Self-confidence and personal motivation. Q. J. Econ. 117, 871–915. https://doi.org/10.1162/003355302760193913.

Benos, A.V., 1998. Aggressiveness and survival of overconfident traders. J. Financ. Mark. 1, 353–383. https://doi.org/10.1016/S1386-4181(97)00010-4.

Bernardo, A.E., Welch, I., 2001. On the evolution of overconfidence and entrepreneurs. J. Econ. Manag. Strategy 10, 301–330. https://doi.org/10.1111/j.1430-9134.2001.00301.x.

Berner, E.S., Graber, M.L., 2008. Overconfidence as a cause of diagnostic error in medicine. Am. J. Med. 121, S2–S23. https://doi.org/10.1016/j.amjmed.2008.01.001.

Chow, E.H., Lin, H.M., Lin, Y.M., Weng, Y.C., 2011. The performance of overconfident fund managers. Emerg. Mark. Financ. Trade 47, 21–30. https://doi.org/10.2753/REE1540-496X470202.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. J. Finance 53, 1839–1885. https://doi.org/10.1515/9781400829125-016.

Darling, D.A., Siegert, A.J.F., 1953. The first passage problem for a continuous Markov process. Ann. Math. Stat. 24, 624–639. https://www.jstor.org/stable/2236785.

Eigen, Z.J., Listokin, Y., 2012. Do lawyers really believe their own hype, and should they? A natural experiment. J. Leg. Stud. 41, 239–267. https://doi.org/10.1086/667711.

Gervais, S., Heaton, J.B., Odean, T., 2011. Overconfidence, compensation contracts, and capital budgeting source. J. Finance 66, 1735–1777. https://doi.org/10.1111/j.1540-6261.2011.01686.x.

Goel, A.M., Thakor, A.V., 2008. Overconfidence, CEO selection, and corporate governance. J. Finance 63, 2737–2784. https://doi.org/10.1111/j.1540-6261.2008.01412.x.

Goodman-Delahunty, J., Granhag, P.A., Hartwig, M., Loftus, E.F., 2010. Insightful or wishful: lawyers' ability to predict case outcomes. Psychol. Public Policy Law 16, 133–157. https://doi.org/10.1037/a0019060.

Gort, C., Wang, M., Siegrist, M., 2008. Are pension fund managers overconfident? J. Behav. Finance 9, 163–170. https://doi.org/10.1080/15427560802341616.

Heidhues, P., Kőszegi, B., Strack, P., 2018. Unrealistic expectations and misguided learning. Econometrica 86, 1159–1214. https://doi.org/10.3982/ECTA14084.

Kahneman, D., Tversky, A., 1972. Subjective probability: a judgment of representativeness. Cogn. Psychol. 3, 430–454. https://doi.org/10.1016/0010-0285(72)90016-3.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., Mullainathan, S., 2018. Human decisions and machine predictions. Q. J. Econ. 133, 237–293. https://doi.org/10.1093/qje/qjx032.

Kőszegi, B., 2014. Behavioral contract theory. J. Econ. Lit. 52, 1075–1118. https://doi.org/10.1257/jel.52.4.1075.

Kyle, A.S., Wang, F.A., 1997. Speculation duopoly with agreement to disagree: can overconfidence survive the market test? J. Finance 52, 2073–2090. https://doi.org/10.1111/j.1540-6261.1997.tb02751.x.

Loftus, E.F., Wagenaar, W.A., Loftus, E.F., Wagenaarf, W.A., 1988. Lawyers' predictions of success. Jurimetrics 28, 437–453. https://www.jstor.org/stable/29762095.

López-Pérez, R., Rodriguez-Moral, A., Vorsatz, M., 2021. Simplified mental representations as a cause of overprecision. J. Behav. Exp. Econ. 92, 101681. https://doi.org/10.1016/j.socec.2021.101681.

Malmendier, U., Tate, G., 2009. Superstar CEOs. Q. J. Econ. 124, 1593–1638. https://doi.org/10.1162/qjec.2009.124.4.1593.

