

DISCUSSION PAPER SERIES

DP12740

THE CULTURE OF OVERCONFIDENCE

Venkataraman Bhaskar and Caroline Thomas

INDUSTRIAL ORGANIZATION



THE CULTURE OF OVERCONFIDENCE

Venkataraman Bhaskar and Caroline Thomas

Discussion Paper DP12740
Published 20 February 2018
Submitted 20 February 2018

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INDUSTRIAL ORGANIZATION**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Venkataraman Bhaskar and Caroline Thomas

THE CULTURE OF OVERCONFIDENCE

Abstract

Why do political leaders or managers persist with their pet projects and policies despite bad news? When project continuation is a more informative experiment than project termination, a reputationally concerned leader is biased towards continuation, as it enables her to disclose her private information. Perceived overconfidence on the part of the leader aggravates this tendency, even when the leader is not, in fact, overconfident. Higher-order beliefs regarding overconfidence can induce inefficient equilibrium selection even when it is "almost common knowledge" that the leader is not overconfident. Thus, a culture where leaders are expected to be overconfident can have undesirable effects even upon leaders who have correct beliefs.

JEL Classification: C73, D82, D72

Keywords: overconfidence, Policy persistence, Mis-specified models, Non-common priors, Higher-order beliefs

Venkataraman Bhaskar - v.bhaskar@austin.utexas.edu
University of Texas at Austin and CEPR

Caroline Thomas - caroline.thomas@austin.utexas.edu
University of Texas at Austin

The Culture of Overconfidence*

V Bhaskar[†] Caroline Thomas[‡]

February 17, 2018

Abstract

Why do political leaders or managers persist with their pet projects and policies despite bad news? When project continuation is a more informative experiment than project termination, a reputationally concerned leader is biased towards continuation, as it enables her to disclose her private information. Perceived overconfidence on the part of the leader aggravates this tendency, even when the leader is not, in fact, overconfident. Higher-order beliefs regarding overconfidence can induce inefficient equilibrium selection even when it is “almost common knowledge” that the leader is not overconfident. Thus, a culture where leaders are expected to be overconfident can have undesirable effects even upon leaders who have correct beliefs.

JEL codes: C73, D82 , D72

Keywords: Overconfidence, Policy persistence, Mis-specified models, Non-common priors, Higher-order beliefs.

*We are grateful to Drew Fudenberg, François Lagarde, Stephen Morris and Muhamet Yildiz for helpful comments. Bhaskar thanks the National Science Foundation for its support via grant 201503942.

[†]Department of Economics, University of Texas at Austin. v.bhaskar@austin.utexas.edu.

[‡]Department of Economics, University of Texas at Austin. caroline.thomas@austin.utexas.edu.

To those waiting with bated breath for the ‘U-turn’, I have only one thing to say:

‘You turn, if you want to. The lady’s not for turning.’

Margaret Thatcher, 10 October 1980.

When the facts change, I change my mind. What do you do, Sir?

Attributed to John Maynard Keynes.

1 Introduction

Why do managers and political leaders persist with their pet projects when faced with negative evidence? Margaret Thatcher’s riposte , “the lady is not for turning”, to the critics of her monetarist policy, made clear that record unemployment of over two million was not going to make her change course. Later, Thatcher persisted with the poll tax when confronted by widespread opposition, a stubbornness that led to her being replaced as Prime Minister by her own party. Mao Tse-Tung intensified the Great Leap Forward, despite reports of widespread starvation in rural China, with over 30 million people dying in the consequent famine (see Dikötter (2010)).

A popular explanation, going back ancient Greece, is *hubris*, or overconfidence. In Aeschylus’ *The Persians* and Sophocles’ *Antigone*, the protagonist kings, Xerxes and Creon, are consumed with ambition and arrogance. They ignore advice and omens, and persist in their chosen paths, precipitating their demise. Modern-day historians and political scientists invoke hubris to explain Napoleon’s doomed march on Moscow (Kroll, Toombs, and Wright (2000)), Hitler’s over-ambitious expansion (Kershaw (2000)), Thatcher’s failure to reverse the poll tax (Owen and Davidson (2009)), and Mao’s persistence with the Great Leap Forward (Dikötter (2010)). Owen and Davidson (2009) argue that “hubris syndrome” afflicts many prominent politicians, and provide a detailed portrait of several US presidents and British prime ministers to support their claim. Hubris has similarly been used to explain the behavior of CEOs and their tendency to persist with bad decisions. Roll (1986), Morck, Shleifer, and Vishny (1990), Malmendier and Tate (2008), and Malmendier and Tate (2005) argue that managerial overconfidence is a major factor behind corporate takeovers, and explains the adverse effects of takeovers on the acquiring firm’s value.

This paper shows that reputational concerns can induce a leader to continue an ineffectual project, even when she is not overconfident and has correct beliefs regarding the quality of the project. Our main contribution is to show that overconfidence, or even perceptions of overconfidence, magnify the resulting inefficiencies. We conclude that a culture where leaders

are expected to be overconfident can have adverse consequences even when the leader in question is not, in fact, overconfident.

We begin with a rational explanation for a leader's obstinacy. We set out a model where she has social as well as reputational concerns, with the novelty that different actions differentially affect the disclosure of public information. Because even the expertise of a Keynes does not preclude making a choice that turns out badly, changing one's mind should not be a sure sign of incompetence. In keeping with this requirement, we shall assume that even the best leaders are fallible. The decision maker (DM) — the manager of a firm, or a political leader — receives (inconclusive) information regarding the viability of a project that she has initiated, and must decide whether to continue the project to its conclusion, or to abandon it. The information received is private, and so unobserved by outsiders who evaluate the DM's performance.¹ Abandoning the project stops the arrival of further information, while continuing publicly reveals the underlying state of the world, resolving all uncertainty.

Our explanation hinges on the *disclosure value of continuation*.² By continuing a project to its conclusion, a leader publicly verifies her private belief about its quality: if she believes that a project has a probability μ of being good, the project will result in a public success with probability μ and a public failure with probability $1 - \mu$. By terminating the project, she prevents any further learning, and only reveals that she thought the project bad enough to warrant termination. Thus, she faces a reputational cost when she is moderately pessimistic about the project, inducing her to continue projects that are socially wasteful. Furthermore, there are strategic complementarities between the DM's continuation decision and the outside observer's inference; if the DM is more stubborn and continues the project at worse beliefs, the observer's inference is more adverse in the event that she terminates the project, increasing the reputational penalty for stopping. Consequently, there can be multiple equilibria, and an equilibrium where the project is stopped more often (i.e. at a higher belief) is more efficient. Thus, the equilibrium where the DM continues the project least often is the most efficient one.

We go on to examine the implications of the DM's overconfidence, in particular the interaction between her private information and overconfidence. We assume that the DM and outside observer have different prior beliefs on the project's chances of success; the DM

¹It is straightforward to extend our analysis to allow for public information about the project, as long as the DM also gets private information.

²The disclosure value pertains to risky options, and arises for much the same reasons as the familiar option value attached to experimenting with risky options: the accrual of information about an uncertain payoff. But it has very different economic consequences.

has a prior, q , that is strictly larger than the observer's prior, p .

Because continuing the project results in more information revelation, it becomes even more appealing for an overconfident DM. There exist signals after which she continues the project although she would stop if she had the same prior belief as the outside observer. Moreover, there is a second, more subtle effect, that induces excessive project continuation. Since the observer believes that the DM is overconfident, he draws even more negative inferences about her private information in the event that she terminates the project. This raises the reputational cost of stopping the project. Indeed, this reputational cost obtains even when the DM is not overconfident, but is merely *perceived* to be so, since the negative inference from project termination is identical to that drawn when she is, in fact, overconfident. In summary, perceived overconfidence, in conjunction with the DM's private information, induces a greater tendency towards stubbornness, exacerbating inefficiencies.

One may take this argument further. Suppose that the DM is not overconfident, and is also not perceived to be overconfident. But she believes that she is perceived to be overconfident. More generally, there may be mutual knowledge that the DM is *not* overconfident up to many levels, but not common knowledge: at some level, there is a perception that the DM is (believed to be) overconfident. How is the situation to be analyzed? We provide an answer to this methodological question. Equilibria in such a game are anchored to equilibria in the game where the DM is *in fact* overconfident. Consequently, the DM's tendency to continue projects is exacerbated, relative to the situation where it is commonly known that she is not overconfident. However, the higher the level of mutual knowledge that she is not overconfident, the less pronounced is this effect.

We then consider the limiting case, where there are arbitrarily high levels of mutual knowledge that the DM is not overconfident, so that it is "almost common knowledge" that the DM is not overconfident. We show that the equilibrium in the game with overconfidence plays a powerful role in determining limit outcomes. If possible overconfidence is large enough, then there is a unique limit equilibrium — the most inefficient equilibrium of the game with (common knowledge of) a common prior.³

These results show that a culture where business and political leaders are expected to be substantially overconfident may have pernicious effects even for leaders who are not overconfident. Indeed, it can be mutual knowledge to a high degree that the leader in question has the right beliefs. Nonetheless, the lack of common knowledge entailed by the

³Indeed, this is the unique rationalizable outcome of that game. We also show that even small possible overconfidence restricts equilibria.

culture ensures that the most inefficient equilibrium is selected.

Our theory highlights a novel and important channel through which cultural stereotypes might determine outcomes: via higher-order beliefs. A culture where leaders are expected to be overconfident plays a powerful role even when a leader is not in fact overconfident. This can be contrasted with a different view of the role of culture, namely that it coordinates expectations, and thereby selects among equilibria, much as history does.⁴

One possible normative implication of our results is that individuals who are not stereotypically expected to be overconfident might prove better leaders, as they are under less pressure to pursue unprofitable projects. In many environments, women are less likely to be overconfident, and they are also *perceived* as not being overconfident.⁵

After discussing the related literature, we set out the model with common priors in Section 2, and show that the DM continues projects excessively, due to reputational concerns. Section 3 examines overconfidence, and higher-order beliefs regarding overconfidence. Section 4 uses our model to examine a related question: does a newly appointed DM have a reputational incentive to cancel the projects of her predecessor? We find that the answer is yes, but only if reputational concerns are moderate. The final section concludes. For the sake of exposition, all proofs are presented in appendices A and B. Appendix C presents several extensions of the baseline model with common priors that are of independent interest, but whose presentation would delay our study of the interaction between the perceptions of overconfidence and reputational concerns.

1.1 Related Literature

There is abundant evidence that managers and political leaders are reluctant to abandon their projects in the face of bad news. Explanations fall into two broad classes: hubris, and rational-choice explanations based on incentives.

A leading rational-choice explanation for the persistence of bad policies is “gambling for

⁴For instance, in a society where women habitually defer to men, it is focal to coordinate on the men-preferred equilibrium in the battle of the sexes. Holm (2000) presents experimental evidence that supports this claim.

⁵Barber and Odean (2001) compare the common stock investments of men and women and find that men trade 45 percent more than women and earn lower returns, a difference the authors attribute to men’s overconfidence. In an influential experiment, Gneezy, Niederle, and Rustichini (2003) show that girls perform worse than boys in a competitive contest, even though their performance is no different when the situation is not competitive. This difference is pronounced when girls compete against boys. They suggest that this may be due to mis-specified beliefs in a strategic context — if girls are less confident in their abilities, they will exert less effort in a winner-take-all contest.

resurrection” (see Downs and Rocke (1994)) — a political leader who is likely to be voted out of office may be tempted to go to war in order to gamble on a major, if unlikely, change in fortune. Similarly for banks and corporations that are protected by limited liability; Freixas, Parigi, and Rochet (2000) argue that “moral hazard and gambling for resurrection are typical behaviors for banks experiencing financial distress.” Although a straightforward explanation, convex incentives are not always apposite — the phrase gambling for resurrection is suggestive of desperate times. A political leader who is popular and expects to be re-elected has concave preferences and, seeking to minimize risks, should cancel a dubious project. Managers are also likely to be risk averse with respect to income, leading them to be cautious.

An alternative explanation, based on incentives to signal ability, has been proposed by Dur (2001) and Majumdar and Mukand (2004). The underlying idea is that competent leaders always select good projects, while incompetent leaders sometimes select bad ones and may subsequently receive information about the state of the project. In both papers, cancelling a project reveals that a leader has changed her mind, and that she must therefore be the unskilled type. In Dur (2001), voters do not observe the project’s outcome, and draw inferences only from the policy maker’s decision. Stopping reveals that the project is bad, whereas continuing a project privately known to be bad hides its failure from the electorate.

Our model differs from the prior rational choice literature in several ways. First, we assume that no leader is infallible, and therefore that it is wise to sometimes reverse one’s decisions. Second, information is revealed to the public when a project or policy is continued, and obfuscated when it is aborted. Thus, continuing the policy is an informative experiment, and is analogous to disclosure with a verifiable type, as in Grossman (1981) or Milgrom (1981). We show that these two factors result in a tendency towards excessive continuation, but that the resulting inefficiency is mitigated. The policy maker’s incentives to continue are tempered by the fact that she cannot hide the truth. Even a DM who is extremely concerned with her reputation will cancel a policy if the news is bad enough.

The richness of our base model allows us to examine the interaction between overconfidence and the private information of the DM. We find that overconfidence aggravates the tendency to persist with bad projects, and that, more subtly, even the perception of overconfidence has a deleterious impact. Indeed, a culture where hubris is the norm may mar decisions even when the DM is not overconfident, and even when there are high degrees of mutual knowledge that the DM is not overconfident.

We have already discussed the empirical evidence that political and business leaders are

overconfident. This may be the result of selection: such individuals attain their leadership positions by being successful, and they may underestimate the role of luck in their success. Van den Steen (2004) provides a rational explanation for overconfidence: individuals have heterogeneous prior beliefs regarding the efficacy of different actions, and each person chooses the action that she thinks best. Consequently, her estimation of success given her choice exceeds that of an outside observer. Santos-Pinto and Sobel (2005) present a related explanation, where skills are multidimensional, and individuals weight these dimensions differently when computing a scalar measure of ability. Since an individual invests in the skills she considers more important, she has a perception of her own ability that is higher than that of others who use different weights and invest differently. Bénabou and Tirole (2002) argue that beliefs play a motivational role and that, consequently, individuals may find it optimal to disregard or forget unfavourable information. These considerations are likely to be particularly important for CEOs or political leaders, who must motivate their followers as well as themselves.

Models with heterogeneous priors are increasingly used to examine a wide range of issues. The implications for financial markets are studied by Harrison and Kreps (1978), Scheinkman and Xiong (2003) and Geanakoplos (2010). Yildiz (2003) examines bargaining when both parties are excessively optimistic about their bargaining power. Sethi and Yildiz (2016) examine the exchange of opinions when individuals have different priors. Heterogeneous priors can be viewed as an instance of agents having a mis-specified model of the world, as studied in Esponda and Pouzo (2016). We are not aware of any other work that examines the implications of higher-order beliefs regarding non-common priors.

Dekel, Fudenberg, and Levine (2004) argue against the use of Nash equilibrium as a solution concept in games where agents have different priors, and advocate the use of non-equilibrium notions, such as rationalizability. Our analysis takes on board this criticism — in Section 3.2.1 we show that our main results on the negative effects of the culture of overconfidence, also applies for rationalizable strategies. Dekel, Fudenberg, and Morris (2007) provide a connection between Bayes Nash equilibria in games without a common prior and correlated rationalizability.

Our results on the effects of higher-order beliefs regarding the lack of a common prior are reminiscent of the electronic mail game (Rubinstein (1989)), and the subsequent literature, notably Carlsson and van Damme (1993), Morris and Shin (1998) and Weinstein and Yildiz (2007). The strategic implications of mis-specified beliefs have been recently explored by Chen, Di Tillio, Faingold, and Xiong (2017).

2 The Baseline Model

2.1 Setup

We study the interaction between a decision maker (DM) who undertakes a project, and is concerned with its profitability and with an outside observer's perception of her ability. The DM can be thought of as the manager of a firm or a political leader. The observers may be many, but are modelled as a single agent, since they share a common belief. For a manager, the observer stands for the shareholders of the firm or potential employers in the managerial labor market. For a political leader, the observer may represent voters or her political followers.

We consider a model with three dates, 0, 1 and 2. At $t = 0$, nature chooses the ability $\tau \in \{H, L\}$ of the DM and the quality $\omega \in \{G, B\}$ of her project. High ability, H , is chosen with probability $\lambda \in (0, 1)$, and a more able DM is more likely to be endowed⁶ with a good project so that $1 > p_H > p_L > 0$, where $p_\tau := \Pr(\omega = G|\tau)$ denotes the probability with which type τ has a good project. Neither the DM nor the observer observe the DM's ability or the quality of the project. They therefore share a common prior $p := \lambda p_H + (1 - \lambda)p_L$ that the project is a good one.

At $t = 1$, the DM privately observes a signal that is informative about the quality of the project, and decides whether to continue the project or to terminate it. If she continues the project (action Y), a cost c is incurred, and the project's outcome is publicly realized in period two. The outcome resolves all uncertainty about the project quality. It is a success if the project is good, yielding a return v , and a failure if the project is bad, yielding zero returns. If the DM terminates the project (action N), this is publicly observed, and there is no further learning about the project's quality. The costs and returns should be interpreted as either accruing to the firm (in the manager example) or to society as large (in the politician context).

At the end of period $t = 2$ (after observing the project's outcome in the event that it was not cancelled at $t = 1$), the observer chooses an action in $[0, 1]$, where his optimal action equals his posterior belief about the DM's ability.⁷ Let β denote the posterior probability which the observer assigns to the project being good, and let ν denote the posterior probability he assigns to the DM having high ability. In Appendix A.1 we show that ν is affine in

⁶In Appendix C.2, we consider a variant of the model where the DM must first decide whether to initiate the project at $t = 0$, when project initiation incurs a cost of $k > 0$.

⁷The observer may be minimizing a quadratic loss function, or the DM's wage could be determined in a competitive labor market, where the value of the DM to any employer is her expected ability.

β , i.e.

$$\nu(\beta) = \nu(0) + \gamma\beta, \tag{1}$$

where $\gamma := \nu(1) - \nu(0) < 1$.

The DM maximizes the sum of the social payoff from the project and of the observer's action — interpreted as the DM's reputation — the latter being weighted by a constant parameter $\alpha \geq 0$ that measures the intensity of her reputational concerns. The resulting *relative* weight on β , the observer's belief at $t = 2$ about the project's quality, is $\alpha\gamma$, and is smaller:

- the greater the weight, α , given to the social payoff in the DM's objective, either due to social concerns or explicit performance pay (for a firm manager).
- the smaller the correlation between the DM's ability and the project's quality, i.e. the smaller the value of γ , which in turn depends on the difference between p_H and p_L .

That managers of firms have career concerns is well established since Holmström (1999), and there is a large subsequent literature on its implications.⁸ Similarly, many politicians enter politics motivated by social concerns, and these concerns may remain even if moderated by re-election preoccupations. The assumption that the DM's payoff is linear in both dimensions, the social payoff and her reputation, is a strong one and we make it for analytical clarity, so as to be able to focus on the differential informational content of the two experiments, stopping and continuing the project. If the DM's evaluation of either the social payoff or her reputation were convex, this would automatically bias her towards continuing the project in our model. Similarly, concavity on either dimension would bias her towards stopping the project.

The signal privately observed by the DM at the beginning of period $t = 1$ induces a cumulative distribution F over the DM's posterior belief, μ , that the project is good. For $z \in (0, 1]$, let $F(z^-) := \lim_{x \uparrow z} F(x)$, and let $\Delta(z)$ denote the size of the atom at z . Thus $F(z) = F(z^-) + \Delta(z)$. Let $\mathcal{C}(F)$ denote the support of the distribution of posterior beliefs induced by the signal, and let $\underline{\mu} := \min \mathcal{C}(F)$ and $\bar{\mu} := \max \mathcal{C}(F)$. The signal satisfies the criterion of Bayes-plausibility, i.e. $\int_0^1 \mu dF = p$. Let $\mu^{**} := \frac{c}{v}$. The optimal decision, from a social point of view, is to continue with the project at the end of $t = 1$ if $\mu > \mu^{**}$, and to abandon it if the inequality is reversed. We assume that the common prior about the

⁸Meyer and Vickers (1997) study the interaction of career concerns and explicit incentives. Prat (2005) examines the role of transparency, i.e. the information available to the observer, when a manager has career concerns.

project's quality satisfies $p > \mu^{**}$, so that the project should be continued at $t = 1$ if no information is revealed. Then, for any Bayes-plausible information structure, there must be some posterior belief such that it is socially optimal to continue the project. The signal observed by the DM is *decision-relevant* if there are some beliefs such that it is strictly optimal to stop, i.e. $F(\mu^{**}) > 0$. We shall assume throughout that the signal is decision-relevant, since otherwise the DM's decision problem at $t = 1$ is trivial — she should always continue with the project.

Definition 1 *The signal observed by the DM is rich if it is decision-relevant and if the conditional distribution $F(\mu|\mu < \mu^{**})$ is non-degenerate, i.e. it does not assign probability one to a single value of μ .*

2.2 Equilibrium Analysis

We analyze Perfect Bayesian Equilibria. A mixed strategy, σ , for the DM maps her private posterior belief, $\mu \in [0, 1]$, about the project's quality to a probability of stopping the project at $t = 1$. A strategy for the observer specifies an action in $[0, 1]$ at $t = 2$ for each public event: success, failure, and N (project termination). Since the observer must take action $\nu(1)$ in the case of success, and $\nu(0)$ in the case of failure, it suffices to specify the observer's action when the project is cancelled. We let $\rho(\beta)$ denote this action when β is the observer's belief about project quality, and identify it with the observer's strategy. Sequential rationality implies that it equals the observer's belief, $\nu(\beta)$, given in (1), about the DM's ability when he observes cancellation. In equilibrium, β must be consistent with the DM's strategy, σ . We require $\sigma(\mu)$ to be a best response to ρ for every belief $\mu \in [0, 1]$, including beliefs that are not in the support of F — this is without loss of generality and simplifies exposition.

Fix an equilibrium strategy profile, (σ, ρ) . If the DM's private belief about the project is μ , then her expected payoff from continuing the project is

$$U(Y, \mu) := \mu [v + \alpha\gamma] - c + \alpha\nu(0). \quad (2)$$

With probability μ , the project succeeds, so that the social value rises by v and the observer's belief jumps to $\nu(1)$, since success is perfectly informative of the project's quality. With the complementary probability the project fails and the observer's belief falls to $\nu(0)$. Regardless of whether the project succeeds or not, the cost c is incurred. Thus the payoff from continuation is an increasing, affine function of μ . Observe that the payoff from continuation is *independent* of σ and does not depend on the observer's beliefs about the strategy played

by the DM. In other words, since project continuation fully reveals the project's quality, it is analogous to *disclosure with a verifiable type*.⁹

However, the DM cannot verifiably disclose the project quality when she terminates the project, so her payoff from termination does depend on σ . The net effect on the project's value is zero, since there is neither a cost nor any return; so the payoff from stopping equals the reputational payoff. Suppose that, in equilibrium, the DM cancels the project on a set Ω of beliefs that has positive F -measure. Then $\mathbb{E}(\mu|\Omega)$ is well defined and equals the observer's belief, β , about the quality of the project following cancellation. Thus, the DM's payoff from cancelling the project is independent of her private belief, μ . Since the payoff from continuing the project, in (2), is strictly increasing in μ , it follows that, in any equilibrium, the DM follows a threshold strategy.¹⁰ We therefore restrict attention to strategies, σ , that can be identified with a pair (x, θ) . The threshold $x \in [0, 1]$, is such that the DM cancels the project if her private belief, μ , satisfies $\mu < x$, and continues if $\mu > x$. The DM may possibly randomize at the belief x . In this case, $\theta \in [0, 1]$ denotes the probability with which she cancels the project if her belief, μ , equals the threshold, x . This randomization is of consequence only when F has an atom, $\Delta(x) > 0$, at x .

Let $U(N, x, \theta)$ denote the DM's payoff from cancelling the project under that strategy:

$$U(N, x, \theta) := \alpha [\gamma \mathbb{E}(\mu|N; x, \theta) + \nu(0)], \quad (3)$$

where the term in square brackets is the observer's best response to project cancellation under the strategy (x, θ) , and where

$$\mathbb{E}(\mu|N; x, \theta) := \frac{\int_0^{x^-} \mu dF + \theta \Delta(x) x}{F(x^-) + \theta \Delta(x)} \quad (4)$$

denotes the observer's belief about the project's quality, conditional on the project being cancelled at $t = 1$, when the DM uses the strategy (x, θ) . It is straightforward to verify that if x is a mass point of F , then $U(N, x, \theta)$ is continuously increasing in θ — by choosing θ appropriately, any value between $U(N, x, 0)$ and $U(N, x, 1)$ can be achieved.

It will be convenient to henceforth work with the normalized payoffs $\tilde{U}(Y, x) := \frac{1}{\alpha} U(Y, x)$ and $\tilde{U}(N, x, \theta) := \frac{1}{\alpha} U(N, x, \theta)$. An equilibrium strategy for the DM is a pair (μ^*, θ^*) satis-

⁹As in Grossman (1981) or Milgrom (1981). The DM's expected reputation from continuing is the same as if she were able to verifiably disclose her private posterior, μ , to the observer.

¹⁰Even if Ω is empty, the observer's belief depends only on the observed cancellation, and is therefore independent of the DM's own belief, and so an optimal strategy must be a threshold strategy.

fyng

$$\tilde{U}(Y, x) = \tilde{U}(N, x, \theta). \quad (5)$$

Equilibrium existence follows from standard arguments and is proved in Appendix A.2. An equilibrium is the triplet $(\mu^*, \theta^*, \rho^*)$: the DM's threshold belief is μ^* ; the probability with which she cancels the project when her belief is at the threshold is θ^* ; and the observer's action about the DM's ability when the DM stops is ρ^* . If there is no mass point at the threshold μ^* , we adopt the convention that θ^* equals a fixed value $\tilde{\theta}$.¹¹

Our focus here is on multiple equilibria, which arise naturally, due to strategic complementarities — the DM's optimal threshold strategy is increasing in the observer's strategy, and vice versa. Also, the payoff from stopping given a threshold, x , is not, in general, well behaved, since the conditional expectation $\mathbb{E}(\mu|N; x, \theta)$ has upward jumps at mass points of F , facilitating multiple equilibria. Mass points occur naturally with discrete news events, such as arise under Poisson news. An example with a mass point is depicted in Figure 1, which graphs the observer's best response to (x, θ) , $\gamma \mathbb{E}(\mu|N; x, \theta) + \nu(0)$, for every threshold belief x . This affine transformation of the conditional expectation is increasing, and is set-valued at any mass point of the distribution F , such as μ_2^* . Recall from (3) that this equals $\tilde{U}(N, x, \theta)$, the scaled payoff from stopping under the strategy (x, θ) . The scaled payoff from continuing at belief $\mu = x$, $\tilde{U}(Y, x)$, is affine in x by (2). The figure shows three intersections of the two curves, at $\underline{\mu}^*, \mu_2^*$ and $\bar{\mu}^*$, corresponding to three distinct equilibrium thresholds, and the corresponding equilibrium actions of the observer, $\underline{\rho}^*, \rho_2^*$ and $\bar{\rho}^*$. At μ_2^* , the equilibrium requires that the DM randomize between stopping and continuing. Mass points are not necessary for multiplicity — it suffices that the conditional expectation, $\mathbb{E}(\mu|N; x, \theta)$, increases rapidly enough over some interval.