Moore, D.A., Healy, P.J., 2008. The trouble with overconfidence. Psychol. Rev. 115, 502–517. https://doi.org/10.1037/0033-295X.115.2.502.

Moore, D.A., Tenney, E.R., Haran, U., 2015. Overprecision in judgment, chapter 6. In: Keren, G., Wu, G. (Eds.), Wiley Blackwell Handbook of Judgment and Decision Making, II. John Wiley & Sons, Ltd, Chichester, UK, pp. 182–209.

Odean, T., 1998. Volume, volatility, price, and profit when all traders are above average. J. Finance 53, 1887–1934. https://doi.org/10.1111/0022-1082.00078.
Ortoleva, P., Snowberg, E., 2015. Overconfidence in political behavior. Am. Econ. Rev. 105, 504–535. https://doi.org/10.1257/aer.20130921.
Puetz, A., Ruenzi, S., 2011. Overconfidence among professional investors: evidence from mutual fund managers. J. Bus. Finance Account. 38, 684–712. https://doi.org/10.1111/j.1468-5957.2010.02237.x.
Ramachandran, V.S., Blakeslee, S., 1998. Phantoms in the Brain: Probing the Mysteries of the Human Mind. Morrow/HarperCollins.
de la Rosa, L.E., 2011. Overconfidence and moral hazard. Games Econ. Behav. 73, 429–451. https://doi.org/10.1016/j.geb.2011.04.001.
Santos-Pinto, L., 2008. Positive self-image and incentives in organisations. Econ. J. 118, 1315–1332. https://doi.org/10.1111/j.1468-0297.2008.02171.x.
Scheinkman, J.A., Xiong, W., 2003. Overconfidence and speculative bubbles. J. Polit. Econ. 111, 1183–1220. https://doi.org/10.1086/378531.
Shiryaev, A.N., 1978. Optimal Stopping Rules. Stochastic Modelling and Applied Probability, vol. 8. Springer, New York, New York, NY.
Tversky, A., Kahneman, D., 1971. Belief in the law of small numbers. Psychol. Bull. 76, 105–110. https://doi.org/10.1037/h0031322.

Additional references for Appendices B and C

Baricz, A., 2008. Mills' ratio: monotonicity patterns and functional inequalities. J. Math. Anal. Appl. 340, 1362–1370. https://doi.org/10.1016/j.jmaa.2007.09.063.
Birnbaum, Z.W., 1942. An inequality for Mill's ratio. Ann. Math. Stat. 13, 245–246. https://doi.org/10.1214/aoms/1177731611.
Chu, J.T., 1955. On bounds for the normal integral. Biometrika 42, 263–265. https://doi.org/10.2307/2333443.
Komatu, Y., 1955. Elementary inequalities for Mills' ratio. Rep. Stat. Appl. Res. UJSE 4, 69–70.
Neuman, E., 2013. Inequalities and bounds for the incomplete gamma function. Results Math. 63, 1209–1214. https://doi.org/10.1007/s00025-012-0263-9.
Revuz, D., Yor, M., 1999. Continuous Martingales and Brownian Motion, 3 ed. Grundlehren der Mathematischen Wissenschaften, vol. 293. Springer Berlin Heidelberg, Berlin, Heidelberg.
Sampford, M.R., 1953. Some inequalities on Mill's ratio and related functions. Ann. Math. Stat. 24, 130–132. https://doi.org/10.1214/aoms/1177729093.
Szarek, S.J., Werner, E., 1999. A nonsymmetric correlation inequality for Gaussian measure. J. Multivar. Anal. 68, 193–211. https://doi.org/10.1006/jmva.1998.1784.
Tikochinsky, Y., Tishby, N.Z., Levine, R.D., 1984. Alternative approach to maximum-entropy inference. Phys. Rev. A 30, 2638–2644. https://doi.org/10.1103/PhysRevA.30.2638.