In general, we will use $\underline{\mu}^*$ to denote the smallest equilibrium threshold, and $\bar{\mu}^*$ to denote the largest equilibrium threshold. These are the *extremal* equilibria, and coincide if the equilibrium is unique. The following definition will be of use in our later analysis. Fix an equilibrium threshold μ^* , and define the following function:

$$\xi(\mu) := [\tilde{U}(N, \mu, 0) - \tilde{U}(Y, \mu)][\mu^* - \mu].$$

Definition 2 *An equilibrium with threshold μ^* is:*

- *Left-stable if there exists an open interval $(\tilde{\mu}, \mu^*)$ such that $\xi(\mu) > 0$ for all $\mu \in (\tilde{\mu}, \mu^*)$.*

¹¹Note that the threshold μ^* need not belong to the support of F , since there may be gaps in the distribution.

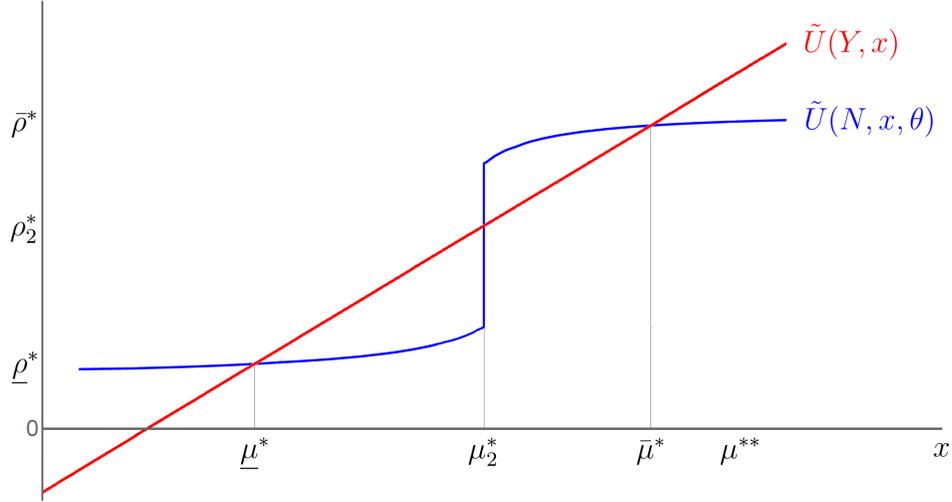


Figure 1: The DM's scaled payoff from continuing at the belief $\mu = x$, $\tilde{U}(Y, x)$, and her scaled payoff, $\tilde{U}(N, x, \theta)$, from stopping under the strategy $\sigma = (x, \theta)$.

- *Right-stable if there exists an open interval $(\mu^*, \tilde{\mu})$ such that $\xi(\mu) > 0$ for all $\mu \in (\mu^*, \tilde{\mu})$.*
- *Stable if it is both left-stable and right-stable.*
- *Unstable if there exists an open interval containing μ^* such that $\xi(\mu) < 0$ for all $\mu \neq \mu^*$ in this interval.*

Intuitively, stability corresponds to $\tilde{U}(N, \mu, 0)$ being flatter than $\tilde{U}(Y, \mu)$ at μ^* . In Figure 1, the equilibria with the largest and smallest thresholds are stable, while the mixed strategy equilibrium is unstable. More generally, any mixed strategy equilibrium is necessarily unstable, since the “slope” of the conditional expectation is infinite at the corresponding threshold.

Proposition 1 $\underline{\mu}^* > \underline{\mu}$ and $\bar{\mu}^* < \mu^{**}$, so that the set of equilibrium thresholds is contained in $(\underline{\mu}, \mu^{**})$. Both extremal equilibria are in pure strategies, so that a pure strategy equilibrium exists. The equilibrium with the smallest threshold, $\underline{\mu}^*$, is left-stable and the equilibrium with the largest threshold, $\bar{\mu}^*$, is right-stable. Thus, if an equilibrium is unique, it is in pure strategies and is stable.

The DM's optimal stopping threshold, x , is increasing in ρ , the observer's action. Also, ρ is increasing in x . Thus our game is supermodular, and the fact that the extremal equilibria

are in pure strategies can be deduced from the results in Milgrom and Roberts (1990). The notion of left-stability is novel, and will play an important role when we consider higher-order beliefs regarding overconfidence.

Stability has implications for comparative statics. If an equilibrium is either stable or unstable, then there exists a nearby equilibrium if parameters are changed by a small amount. Consider, for example, greater reputational concerns, i.e. a larger value of α . At the initial equilibrium threshold, μ^* , this increases the payoff from continuation, but does not change the payoff from stopping. In Figure 1, the line $\tilde{U}(Y, \cdot)$ swivels around the point $(\mu^{**}, \tilde{U}(Y, \mu^{**}))$, becoming flatter, so that the payoff from continuation exceeds the payoff from stopping at any initial equilibrium. If the equilibrium is stable, then the equality in payoffs can only be restored by a decrease in the equilibrium threshold, so that the DM continues the project more often. At an unstable equilibrium, the comparative statics are reversed, and the threshold must increase. These claims are proved, slightly more generally, in Appendix A.5. As a consequence, if equilibrium is unique, greater reputational concerns increase the tendency to continue unprofitable projects.

2.3 Efficiency

A social planner who seeks to maximize the value of the project uses a strategy with threshold μ^{**} . In any equilibrium, the observer's posterior belief is a martingale, and thus its expectation at date zero must equal the prior, p . Consequently, the DM's ex-ante payoff in any equilibrium, before she observes her private information, equals the payoff from the project under this equilibrium, plus her expected reputational payoff. The latter does not vary across equilibria. Therefore, the equilibrium that maximizes the project's payoff is also the one that is best for the DM. From Proposition 1, $\bar{\mu}^* < \mu^{**}$, so profitable projects are never inefficiently terminated. Nonetheless, this strict inequality does not immediately imply excessive continuation of unprofitable projects, since there may be a gap in the support of F . The following proposition identifies the conditions under which there is inefficient excessive continuation — notably, when the DM's beliefs are rich and reputational concerns are large enough.

Proposition 2 *1. When the distribution of beliefs is rich, then there is inefficient continuation in every equilibrium if reputational concerns are sufficiently large, i.e. if α is large enough.*

*2. When the distribution of beliefs is continuous and has no gap immediately below μ^{**} ,*

there is inefficient continuation in every equilibrium, no matter how small reputational concerns are.

3. When the distribution of beliefs is not rich, so that there is only a single belief below μ^{**} in the support of F , then the unique equilibrium is efficient.
4. Inefficiency is bounded: no matter how important reputational concerns are, the DM never continues the project at the lowest belief, $\underline{\mu}$.
5. Inefficiency is one-sided: the DM never terminates a profitable project.
6. When there are multiple equilibria, the most efficient equilibrium is the equilibrium with the largest threshold, $\bar{\mu}^*$.

Ideally, players would coordinate on the equilibrium with the largest threshold, since it is the most efficient one. However, this paper will show that a culture of overconfidence leads them to play the most inefficient one, even when there is a high level of mutual knowledge that there is no overconfidence. This is discussed in Section 3.2.

3 Overconfident Leaders

Suppose that the DM and the observer have different priors on the competence of the DM, and therefore, on the quality of the project. Specifically, assume that the DM is overconfident, and has a prior q on the quality of the project, that is strictly greater than p , the prior of the observer. This defines G^0 , a game with non-common priors. For the remainder of this section, we make the following assumption, for the sake of exposition — the appendix shows that our arguments apply without this assumption.

Assumption 3 *The distribution of F is atomless.*

Suppose that the DM observes a signal s with likelihood ratio ℓ .¹² Her posterior belief equals

$$\pi(\ell) := \frac{q\ell}{q\ell + 1 - q}. \quad (6)$$

The posterior belief of the observer, if he were to observe the same signal, would be $\mu(\ell)$:

$$\mu(\ell) := \frac{p\ell}{p\ell + 1 - p}. \quad (7)$$

¹²Here, $\ell = \frac{h(s|G)}{h(s|B)}$, where $h(s|\omega)$ denotes the value of the probability density function at s when the project quality is ω .

In the interest of parsimony, we identify a signal with its associated likelihood ratio, and define the above belief-updating functions for all $\ell \in [0, \infty)$, with $\pi(\infty) = \mu(\infty) = 1$.

The DM's payoff from continuing the project, $\tilde{U}(Y, \pi)$, is, as before, an affine function of her posterior belief, π . Her payoff from stopping cannot depend upon her private belief π . Thus, any equilibrium must be in threshold strategies, and given Assumption 3, it is in pure strategies. Thus there exists a likelihood ratio, denoted ℓ_0 , such that the DM continues the project for $\ell \geq \ell_0$ and terminates it for $\ell < \ell_0$. Let $\mu_0 := \mu(\ell_0)$.

It will be analytically convenient to identify the DM's threshold strategy with the threshold belief μ_0 . However, observe that this does not refer to the DM's own belief upon observing the threshold signal, which is given by $\pi(\ell_0)$. Rather, $\mu_0 := \mu(\ell_0)$ would be the posterior belief held by the observer if he were to observe the threshold signal, ℓ_0 .

Fix a threshold strategy for the DM, with threshold $x \in [0, 1]$. If the DM terminates the project, the observer forms the belief $\mathbb{E}(\mu|N; x) = \mathbb{E}_F(\mu|\mu < x)$, about the project quality, where the expectation is taken with respect to F , the distribution of beliefs under the prior p .¹³ Consequently, the DM's payoff from termination under the strategy x is given by

$$\tilde{U}(N, x) := \gamma \mathbb{E}(\mu|N; x) + \nu(0). \quad (8)$$

Let us define $\pi^\dagger(\mu)$ to be the posterior belief of the DM after a signal that would induce in the observer the posterior belief μ .¹⁴ That is $\pi^\dagger : [0, 1] \rightarrow [0, 1]$ is obtained by using the inverse of (7) as the argument in (6). This has the explicit form:

$$\pi^\dagger(\mu) = \frac{q(1-p)\mu}{q(1-p)\mu + (1-q)p(1-\mu)}. \quad (9)$$

Lemma 1 *The function π^\dagger is strictly increasing and strictly concave. For any $\mu \in (0, 1)$, $\pi^\dagger(\mu) > \mu$.*

The DM's private belief after observing the threshold signal ℓ_0 is $\pi^\dagger(\mu_0)$, and her payoff from continuing the project at that belief is

$$\tilde{U}(Y, \pi^\dagger(\mu_0)) = \frac{1}{\alpha} [\pi^\dagger(\mu_0) [v + \alpha\gamma] - c + \alpha\nu(0)]. \quad (10)$$

¹³Let $H_{s|\omega}$ denote the cumulative distribution function of the DM's private signals in the state ω . From the point of view of the observer, whose prior belief about the state $\omega = G$ is p , $H_p := p H_{s|G} + (1-p) H_{s|B}$ is the distribution over the signals that the DM observes. We let F denote the corresponding distribution over the resulting posterior beliefs, $\mu(\ell(s))$.

¹⁴Equivalently, $\pi^\dagger(\mu)$ is the posterior belief under prior q after a signal that would induce the posterior belief μ under the prior p .

The threshold μ_0 is an equilibrium strategy for the DM if

$$\tilde{U}(Y, \pi^\dagger(\mu_0)) = \tilde{U}(N, \mu_0), \quad (11)$$

and, by (8), the right-hand side equals ρ_0 , the observer's equilibrium action.

Equilibrium existence follows from the same arguments as in Appendix A.2, the only difference being that if overconfidence is sufficiently acute (i.e. the difference $q - p$ is sufficiently large), the DM may continue the project even at the lowest possible belief.

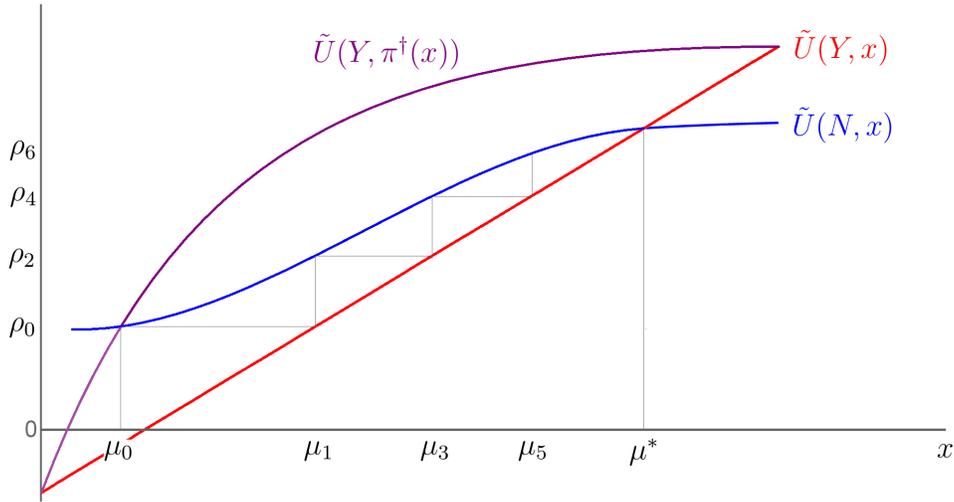


Figure 2:

Figure 2 illustrates an equilibrium of the game G^0 . The straight line, $\tilde{U}(Y, x)$, shows the payoff from continuation without overconfidence, i.e. when the posterior belief under the prior p equals the threshold x . The concave function $\tilde{U}(Y, \pi^\dagger(x))$ shows the DM's payoff from continuation with overconfidence, i.e. when the DM's posterior belief is $\pi^\dagger(x)$. By Lemma 1, this lies above $\tilde{U}(Y, x)$. The equilibrium threshold with overconfidence, μ_0 , satisfies (11). It lies to the left of μ^* , the equilibrium threshold under a common prior, which satisfies $\tilde{U}(Y, \mu^*) = \tilde{U}(N, \mu^*)$. Thus, overconfidence leads to excessive continuation, over and above that arising under a common prior. This has two reasons. The first is straightforward: since the DM's belief is $\pi^\dagger(\mu_0) > \mu_0$, the DM continues because she has more optimistic beliefs in G^0 than in the game with a common prior. The second reason is more subtle. Since in G^0 the observer knows that the DM is overconfident (as compared to his own prior), his inference upon project termination is more adverse. Thus the DM knows that she will be

penalized more for terminating the project in G^0 than in the game with a common prior.

The argument, that greater confidence (or overconfidence) aggravates the tendency to continue bad projects, is more general. In the appendix, we show that the game G^0 is supermodular: the observer's best response is increasing in the DM's threshold, and the DM's optimal threshold is increasing in the observer's strategy. In consequence, the results of Milgrom and Roberts (1990) imply that the belief thresholds of the DM at the extremal equilibria are decreasing in q , the prior belief of the DM. Furthermore, at any stable equilibrium, the effect of a small increase in q is to reduce the equilibrium threshold.

3.1 Perceived Overconfidence

We now define G^1 , a game where the DM and observer share the same prior, p , about the project. However, the outside observer *believes* that the DM is overconfident, i.e. that she has a prior $q > p$. We assume that the DM is aware of this belief. Specifically, assume:

- The DM and the observer share the prior p .
- **(T1)** The observer believes that the DM's prior is $q > p$.
- The observer's second-order belief is known to the DM, i.e. the DM knows T1.

Recalling that G^0 refers to the game with actual overconfidence discussed in the previous section, an alternative formalization of the game G^1 is as follows:

- The DM and the observer share the prior p .
- The observer believes that the game G^0 is being played.
- The DM knows that the observer believes that G^0 is being played.

Let (μ_0, ρ_0) be an equilibrium of the game G^0 . In any equilibrium of the game G^1 , the observer's strategy must be the same as in an equilibrium of the game G^0 , since he believes that G^0 is being played. In contrast, the DM's strategy differs across the two games. This is a feature that will recur in our later analysis. Thus, an equilibrium of the game G^1 is a triple $(\mu_0, \rho_0, \sigma_1)$ consisting of an equilibrium, (μ_0, ρ_0) , of the game G^0 , and σ_1 , a best response to ρ_0 given the DM's prior, p .

The optimal strategy for the DM, σ_1 , must be a threshold strategy, for the same reasons as before. Let ℓ_1 denote the equilibrium cut-off signal for the DM in the game G^1 , and let

$\mu_1 := \mu(\ell_1)$ be the corresponding threshold belief, derived according to (7). It must satisfy the equilibrium condition:

$$\tilde{U}(Y, \mu_1) = \rho_0. \quad (12)$$

Since the left-hand side of (12) is strictly increasing in μ_1 , there is a unique solution. In other words, for any equilibrium (μ_0, ρ_0) of the game G^0 with overconfidence, there is a unique equilibrium, (μ_0, ρ_0, μ_1) , in the game G^1 with perceived overconfidence. This is illustrated in Figure 2. Here, the unique equilibrium of the game G^0 has threshold, μ_0 , satisfying (11). In G^1 , the unique corresponding threshold, μ_1 , satisfies (12), where the right-hand side equals the observer's action, ρ_0 , in both equilibria. The equilibrium *outcome*¹⁵ in the game G^1 is given by the pair (μ_1, ρ_0) . In the game G^0 , the equilibrium outcome is (μ_0, ρ_0) .

Observe from (11) and (12) that $\mu_1 = \pi^\dagger(\mu_0)$. Thus, the DM's *equilibrium cut-off beliefs are identical* in the two games, with actual overconfidence and perceived overconfidence. Lemma 1 then implies that $\mu_1 > \mu_0$, or equivalently $\ell_1 > \ell_0$, so a DM who is indeed overconfident requires more adverse news to cancel the project than a DM who is only perceived to be overconfident. Nonetheless, $\mu_1 < \mu^*$, so that, when compared to the game with a common prior analyzed in Section 2.2, the perception of overconfidence penalizes the reputation of a DM who is not, in fact, overconfident, and makes her more reluctant to cancel the project.

Consider now the case where there are multiple equilibria in the game G^0 . Any such equilibrium, (μ_0, ρ_0) , induces a unique threshold in G^1 , $\mu_1 > \mu_0$, satisfying $\mu_1 = \pi^\dagger(\mu_0)$. Thus the equilibrium behavior of the DM in G^1 is uniquely determined by the equilibria in G^0 . Consequently, there is the same number of equilibria in G^1 as in G^0 (with each induced threshold strictly greater than the one inducing it).

3.2 Higher-Order Beliefs about Overconfidence

Suppose that the DM is not overconfident and the observer knows this. However, the DM believes that the observer believes her to be overconfident. In this case, we can define the game G^2 with perceived perceived overconfidence. More generally, let us consider the game G^N , where the DM is not overconfident, but this is not common knowledge. Define the following statements:

- (S1) The observer believes that the DM's prior is p .

¹⁵The *outcome* of a game is the distribution over terminal nodes, i.e. a joint distribution over states and player actions.

- **(T1)** The observer believes that the DM's prior is $q > p$.

and for every integer $K > 1$,

- **(SK)** (K even) The DM believes that $S(K-1)$ is true.
- **(SK)** (K odd) The observer believes that $S(K-1)$ is true.
- **(TK)** (K even) The DM believes that $T(K-1)$ is true.
- **(TK)** (K odd) The observer believes that $T(K-1)$ is true.

Then, for every integer $N > 1$, the game G^N is defined by:

- The DM and the observer share the prior p .
- The statements $S1$ to $S(N-1)$ are true.
- The statement TN is true.
- Both players know TN .

The games G^0 and G^1 were defined in earlier sections. Let G denote the game with common priors analyzed in Section 2.2.

Consider the sequence of games, (G^N) , $N \in \{0\} \cup \mathbb{N}$. Let \mathcal{E} denote the set of even numbers, and let \mathcal{O} denote the set of odd numbers. We define a sequence of strategies $(\mu_0, \rho_0), (\mu_n)_{n \in \mathcal{O}}, (\rho_n)_{n \in \mathcal{E}}$. An equilibrium of the game G^N consists of the sequence truncated at N , with the property that:

- (μ_0, ρ_0) is an equilibrium of the game G^0 .
- For any $n \in \mathcal{O}, n \leq N$, μ_n is a best response to ρ_{n-1} .
- For any $n \in \mathcal{E}, n \leq N$, ρ_n is a best response to μ_{n-1} .

Thus for any odd n , μ_n is uniquely determined by ρ_{n-1} as follows:

$$\tilde{U}(Y, \mu_n) = \rho_{n-1}. \quad (13)$$

When n is an even number, ρ_n is uniquely determined by μ_{n-1} via:

$$\rho_n := \tilde{U}(N, \mu_{n-1}). \quad (14)$$

The equilibrium construction is best illustrated in Figure 2. The pair (μ_0, ρ_0) represents the equilibrium strategies in G^0 , the game with overconfidence. In G^1 , the game with

perceived overconfidence, the observer believes that G_0 is being played, and therefore plays ρ_0 . The DM's equilibrium strategy in G^1 is μ_1 , the best response to ρ_0 , satisfying (13). In G^2 , the game with perceived perceived overconfidence, the DM believes that G^1 is being played, and therefore plays μ_1 . The observer's equilibrium strategy is ρ_2 , his best response to μ_1 , which satisfies (14). Thus in any game G^n , the equilibrium outcomes are (μ_n, ρ_{n-1}) if n is odd and (μ_{n-1}, ρ_n) if n is even.

Observe that the the equilibrium values μ_n for n odd and ρ_n for n even are defined iteratively and converge to (μ^*, ρ^*) , the equilibrium the game with common priors. Also, even if there were additional equilibria to the right of (μ^*, ρ^*) , the iterative process starting at (μ_0, ρ_0) cannot go to the right of (μ^*, ρ^*) .

This last point can be made more precise as follows. Fix an equilibrium (μ_0, ρ_0) in the game with overconfidence, G^0 . Let $\mu_+^*(\mu_0)$ denote the smallest equilibrium threshold in the game G that is larger than μ_0 , i.e.

$$\mu_+^*(\mu_0) = \min\{\mu > \mu_0 : \tilde{U}(Y, \mu) = \tilde{U}(N, \mu)\}. \quad (15)$$

Let $\rho_+^*(\mu_0)$ denote the corresponding equilibrium strategy for the observer. Thus $(\mu_+^*(\mu_0), \rho_+^*(\mu_0))$ is the smallest equilibrium in G that is larger than (μ_0, ρ_0) . In the appendix, we show that such an equilibrium is necessarily left-stable, and the following proposition shows that the sequence of equilibria in the games $(G^n)_{n \geq 0}$ converges to it.

Proposition 3 *Fix an equilibrium, (μ_0, ρ_0) , of G^0 , the game with overconfidence. This induces sequences $(\mu_n)_{n \in \mathcal{O}}$ and $(\rho_n)_{n \in \mathcal{E}}$, such that for $n \in \mathcal{O}$, in any game G^n , the equilibrium outcome is (μ_n, ρ_{n-1}) , and in any game G^{n+1} , the equilibrium outcome is (μ_n, ρ_{n+1}) . The sequences (μ_n) and (ρ_n) are both increasing, and converge to μ_+^* and ρ_+^* respectively as defined in (15), where (μ_+^*, ρ_+^*) is a left-stable pure strategy equilibrium of G , the game with common priors.*

Observe that, in the absence of common knowledge of common priors, no unstable equilibrium can be approximated. In the appendix, we show that this extends to mixed equilibria, which are necessarily unstable.

What are the implications of perceived under-confidence, i.e. of the DM having a prior, q , that is strictly less than p , the prior of the observer? The analysis is symmetric and with higher-order beliefs, the only modification to Proposition 3 is that the equilibrium of the game G with common priors that is selected must be right-stable. In conclusion, generically,

only stable equilibria are selected if there is lack of common knowledge of common priors, regardless of how priors diverge.

3.2.1 Large Overconfidence

Overconfidence, when it exists, is likely to be substantial. In the context of our model, that translates to the difference in priors, $q - p$, being large. The direct consequences of large overconfidence, or large perceived overconfidence, are straightforward — the tendency to continue projects is aggravated. We now examine the implications of this assumption for the limit outcomes in the game G with common knowledge of common priors. Our first result is substantive — only the most inefficient equilibrium of G is selected. Our second result addresses possible conceptual concerns — the result remains true if our solution concept is weakened to rationalizability. That is, it is not essential to assume that players play a Nash equilibrium in any of the games where there is a lack of common knowledge of common priors. This latter result addresses the concern raised in Dekel, Fudenberg, and Levine (2004) that there is a conceptual inconsistency in imposing equilibrium requirements in a setting where agents have different priors.

The intuition for our results is as follows. Consider $(\underline{\mu}^*, \underline{\rho}^*)$ the equilibrium with the smallest threshold in the game with common priors, G .¹⁶ If q is sufficiently large, then $\pi^\dagger(\mu)$ becomes sufficiently large that the DM never wants to stop the project at any $\mu \geq \underline{\mu}^*$. Thus *every* equilibrium threshold in the game G^0 lies below $\underline{\mu}^*$. Since the game G^0 is supermodular, any rationalizable strategy ρ_0 for the observer has $\rho_0 \leq \underline{\rho}^*$. Thus any sequence of rationalizable thresholds, $(\mu_n)_{n \in \mathcal{O}}$, must start to the left of $\underline{\mu}^*$, and must therefore converge to $\underline{\mu}^*$. Similarly, any sequence of rationalizable actions, $(\rho_n)_{n \in \mathcal{E}}$, must converge to $\underline{\rho}^*$. Thus the limit points are unique, even if one considers rationalizable profiles, and does not require that players coordinate on an equilibrium.

Proposition 4 *If $q - p$ is sufficiently large, then any sequence of rationalizable DM strategies, $(\mu_n)_{n \in \mathcal{O}}$, converges to $\underline{\mu}^*$, and any sequence of rationalizable observer strategies, $(\rho_n)_{n \in \mathcal{E}}$, converges to $\underline{\rho}^*$.*

This result shows that a culture where business and political leaders are expected to be substantially overconfident may have pernicious effects even for leaders who are not overconfident. Indeed, it can be mutual knowledge to a high degree that the leader in

¹⁶By Proposition 1, this equilibrium must be in pure strategies, even if we allow for mass points in the distribution of F .

question has the right beliefs. Nonetheless, the lack of common knowledge entailed by the culture ensures that the most inefficient equilibrium is selected.

The implications of large underconfidence (i.e. q substantially smaller than p) are in the opposite direction: the equilibrium selected is the most efficient equilibrium in G . Our theory highlights a novel and important channel through which cultural stereotypes might determine outcomes: via higher-order beliefs. It may also have normative implications. In many environments, women are *perceived* as being less confident than men. Indeed, there is evidence that they are, in fact, less confident.¹⁷ Thus, a female CEO may benefit from the stereotyping and may feel less pressure to pursue unprofitable projects than her male counterparts, since she will be penalized less for cancellations.

4 Trashing a Predecessor's Reputation

Politicians and CEOs often scrap the projects or policies of their predecessors, even when the policy or project in question is not ideological, and the environment is one of common values. Our model can provide a rational explanation for this phenomenon.

Suppose that the DM wants to *minimize the reputation of her predecessor*. Such a motivation arises naturally, for purely rational reasons. In the political context, if the predecessor is from a different party, then a political leader mitigates competition by depicting her opponents as incompetent. In the context of a firm, it is plausible that the actions of a CEO has persistent effects on the firm's profits. If the predecessor is perceived as being of low ability, then any improvement in firm value will be attributed to the current CEO's ability, and will directly increase her own reputational payoff.

Assuming that the DM's payoff is linear in the perceived ability of the predecessor, her objectives are then captured by our baseline model in Section 2, if we now assume that the coefficient α is negative. Let us assume that the reputational concern is not too large, so that $v + \alpha\gamma > 0$. Then the payoff from continuing the project, $U(Y, \mu)$, defined in (2), is strictly increasing in μ , the DM's private belief about the project quality. Fix an equilibrium σ . Since the payoff from termination must be measurable with respect to the stopping decision, it is constant with respect to μ . Thus any equilibrium must be in threshold strategies, where the DM continues the project for beliefs above the threshold x , stops it for beliefs below, and possibly randomizes, stopping with probability θ , at the threshold. Since $\alpha < 0$, the

¹⁷See, for example, the papers by Barber and Odean (2001) and Gneezy, Niederle, and Rustichini (2003) discussed in the introduction.

payoff from terminating the project, $U(N, x, \theta)$ defined in (3), is now decreasing in x . It is also decreasing in θ at any x that is a mass point of F .

Given our assumption that the signal is decision-relevant, so that it is optimal to stop at $\underline{\mu}$, we have $U(Y, \underline{\mu}) < U(N, \underline{\mu}, \theta)$ for every θ . There are two possibilities. Either $U(Y, x) < U(N, x, \theta)$ over the entire support, so that the DM always cancels the project, or there is an interior equilibrium, (μ^*, θ^*) , satisfying

$$U(Y, \mu^*) = U(N, \mu^*, \theta^*).$$

In either case, the equilibrium is unique, and the set of beliefs at which the DM cancels the project is non-empty, so that the observer's belief upon termination is well defined, by (4).

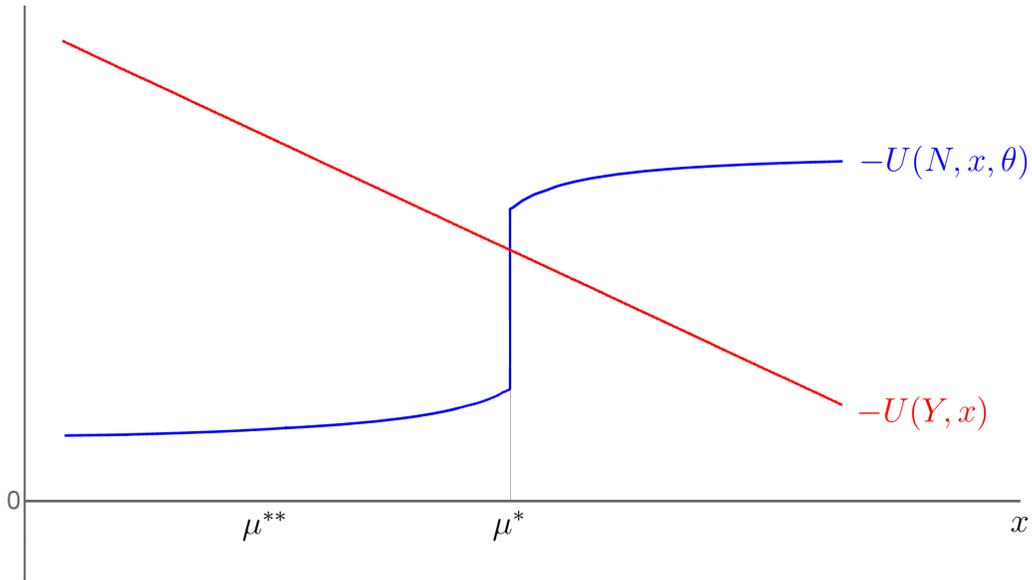


Figure 3:

Figure 3 depicts the equilibrium. The straight line, $-U(Y, x)$, represents the *negative* of the payoff from continuation, while $-U(N, x, \theta)$ represents the *negative* of the payoff from cancellation. Since the first is downward sloping and the second is upward sloping, uniqueness follows — there is either one intersection point, or first is always above the second, implying that continuation is always worse than stopping.

Observe that the DM now gets a reputational *premium* from cancellation at the threshold, since $\mathbb{E}(\mu|N; \mu^*, \theta^*)$ is always weakly less than μ^* . That is, at the threshold, by cancelling the project the DM permits a more adverse inference on her predecessor than she would be

continuing the project. It follows that $\mu^* > \mu^{**}$, where $\mu^{**} := \frac{c}{v}$ is the socially efficient cut-off. In equilibrium, the DM repeals a project even though she deems it socially worthwhile.

Since equilibrium is unique and regular, the comparative statics properties are also straightforward. Suppose that the absolute value of α increases, so that the reputational component is given greater weight. The straight line swivels around the point $(\mu^{**}, -U(Y, \mu^{**}))$, becoming flatter. Thus the new equilibrium threshold is (weakly) greater, and the inefficiency increases. Thus, greater reputational concerns give rise to more destructive behavior towards a predecessor's projects.

However, there is a qualitative change in equilibrium once reputational concerns become sufficiently large. Suppose now that α is sufficiently small that $v + \alpha\gamma < 0$. Now the payoff from continuing the project, $U(Y, \mu)$, is decreasing in μ . Since the payoff from stopping is constant, the equilibrium must be in threshold strategies, but with the property that the DM stops the project above the threshold and continues below the threshold. We now show that there is a unique equilibrium, where the DM always continues the project.

Let us show that always continuing the project is an equilibrium (the proof of uniqueness is in Appendix A.6). Since cancelling the project is never observed in this equilibrium, we must specify the observer's beliefs when a cancellation does occur. We stipulate that the observer believes that the DM observed the highest signal and holds the belief $\bar{\mu}$ — we will see that these beliefs satisfy the D1 refinement (Cho and Kreps (1987)).

To verify that this assessment is an equilibrium, consider the choice of the DM at $\bar{\mu}$. The project is profitable at $\bar{\mu}$ (since the signal is decision-relevant), and the reputational payoff is $\bar{\mu}$ under both continuation and cancellation. We therefore have that $U(Y, \bar{\mu}) > U(N, \bar{\mu})$, where $U(N, \beta) := \alpha[\gamma\beta + \nu(0)]$ is the DM's payoff from cancelling the project, if in response the observer assigns the belief β to the project being good. Since $U(Y, \mu) > U(Y, \bar{\mu})$ for every $\mu < \bar{\mu}$, whereas the payoff from cancellation is constant in μ , continuation is strictly better than termination for a DM with private belief $\mu < \bar{\mu}$.

Let us now show that the observer's belief satisfies D1. Let $\tilde{\mu}$ denote the belief satisfying $U(Y, \bar{\mu}) = U(N, \tilde{\mu})$. Then, a DM with posterior belief $\bar{\mu}$ is indifferent between continuation and cancellation, if, upon cancellation, the observer assigns the belief $\tilde{\mu}$ to the project being good. The payoff from stopping is $U(N, \bar{\mu})$, irrespective of the DM's private belief μ , but the payoff from continuing is strictly decreasing in μ . Thus, for every belief $\mu < \bar{\mu}$, the DM strictly prefers to continue. This verifies that the beliefs satisfy D1.

We therefore have the following proposition.

Proposition 5 *Suppose that $\alpha < 0$, so that the DM seeks to minimize the observer's beliefs*

about project quality. If reputational concerns are moderate, so that $v + \alpha \gamma > 0$, then there is a unique equilibrium, with threshold $\mu^ > \mu^{**}$, so that the DM terminates the project too often. If reputational concerns are extreme, so that $v + \alpha \gamma < 0$, then in the unique equilibrium, the DM always continues the project.*

Suppose that the DM and the observer have different prior beliefs about the quality of the project — the DM’s prior, q , being less than p , the prior of the observer. That is, assume the DM is excessively pessimistic about the quality of the project, and that these priors are common knowledge. Assume also that $v + \alpha \gamma > 0$, so that reputational concerns are moderate. The first-order effect is that the DM cancels the project more often than the outside observer would, since she is more pessimistic. However, the observer draws less negative inferences about the project upon cancellation, thereby reducing the DM’s payoff. Thus, in equilibrium, the observer’s perception of the DM’s excess pessimism mitigates the direct effect of excess pessimism. Since the strategies of the DM and of the observer are strategic substitutes, the effect of heterogeneous priors is mitigated. This is in contrast with the case of $\alpha > 0$, where strategic complementarity meant that the perception of overconfidence amplified its direct effect. Finally, since equilibrium is unique in the game with common priors, higher-order beliefs cannot play a role in selecting equilibria.

To summarize: reputational concerns can explain the tendency of a new manager or political leader to cancel her predecessor’s projects, but only if these reputational concerns are tempered and moderate. If a leader is known to have extreme reputational concerns (or, equivalently, a disregard for social welfare), then one has the opposite inefficiency — every project of the predecessor is carried out to conclusion.

5 Conclusion

This paper has two contributions. First, we show that a reputationally concerned leader who receives private information will be biased towards more informative experiments, inducing her to pursue her chosen projects to the end, even when she does not deem them socially worthwhile. Second, we show that even if the agent is not overconfident, being perceived to be overconfident penalizes her. Indeed, the lack of common knowledge that she is not overconfident can have substantive effects, and aggravate inefficiency, by selecting the worst equilibrium in the game with common priors. Thus a culture where leaders are expected to be overconfident has negative consequences, even when a leader has properly calibrated beliefs and is known to have correct beliefs.

A Appendix: Proofs relating to Sections 2 and 4

A.1 Relation between ν and β

The relation between ν and β is as follows. When the project succeeds, so that $\beta = 1$,

$$\nu(1) = \frac{\lambda p_H}{\lambda p_H + (1 - \lambda) p_L}.$$

When the project fails, $\beta = 0$ and

$$\nu(0) = \frac{\lambda(1 - p_H)}{\lambda(1 - p_H) + (1 - \lambda)(1 - p_L)}.$$

ν must satisfy the martingale property for a (hypothetical) experiment which reveals project quality, so that

$$\nu(\beta) = \beta \nu(1) + (1 - \beta) \nu(0),$$

which verifies that $\nu(\beta)$ is given by (1).

A.2 Proof of Equilibrium Existence

Consider the correspondence $g : [\underline{\mu}, \mu^{**}] \rightrightarrows \mathbb{R}$ defined by

$$g(x) = \tilde{U}(Y, x) - \tilde{U}(N, x, \theta), \quad \theta \in [0, 1]. \quad (\text{A.16})$$

The correspondence g is singleton-valued and continuous at any point, μ , that is not a mass point of F , and convex-valued and upper-hemicontinuous at any mass point of F . At $\underline{\mu}$, the set $g(\underline{\mu})$ lies strictly below zero, since there is no reputational loss from cancelling the project, and a loss to social value from continuing it. The set $g(\mu^{**})$ lies strictly above zero, since the social payoffs from continuing or stopping are equal, and there is a reputational loss from stopping (since the signal is decision-relevant). Thus, there exists μ^* such that $0 \in g(\mu^*)$. As already noted, μ^* need not be in the support of F , since there may be gaps in the distribution. If μ^* is a mass point of F , then there is a unique θ^* such that (μ^*, θ^*) is an equilibrium strategy for the DM. If μ^* is not a mass point, the value of θ^* is irrelevant and equals $\tilde{\theta}$ by our convention.

A.3 Proof of Proposition 1

Let μ^* be an equilibrium threshold. We show first that $\underline{\mu} < \mu^* < \mu^{**}$ for every $\alpha > 0$. Consider the threshold $x \geq \mu^{**}$. Since the signal observed by the DM is decision-relevant, $\mathbb{E}(\mu|N; x, \theta) < x$, so the DM incurs a reputational loss by terminating the project when her posterior belief, μ , equals the threshold, x . In addition, there is a social loss from termination when $\mu = x$. Consequently, any equilibrium threshold μ^* must be strictly below μ^{**} . Now consider the threshold $x = \underline{\mu}$. Since $\mathbb{E}(\mu|N; \underline{\mu}, \theta) = \underline{\mu}$, there is no reputational loss from termination when $\mu = x$. However, since $\mu^{**} > \underline{\mu}$ for any decision-relevant signal, there is a strict social loss from continuation when $\mu = x$. Consequently, any equilibrium threshold μ^* must be strictly greater than $\underline{\mu}$.

We show next that an equilibrium with threshold $\underline{\mu}^*$ is left-stable and in pure strategies. Consider a strictly increasing sequence, (x_n) , with $x_1 = \underline{\mu}$ and which converges to $\underline{\mu}^*$. We have just argued that $g(\underline{\mu}) < 0$. Moreover, this inequality must hold all along the sequence: $g(x_n) < 0$ for every n . Suppose instead that for some n , $g(x_n) = 0$. Then, there exists an equilibrium with a threshold lower than $\underline{\mu}^*$, a contradiction. Suppose that for some n , $g(x_n) > 0$. Then, since g is upper-hemicontinuous, there exists $x < x_n$ such that $g(x) = 0$, again contradicting the definition of $\underline{\mu}^*$. This proves left-stability. The equilibrium is in pure strategies if there is no mass point at $\underline{\mu}^*$. Suppose that F has a mass point at $\underline{\mu}^*$. Then the conditional expectation $\mathbb{E}(\mu|N; x, \theta)$ has an upward jump at $x = \underline{\mu}^*$. Since $U(Y, x) < U(N, x, \theta)$ for every $x < \underline{\mu}^*$, and since $U(Y, x)$ is continuous in x , $U(Y, \underline{\mu}^*)$ can equal $U(N, \underline{\mu}^*, \theta)$ only if $\theta = 0$. This proves that the equilibrium with the lowest threshold is in pure strategies. The proof that the equilibrium with threshold $\bar{\mu}^*$ is right-stable and in pure strategies is similar, and we omit the details.

A.4 Proof of Proposition 2

If the signal observed by the DM is not rich, then $\underline{\mu}$ is the only point in $\mathcal{C}(F)$ below μ^{**} . Therefore, the equilibrium is unique, since the conditional expectation $\mathbb{E}(\mu|N; x, \theta)$ is constant for $x \in (\underline{\mu}, \mu^{**})$. By Proposition 1, the equilibrium threshold is strictly greater than $\underline{\mu}$. Therefore, the equilibrium is efficient, proving (3). When there is no gap in $\mathcal{C}(F)$ immediately below μ^{**} , then the interval $(\bar{\mu}^*, \mu^{**})$ has positive F -measure, and there is inefficient continuation on this interval, thereby proving (2). Assume that there is a gap in $\mathcal{C}(F)$ immediately below μ^{**} , and let $\hat{\mu} = \sup\{\mu \in \mathcal{C}(F), \mu < \mu^{**}\}$. By the richness assumption, $\hat{\mu} > \underline{\mu}$, and by the same assumption, $\hat{\mu} > \mathbb{E}(\mu|N; \hat{\mu}, \theta)$, so that the DM suffers a reputational

loss from stopping when her belief equals $\hat{\mu}$. Since the social loss from continuing at $\hat{\mu}$ is bounded, there exists $\hat{\alpha}$ such that for all larger values of α , $U(Y, \mu) > U(N, \mu, \theta)$ for every $\mu \geq \hat{\mu}$ and for every θ . Thus for any $\alpha > \hat{\alpha}$, $\bar{\mu}^* < \hat{\mu}$. By the definition of $\hat{\mu}$, the interval $(\bar{\mu}^*, \hat{\mu}]$ has positive F -measure, and there is inefficient continuation on this interval, thereby proving (1). The remaining parts of the proposition are immediate from the characterization in Proposition 1.

A.5 Comparative Statics

We derive the comparative statics of a small increase in α , from α_0 to α_1 , upon any equilibrium threshold, μ_0^* .¹⁸ Observe that $\tilde{U}(N, x, \theta) = \gamma \mathbb{E}(\mu | N; \mu_0^*, \theta_0^*) + \nu(0)$ does not depend upon α , while $\tilde{U}(Y, x)$ does. To take account of this dependence, we re-write it as $\tilde{U}(Y, x; \alpha)$:

$$\tilde{U}(Y, x; \alpha) := x \left[\gamma + \frac{v}{\alpha} \right] - \frac{c}{\alpha} + \nu(0). \quad (\text{A.17})$$

An equilibrium (μ_0^*, θ_0^*) at the original parameter value, α_0 , satisfies

$$\tilde{U}(Y, \mu_0^*; \alpha_0) = \tilde{U}(N, x, \theta). \quad (\text{A.18})$$

The change in parameters only affects the function on the left-hand side of the above since the right-hand side is independent of α . Recall that $\mu^{**} = \frac{c}{v}$, and observe from (A.17) that $\tilde{U}(Y, \mu^{**}; \alpha) = \gamma \mu^{**} + \nu(0)$, which is independent of α . Finally, the intercept of A.17 on the vertical axis, $\tilde{U}(Y, 0; \alpha)$, is increasing in α . Thus an increase in α causes $\tilde{U}(Y, x)$ to swivel around the point $(\mu^{**}, \gamma \mu^{**} + \nu(0))$, becoming flatter. Consequently, $\tilde{U}(Y, x; \alpha_1) > \tilde{U}(Y, x; \alpha_0)$ for every $x < \mu^{**}$. Since $\mu_0^* < \mu^{**}$ by Proposition 1, we deduce that:

$$\tilde{U}(Y, \mu_0^*; \alpha_1) - \tilde{U}(N, x, \theta) > 0. \quad (\text{A.19})$$

If the initial equilibrium is left-stable, then there exists an interval $[\mu^-, \mu_0^*)$ such that for all x in this interval, $\tilde{U}(Y, x; \alpha_0) - \tilde{U}(N, x, \theta) < 0$, for all θ — including $\theta = 0$, where $\tilde{U}(N, x, \theta)$ is minimized. Since $\tilde{U}(Y, x; \alpha)$ is continuous in α , $\tilde{U}(Y, \mu^-; \alpha_1) - \tilde{U}(N, \mu^-, 0) < 0$ if α_1 is sufficiently close to α_0 .

The correspondence $\tilde{U}(Y, x; \alpha_1) - \tilde{U}(N, x, \theta), \theta \in [0, 1]$ is upper-hemicontinuous and convex-valued. We have established that it is strictly negative at μ^- and strictly positive at

¹⁸The comparative statics of a small increase in γ , keeping α fixed, are obviously the same.

μ_0^* . Thus, there must be a (μ_1^*, θ_1^*) , where $\mu_1^* \in (\mu^-, \mu_0^*)$, such that this correspondence takes the value 0. We have therefore proved that, if the original equilibrium is left-stable, a small increase in α reduces the equilibrium threshold.

A similar argument establishes that if the equilibrium threshold μ_0^* is unstable and without a mass point, then a small increase in α must result in a larger equilibrium threshold. Finally, if there is mass point and mixing at the equilibrium (μ_0^*, θ_0^*) , so that $\theta_0^* < 1$ then the mixing probability following an increase in α must be higher, the threshold μ_0^* being unaffected.

A.6 Proof of Proposition 5

We show that when $v + \gamma\alpha < 0$, equilibrium is unique — the rest of the proposition has been proved in the paper. Let $\mu^* < \bar{\mu}$ be an equilibrium threshold, and let θ^* be the cancellation probability at the threshold, so that $\mathbb{E}(\mu|N; \mu^*, \theta^*) := \tilde{\mu} \in (\mu^*, \bar{\mu})$. At $\tilde{\mu}$, the reputational payoffs from stopping and continuing are identical, so that stopping can only be optimal if it is socially optimal, and so $\tilde{\mu} \leq \mu^{**}$. However, $\tilde{\mu}$, being the truncated expectation, must be no less than the unconditional expectation, $\mathbb{E}(\mu)$, which equals p . Since $p > \mu^{**}$ by assumption, we have a contradiction. This establishes that there cannot be any equilibrium with an interior threshold μ^* .

B Appendix: Proofs relating to Section 3

B.1 Proof of Lemma 1

The derivative of $\pi^\dagger(\mu)$ is

$$\frac{q(1-p)p(1-q)}{[q(1-p)\mu + (1-q)p(1-\mu)]^2} > 0.$$

The numerator in the above expression does not depend on π , and the denominator is decreasing, since the derivative of $q(1-p)\mu + (1-q)p(1-\mu)$ equals $q-p > 0$. Thus the derivative of π^\dagger is strictly decreasing. Since π^\dagger is strictly concave, with $\pi^\dagger(0) = 0$ and $\pi^\dagger(1) = 1$, $\pi^\dagger(\mu) > \mu$ for any $\mu \in (0, 1)$.

B.2 Proof of Proposition 3

We prove the proposition without Assumption 3, and allow for mass points in the distribution F . In this case, (8) must be replaced by

$$\tilde{U}(N, x, \theta) := \gamma \mathbb{E}(\mu|N; x, \theta) + \nu(0), \quad (\text{B.20})$$

where $\mathbb{E}(\mu|N; x, \theta)$ is defined in (4).

Instead of Assumption 3, we make the following mild assumption.

Assumption 4 *Let $((\mu^*, 0), \rho^*)$ be any pure strategy equilibrium of G where the DM stops with probability 0 at the threshold. Either (1) or (2) below hold:*

1. *There exists an open interval $(\tilde{\mu}, \mu^*)$ such that F is strictly increasing on this interval.*
2. *There is no mass point at μ^* .*

The assumption will be violated only if the graph of the payoff $\tilde{U}(N, \mu, \theta)$ has a right-angle at μ^* , and if the straight line $\tilde{U}(Y, \mu)$ is tangent to this graph at μ^* . Thus, generically, the assumption will be satisfied. Figure 4, at the end of this proof, illustrates an example where the assumption is violated.

In the game G^0 , an equilibrium strategy for the DM consists of a pair (μ_0, θ_0) , where θ_0 is her probability of stopping at the threshold μ_0 . The corresponding equilibrium strategy for the observer is ρ_0 . The strategy profile $((\mu_0, \theta_0); \rho_0)$ is an equilibrium of G^0 if and only if it satisfies

$$\rho_0 = \tilde{U}(N, \mu_0, \theta_0) = \tilde{U}(Y, \pi^\dagger(\mu_0)). \quad (\text{B.21})$$

If there is a mass point at the belief threshold μ_0 , the corresponding value θ_0 is unique, since the value of the right hand side of the equilibrium condition in B.21 is uniquely defined by μ_0 . If there is no mass point at μ_0 , θ_0 is inconsequential, and we may fix it at some value, $\tilde{\theta}$. Thus, any equilibrium in G^0 defines a unique pair (μ_0, θ_0) .

First, we extend our definition of equilibrium in the games with higher-order beliefs about overconfidence G^n , $n > 0$, to allow for mass points in the distribution of F . Now, we need to allow for mixing if there is a mass point at a threshold belief. For $n \in \mathcal{O}$, the strategy of the DM is given by a pair (μ_n, θ_n) , where θ_n denotes the probability with which she cancels the project if her belief equals the threshold, μ_n . We now turn to determining these values given the equilibrium $((\mu_0, \theta_0); \rho_0)$ in G^0 .

Lemma 1 shows that $\mu_0 < \pi^\dagger(\mu_0)$, which, together with (B.21), implies that $\tilde{U}(N, \mu_0, \theta_0) > \tilde{U}(Y, \mu_0)$. Since $\tilde{U}(Y, x)$ is a strictly increasing affine function of x , it has a strictly increasing affine inverse, and we let $\tilde{U}_Y^{-1}(\cdot)$ denote the inverse. Define μ_1 to be the unique value satisfying $\tilde{U}(Y, \mu_1) = \rho_0$, or equivalently, $\mu_1 = \tilde{U}_Y^{-1}(\rho_0)$. Since $\rho_0 = \tilde{U}(N, \mu_0, \theta_0) > \tilde{U}(Y, \mu_0)$, this implies that $\mu_1 > \mu_0$. Observe that θ_1 can be arbitrary if F has a mass point at μ_1 , in which case there exist a continuum of equilibrium strategies (μ_1, θ_1) , $\theta_1 \in [0, 1]$, in G^1 , corresponding to the equilibrium in G^0 . Any pair (μ_1, θ_1) generates a distinct and unique ρ_2 in G^2 , as follows:

$$\rho_2 = \tilde{U}(N, \mu_1, \theta_1).$$

Similarly, equilibria in higher-order games are defined recursively as follows. For any $n \in \mathcal{O}$, $n > 1$, given any ρ_{n-1} , the threshold μ_n is defined by

$$\tilde{U}(Y, \mu_n) = \rho_{n-1}. \quad (\text{B.22})$$

Observe that the above defines a unique μ_n , since $\tilde{U}(Y, x)$ is a strictly increasing function. The corresponding θ_n can take any value in $[0, 1]$ if F has a mass point at μ_n , and equals $\tilde{\theta}$ otherwise. Given any pair (μ_n, θ_n) , where n is an odd number, ρ_{n+1} is uniquely defined by

$$\rho_{n+1} = \tilde{U}(N, \mu_n, \theta_n). \quad (\text{B.23})$$

Given any equilibrium $((\mu_0, \theta_0); \rho_0)$ in G^0 , equations (B.22) and (B.23) define (possibly multiple, if F has a mass points at some μ_n , $n \geq 1$) sequences

$$\left(((\mu_0, \theta_0); \rho_0), ((\mu_n, \theta_n))_{n \in \mathcal{O}}, (\rho_n)_{n \in \mathcal{E}} \right).$$

These satisfy the inequalities

$$\tilde{U}(N, \mu_{n+2}, \theta_{n+2}) \geq \tilde{U}(Y, \mu_{n+2}) = \tilde{U}(N, \mu_n, \theta_n) \geq \tilde{U}(Y, \mu_n), \quad n \in \{-1\} \cup \mathcal{O} \quad (\text{B.24})$$

where we let $(\mu_{-1}, \theta_{-1}) := (\mu_0, \theta_0)$.

Define the following order on pairs in \mathbb{R}^2 . $(\mu, \theta) \succ (\mu', \theta')$ if either $\mu > \mu'$, or $\mu = \mu'$ and $\theta > \theta'$. $(\mu, \theta) \succeq (\mu', \theta')$ if either $(\mu, \theta) > (\mu', \theta')$ or $(\mu, \theta) = (\mu', \theta')$. We show now that the sequence $(\mu_n, \theta_n)_{n \in \mathcal{O}}$ is weakly increasing according to the above defined order.

We have established that $\mu_1 > \mu_{-1}$. Now consider $n \in \mathcal{O}$, $n \geq 3$, and assume, by the induction hypothesis, that the sequence (μ_m, θ_m) is weakly increasing for all $m < n$, $m \in$

$\{-1\} \cup \mathcal{O}$. Then $(\mu_{n-2}, \theta_{n-2}) \succeq (\mu_{n-4}, \theta_{n-4})$. Since, by (B.20), $\tilde{U}(N, \mu, \theta)$ is increasing in μ and θ , equation B.23 implies that $\rho_{n-1} \geq \rho_{n-3}$. Thus B.22 establishes that $(\mu_n, \theta_n) \succeq (\mu_{n-2}, \theta_{n-2})$.

We have established that (μ_n) is an increasing sequence, and since it is bounded, it must converge to some value, denoted μ^∞ . There are two possibilities:

- (i) either $\mu_n = \mu^\infty$ for some n , or
- (ii) $\mu_n < \mu^\infty$, $\forall n$.

We will show that $(\mu^\infty, 0)$ is an equilibrium strategy for the DM in the game with common priors, G , in both the cases above, by showing that $\tilde{U}(Y, \mu^\infty) = \tilde{U}(N, \mu^\infty, 0)$.

Consider case (i). Let m denote the smallest value of n such that $\mu_n = \mu^\infty$. By the definition of m , and using the inequalities in B.24, we deduce that

$$\tilde{U}(N, \mu_m, \theta_m) = \tilde{U}(Y, \mu_m) = \tilde{U}(N, \mu_{m-2}, \theta_{m-2}) > \tilde{U}(Y, \mu_{m-2}).$$

Thus $\tilde{U}(N, \mu_m, \theta_m) = \tilde{U}(N, \mu_{m-2}, \theta_{m-2})$. But $\tilde{U}(N, \mu, \theta)$ can be constant on the interval $(\mu_{m-2}, \mu_m]$ only if this interval has zero F -measure, and if $\theta_m = 0$ if there is a mass point at μ_m . In other words, $U(N, \mu, \theta)$ must be flat along this interval, and, if there is a mass point of F at μ^∞ , the corresponding value of θ must equal zero. Thus $(\mu^\infty, 0)$ is a pure strategy equilibrium strategy in G , and we have also proved that it is left-stable, since $U(N, \mu, \theta)$ is constant on the interval (μ_{m-2}, μ^∞) , while $\tilde{U}(Y, \mu)$ is strictly increasing.

Next, consider case (ii). Observe that $\tilde{U}(N, \mu, 0)$ defines a left-continuous function of μ . Therefore, $\tilde{U}(Y, \mu_n) \leq \tilde{U}(N, \mu_n, \theta_n)$, and $\mu^\infty > \mu^n$ for all n , imply that $\tilde{U}(Y, \mu^\infty) \leq \tilde{U}(N, \mu^\infty, 0)$. We now prove the reverse inequality. Since the sequence (μ_n) is Cauchy, and since $\tilde{U}(Y, \mu)$ is an affine function, the sequence $(\tilde{U}(Y, \mu_n))$ is a Cauchy sequence. Thus for any $\epsilon > 0$, there exists N such that $\tilde{U}(Y, \mu_{n+2}) - \tilde{U}(Y, \mu_n) < \epsilon$ if $n > N$. Since $\tilde{U}(Y, \mu_{n+2}) = \tilde{U}(N, \mu_n, \theta_n)$, $\tilde{U}(N, \mu_n, \theta_n) - \tilde{U}(Y, \mu_n) < \epsilon$ if $n > N$. Since ϵ was arbitrary, and both functions are left-continuous, $\tilde{U}(Y, \mu^\infty) \geq \tilde{U}(N, \mu^\infty, 0)$. We conclude that $\tilde{U}(Y, \mu^\infty) = \tilde{U}(N, \mu^\infty, 0)$, so that $(\mu^\infty, 0)$ must be an equilibrium threshold of G , the game with common priors.

Having established convergence to a pure strategy equilibrium, we now show that this must correspond to the threshold μ_+^* defined in (15). The equilibrium $((\mu_+^*, 0), \rho_+^*)$ is left-stable by definition: recall $\mu_+^* = \min\{\mu > \mu_0 : \tilde{U}(Y, \mu) = \tilde{U}(N, \mu, 0)\}$, and $\tilde{U}(Y, \mu_0) < \tilde{U}(N, \mu_0, \theta_0)$. To prove convergence, we invoke Assumption 4, so that there are two possible cases: either there is no gap in the support of F to the left of μ_+^* , or this condition fails but there is no mass point at μ_+^* .

Suppose that there is no gap, i.e. there exists an open interval $(\tilde{\mu}, \mu_+^*)$ such that F is strictly increasing on this interval. This implies that $\tilde{U}(N, \mu_+^*, 0) > \tilde{U}(N, \mu, \theta)$ if $\mu < \mu_+^*$. By the definition of μ_+^* , $\tilde{U}(Y, \mu) < \tilde{U}(N, \mu, 0)$ if $\mu \in [\mu_0, \mu_+^*)$. Thus we have the following inequality for any $\mu \in [\mu_0, \mu_+^*)$ and any θ :

$$\tilde{U}(N, \mu_+^*, 0) > \tilde{U}(N, \mu, \theta) > \tilde{U}(Y, \mu).$$

We now show that sequence (μ_n) is *strictly* increasing and that $\mu_n < \mu_+^*$ for all $n \in \mathcal{O}$. Since $\mu_n = \tilde{U}_Y^{-1}[\tilde{U}(N, \mu_{n-2}, \theta_{n-2})]$ while $\mu_+^* = \tilde{U}_Y^{-1}[\tilde{U}(N, \mu_+^*, 0)]$, and the function \tilde{U}_Y^{-1} is strictly increasing, it follows from $\mu_0 < \mu_+^*$ that $\mu_n < \mu_+^*$ for every $n \in \mathcal{O}$.

Now, suppose that there exists an interval $(\tilde{\mu}, \mu_+^*)$ such that F is constant on this interval. Assume therefore that there is no mass point at μ_+^* . Now for any $\mu \in [\mu_0, \mu_+^*)$ and any θ :

$$\tilde{U}(N, \mu_+^*, \tilde{\theta}) \geq \tilde{U}(N, \mu, \theta) > \tilde{U}(Y, \mu).$$

Consequently, it follows that if $\mu_n < \mu_+^*$, then $\mu_{n+2} \leq \mu_+^*$. Suppose that $\mu_n = \mu_+^*$. Since there is no mass point at μ_+^* , it follows that $\mu_{n+2} = \mu_+^*$, and thus the sequence becomes constant, and therefore $\mu_n < \mu_+^*$ for every $n \in \mathcal{O}$.

Thus, we have established that the sequence (μ_n) is an increasing sequence that converges to an equilibrium pure strategy, and also established that $\mu_n \leq \mu_+^*$ for all n . This completes the proof of the proposition.

Finally, Figure 4 provides an example where Assumption 4 is violated. At $\underline{\mu}^*$, there is a mass point, as well as gap in the support of F immediately to the left of $\underline{\mu}^*$. Furthermore, the straight line depicting the payoff $\tilde{U}(Y, \mu)$ touches the graph of $\tilde{U}(N, \mu, \theta)$ at the point $(\underline{\mu}^*, 0)$.

Observe from the figure that $\mu_1 = \underline{\mu}^*$, since $\tilde{U}(N, \mu, \theta)$ is flat on the interval $[\mu_0, \underline{\mu}^*)$. Thus, one possibility is that $(\mu_n, \theta_n) = (\underline{\mu}^*, 0)$ for all $n \in \mathcal{O}$ so that $\rho_n = \rho_0$ for all $n \in \mathcal{E}$. The other possibility is that $\theta_n > 0$ and $\mu_n = \underline{\mu}^*$ for some value of n , so that the sequence converges to the equilibrium $(\bar{\mu}^*, \bar{\rho}^*)$.

Thus, the sequence may converge to the higher equilibrium in G^0 when Assumption 4 is violated.

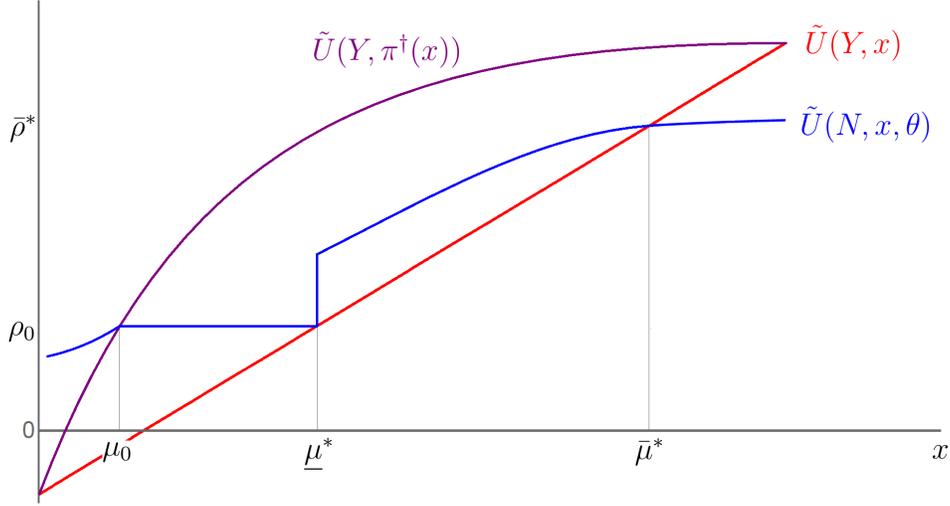


Figure 4: Violation of Assumption 4: F has a mass point at $\underline{\mu}^*$ and a gap in the support to the left.

B.3 Proof of Proposition 4

We show first that the game G^0 is supermodular. If the DM chooses a threshold x so that her threshold belief is $\pi^\dagger(x)$, and if she cancels with a probability θ at x , then the observer's best response correspondence $\hat{\rho}(x) := \{\rho(x, \theta), \theta \in [0, 1]\}$ is increasing in x , i.e. $x > x' \Rightarrow \min \hat{\rho}(x) > \max \hat{\rho}(x')$.

Let the DM's best response threshold be denoted by $\hat{\mu}(\rho) \in [0, 1]$, and satisfy

$$\tilde{U}(Y, \pi^\dagger(\hat{\mu})) = \alpha \rho.$$

Then $\hat{\mu}(\rho)$ is uniquely defined and increasing since both π^\dagger and $\tilde{U}(Y, \cdot)$ are strictly increasing functions. Thus the game G^0 is supermodular.

Let $(\underline{\mu}_0, \underline{\rho}_0)$ denote the smallest Nash equilibrium of G^0 , and let $(\bar{\mu}_0, \bar{\rho}_0)$ denote the largest Nash equilibrium. Since the game is supermodular, the extremal equilibria are in pure strategies, and any rationalizable strategy for the observer lies in the interval $[\underline{\rho}_0, \bar{\rho}_0]$.

From (9), observe that for any interior value of μ , $\pi^\dagger(\mu) \rightarrow 1$ as $q \rightarrow 1$. Thus, for q sufficiently large, any equilibrium threshold in the game G^0 is strictly less than $\underline{\mu}^*$, and therefore $\bar{\mu}_0 < \underline{\mu}^*$.

Consider a rationalizable strategy of the DM in the game G^1 . This must be a best

response to a probability distribution over the rationalizable strategies of the observer in G^0 , i.e. to a distribution with support $[\underline{\rho}_0, \bar{\rho}_0]$. Since the DM's payoff from stopping is linear in ρ , her rationalizable strategies in G^1 are the best responses to elements the set $[\underline{\rho}_0, \bar{\rho}_0]$. Thus the rationalizable thresholds μ_1 in G^1 satisfy

$$\tilde{U}(Y, \mu_1) = \alpha \rho, \quad \rho \in [\underline{\rho}_0, \bar{\rho}_0].$$

For $n \in \mathcal{O}$, let $\underline{\mu}_n$ denote the smallest equilibrium threshold for the DM in G^n , and let $\bar{\mu}_n$ denote the largest. Any rationalizable threshold, μ_n , must belong to the interval spanned by these two thresholds. Similarly, for $n \in \mathcal{E}$, let $\underline{\rho}_n$ denote the smallest equilibrium action for the observer in G^n , and let $\bar{\rho}_n$ denote the largest. Again, any rationalizable action ρ_n belongs to the interval spanned by these two actions.

Consider the sequences (μ_n) and (ρ_n) induced by $(\underline{\mu}_0, \underline{\rho}_0)$. By Proposition 3, these must converge to $\underline{\mu}^*$ and $\underline{\rho}^*$, respectively. Similarly, the sequences induced by $(\bar{\mu}_0, \bar{\rho}_0)$ also converge to the same limits. Since any rationalizable sequence μ_n (resp. ρ_n) is sandwiched between the two induced equilibrium sequences, this proves the proposition.

C Appendix: Extensions to the Model with Common Priors

This appendix presents three extensions to the basic model, with common knowledge of common priors. First, we show that our analysis extends when the DM has private information on her ability. Second, we analyze the decision to initiate the project. Finally, we consider the possibility that the outcome of the project is inconclusive, and also extend the analysis to multiple periods.

C.1 When the DM knows her ability

How is our analysis affected if the DM knows her ability? Suppose that nature's choice of the DM's type in $\{H, L\}$ is observed by the DM at $t = 0$, but not by the observer. We continue to assume that the project's quality is not observed by the DM — since both p_H and p_L are interior, both types of the DM are uncertain about the quality.

Suppose the DM observes the private signal s with likelihood ratio

$$\ell(s) := \frac{h(s|G)}{h(s|B)}, \quad (\text{C.25})$$

where $h(s|\omega)$ denotes the value, evaluated at s , of the probability density function¹⁹ governing the distribution of signals when the project's quality is ω . Then, the type $\tau \in \{G, B\}$ of the DM holds the posterior belief

$$\mu_\tau(s) := \frac{p_\tau \ell(s)}{p_\tau \ell(s) + 1 - p_\tau}.$$

Let F_H and F_L denote the associated distribution of posterior beliefs for the two types. Define $F := \lambda F_H + (1 - \lambda)F_L$. From the point of view of the outside observer, who does not observe the DM's type, F describes the distribution of beliefs of the DM at the beginning of period $t = 2$. Observe that, given any posterior belief, μ , about project quality held by the DM at $t = 1$, the DM's own type is irrelevant in the continuation game. Consequently, both types of the DM will have the same cut-off belief, μ^* , that satisfies the equilibrium condition (5).

The analysis here may be contrasted with that in Majumdar and Mukand (2004), where a DM of type H knows for sure that her project is good (i.e. $p_H = 1$), and therefore does not update her belief on the likelihood of the project succeeding, regardless of the private signal she receives. Thus, a high-ability DM never cancels a project. In other words, changing one's mind is a sure sign of incompetence. In contrast, our analysis does not equate competence with infallibility. There are facts that would change even the most competent DM's mind.

C.2 Initiating the Project

Let us briefly discuss an extension of our model, where the DM with prior p decides at $t = 0$ whether to initiate the project at a cost of $k > 0$. Consider first the choice of a social planner, i.e. a DM who has no reputational concerns. Under the socially optimal policy, the value of project at $t = 1$ equals $\mu v - c$ if $\mu > \mu^{**}$, and 0 otherwise. Thus, the expected value at $t = 0$ from initiating the project equals

$$\tilde{V}(\mu^{**}) := [\mathbb{E}(\mu|\mu > \mu^{**})v - c] [1 - F(\mu^{**})] - k,$$

¹⁹Or the atom at s , if the cumulative distribution function of the signal, conditional on ω , has jump at s .

where the expectation is taken with respect to F .

The project is socially optimal at $t = 0$ if $\tilde{V}(\mu^{**}) \geq 0$. Observe that the option to terminate the project at $t = 1$ makes it more profitable to start it in the first place. Even if the prior were $p < \frac{c+k}{v}$, the expected payoff from the project under the planner policy might still be positive. This option value is well known from the literature on multiarmed bandits.

Now consider an equilibrium of the game with project initiation. In period 0, the DM must decide whether or not to start the project. In an equilibrium with threshold μ^* and no mass point at the threshold, the project is worth initiating at $t = 0$ if

$$\tilde{V}(\mu^*) := [\mathbb{E}(\mu|\mu > \mu^*)v - c] [1 - F(\mu^*)] - k \geq 0.$$

From Proposition 1, we know that, for a reputationally concerned DM, the expected social return from the project will be lower than for the social planner, i.e. $\tilde{V}(\mu^*) < \tilde{V}(\mu^{**})$ since the termination decision is inefficient when the richness assumption is satisfied. Thus reputationally concerned DM will be less likely to initiate the project at $t = 0$, since she knows that she will continue it after some private signals at which a social planner would stop. Thus, the fact that information about the project is privately received *reduces the option value of termination* and makes risky projects less attractive.

Observe that the DM's project initiation decision is constrained efficient — the DM takes the decision that is socially optimal at $t = 0$, given that she is going to behave inefficiently at $t = 1$. Intuitively, the absence of private information at $t = 0$ leads to efficient decisions, even though private information in the future implies future inefficiency.

C.3 Information Revelation Disciplines the DM

Our analysis emphasizes the informational difference between the two experiments, project continuation and termination. The former fully reveals project quality, while the latter hides it. The purpose of this section is to show that if the outcome may be inconclusive when the project is continued, then this aggravates the DM's incentives to continue bad projects. Thus information revelation disciplines the DM. We also extend the model to many periods to examine the dynamics of stubbornness, and show that the DM exhibits behavior that is similar to that induced by the sunk cost fallacy.

We amend the model as follows. Suppose now that when the project is continued, its quality is only revealed with positive probability. Specifically, if the DM continues the project at $t = 1$, then a move of nature determines subsequent events. With probability λ ,

the outcome of the project, success or failure, is publicly realized. With probability $1 - \lambda$, the outcome of the project is not realized, and we assume, for simplicity, that the project is scrapped, so that its continuation social value is zero.

Since the move of nature is independent of the underlying state of the project, the DM's assessment, at $t = 1$, that the project will succeed, conditional on a public outcome, remains equal to μ . There is now a third possible public event at $t = 2$, namely the one where the DM continues the project at $t = 1$ but the project's outcome is not realized. The observer's strategy requires specifying his action in this event, and must equal his belief about the DM's ability conditional on this event.

Suppose that the DM follows a threshold strategy: she continues the project if $\mu \geq \mu^*$. We make two simplifying assumptions: first, the distribution of F is atomless, and second, that the function $H(x) := (1 - \lambda)\mathbb{E}(\mu|\mu > x) - \mathbb{E}(\mu|\mu < x)$ is strictly decreasing in x .²⁰ The DM's payoff from continuation when her private belief is μ equals

$$\lambda \mu [v + \alpha \gamma] + (1 - \lambda) \alpha \gamma \mathbb{E}(\mu|\mu \geq \mu^*) - c + \alpha \nu(0). \quad (\text{C.26})$$

The payoff from terminating the project is, as before, $\alpha [\gamma \mathbb{E}(\mu|\mu < \mu^*) + \nu(0)]$. For μ^* to be an equilibrium threshold, these two payoffs must be equal when $\mu = \mu^*$. In contrast, the efficient solution is to terminate the project at $\mu^{**} = \frac{c}{\lambda v}$.

Keeping λv fixed, let us examine the effects of a decrease in λ . This does not affect the efficient threshold but has the effect of reducing μ^* , if the original equilibrium is stable. The reputational benefit from continuation is now proportional to $\lambda \mu^* + (1 - \lambda)\mathbb{E}(\mu|\mu \geq \mu^*)$, and this is decreasing in λ . If λ is high, so that it is more likely that there is conclusive evidence on the project's quality, the DM has less incentives to continue bad projects.

This has implications for the DM's incentives to continue the project when nearing the period of evaluation. Consider a DM who initiates a project at date 0, and who is up for re-election or re-appointment at the beginning of period T . In each period $\tau \in \{1, 2, \dots, T - 1\}$, the outcome of the project, i.e. its success or failure, is realized with probability λ , conditional on it not having been realized at any previous date. If the outcome of the project is not realized before, the project is terminated at date T . The flow cost of continuing the project is c per period. Suppose that the DM receives private information about the project at a single date τ , and must decide at this point whether to continue with it or terminate it. where continuation entails paying the cost c . Let $t = T - \tau$ denote the number of periods

²⁰The first assumption is for expositional convenience, while the second ensures that the game is super-modular, a property that is used for the comparative statics predictions.

remaining. The socially efficient threshold is independent of t and is given by

$$\mu_t^{**} = \frac{c}{\lambda v} := \mu^{**}.$$

The proof is by induction. When $t = 1$, we have already established that this is the case, from our previous analysis. Now suppose that $\mu_s^{**} = \mu^{**}$ for $s = 1, 2, \dots, t - 1$, and consider the situation with t periods remaining. If $\mu \geq \mu^{**}$, then it is profitable to run the project for one period, and also the continuation value with $t - 1$ periods remaining is also positive. On the other hand, if $\mu < \mu^{**}$, it is unprofitable to run the project for one period, and the continuation value is also negative.

Turning to equilibrium analysis, suppose that the DM receives her private information with t periods remaining. Assume for the moment that the DM makes an irrevocable decision to cancel the project today or continue till the terminal date.²¹ Let μ_t^* be her equilibrium threshold. This is characterized by the condition:

$$[1 - (1 - \lambda)^t]v + \alpha\gamma]\mu_t^*[-\frac{[1 - (1 - \lambda)^t]c}{\lambda} + (1 - \lambda)^t\alpha\gamma\mathbb{E}(\mu|\mu \geq \mu_t^*) = \alpha\gamma\mathbb{E}(\mu|\mu < \mu_t^*).$$

Inspecting the equilibrium condition, we see that the reputational benefit from continuation is proportional to

$$[1 - (1 - \lambda)^t][\mu_t^* - \mathbb{E}(\mu|\mu < \mu_t^*)] + (1 - \lambda)^t[\mathbb{E}(\mu|\mu \geq \mu_t^*) - \mathbb{E}(\mu|\mu < \mu_t^*)].$$

An decrease in t increases the weight on the second term above. The shorter the time horizon t , the greater the reputational benefit from continuation since it is more likely that the outcome of the project will not be realized. Thus, the DM is more reluctant to cancel a sub-optimal project towards the end of the horizon than at the beginning, since uncertainty more likely to be resolved by success/failure at the beginning. More precisely, if we focus on either of the extremal equilibria — either the most efficient one or the least efficient one — μ_t^* is decreasing in t . It is as though the DM exhibits the “sunk-cost fallacy” in her behavior, although this is driven by reputational concerns rather than irrationality.

²¹This assumption is made to simplify exposition — in equilibrium, the DM will never change her mind, even if the decision is revocable.

References

- BARBER, B. M., AND T. ODEAN (2001): “Boys will be boys: Gender, overconfidence, and common stock investment,” *The Quarterly Journal of Economics*, 116(1), 261–292.
- BÉNABOU, R., AND J. TIROLE (2002): “Self-confidence and personal motivation,” *The Quarterly Journal of Economics*, 117(3), 871–915.
- CARLSSON, H., AND E. VAN DAMME (1993): “Global Games and Equilibrium Selection,” *Econometrica*, 61(5), 989–1018.
- CHEN, Y.-C., A. DI TILLIO, E. FAINGOLD, AND S. XIONG (2017): “Characterizing the Strategic Impact of Higher-order Beliefs,” *Review of Economic Studies*, 84, 1424–1471.
- CHO, I.-K., AND D. KREPS (1987): “Signaling Games and Stable Equilibria,” *Quarterly Journal of Economics*, 102(2), 179–221.
- DEKEL, E., D. FUDENBERG, AND D. K. LEVINE (2004): “Learning to Play Bayesian Games,” *Games and Economic Behavior*, 46(2), 282–303.
- DEKEL, E., D. FUDENBERG, AND S. MORRIS (2007): “Interim correlated rationalizability,” *Theoretical Economics*, 2(1), 15–40.
- DIKÖTTER, F. (2010): *Mao’s great famine: The history of China’s most devastating catastrophe, 1958-1962*. Bloomsbury Publishing USA.
- DOWNES, G. W., AND D. M. ROCKE (1994): “Conflict, agency, and gambling for resurrection: The principal-agent problem goes to war,” *American Journal of Political Science*, pp. 362–380.
- DUR, R. A. (2001): “Why do policy makers stick to inefficient decisions?,” *Public Choice*, 107(3), 221–234.
- ESPONDA, I., AND D. POUZO (2016): “Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models,” *Econometrica*, 84(3), 1093–1130.
- FREIXAS, X., B. M. PARIGI, AND J.-C. ROCHET (2000): “Systemic risk, interbank relations, and liquidity provision by the central bank,” *Journal of Money, Credit and Banking*, pp. 611–638.

- GEANAKOPOLOS, J. (2010): “The leverage cycle,” *NBER macroeconomics annual*, 24(1), 1–66.
- GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): “Performance in competitive environments: Gender differences,” *The Quarterly Journal of Economics*, 118(3), 1049–1074.
- GROSSMAN, S. J. (1981): “The informational role of warranties and private disclosure about product quality,” *The Journal of Law and Economics*, 24(3), 461–483.
- HARRISON, J. M., AND D. M. KREPS (1978): “Speculative investor behavior in a stock market with heterogeneous expectations,” *The Quarterly Journal of Economics*, 92(2), 323–336.
- HOLM, H. J. (2000): “Gender-based focal points,” *Games and Economic Behavior*, 32(2), 292–314.
- HOLMSTRÖM, B. (1999): “Managerial Incentive Problems: A Dynamic Perspective,” *Review of Economic Studies*, 66(1), 169–182.
- KERSHAW, I. (2000): *Hitler: 1889-1936 Hubris*. WW Norton & Company.
- KROLL, M. J., L. A. TOOMBS, AND P. WRIGHT (2000): “Napoleon’s tragic march home from Moscow: Lessons in hubris,” *The Academy of Management Executive*, 14(1), 117–128.
- MAJUMDAR, S., AND S. W. MUKAND (2004): “Policy gambles,” *The American Economic Review*, 94(4), 1207–1222.
- MALMENDIER, U., AND G. TATE (2005): “CEO overconfidence and corporate investment,” *The Journal of Finance*, 60(6), 2661–2700.
- (2008): “Who makes acquisitions? CEO overconfidence and the market’s reaction,” *Journal of Financial Economics*, 89(1), 20–43.
- MEYER, M. A., AND J. VICKERS (1997): “Performance Comparisons and Dynamic Incentives,” *Journal of Political Economy*, 105(3), 547–581.
- MILGROM, P. R. (1981): “Good news and bad news: Representation theorems and applications,” *The Bell Journal of Economics*, pp. 380–391.

- MILGROM, P. R., AND D. J. ROBERTS (1990): “Rationalizability, learning, and equilibrium in games with strategic complementarities,” *Econometrica*, 58, 1255–1277.
- MORCK, R., A. SHLEIFER, AND R. W. VISHNY (1990): “Do managerial objectives drive bad acquisitions?,” *The Journal of Finance*, 45(1), 31–48.
- MORRIS, S., AND H. S. SHIN (1998): “Unique equilibrium in a model of self-fulfilling currency attacks,” *American Economic Review*, pp. 587–597.
- OWEN, D., AND J. DAVIDSON (2009): “Hubris syndrome: An acquired personality disorder? A study of US Presidents and UK Prime Ministers over the last 100 years,” *Brain*, 132(5), 1396–1406.
- PRAT, A. (2005): “The wrong kind of transparency,” *American Economic Review*, 95(3), 862–877.
- ROLL, R. (1986): “The hubris hypothesis of corporate takeovers,” *Journal of Business*, pp. 197–216.
- RUBINSTEIN, A. (1989): “The Electronic Mail Game: Strategic Behavior under Almost Common Knowledge,” *American Economic Review*, 79(3), 385–391.
- SANTOS-PINTO, L., AND J. SOBEL (2005): “A model of positive self-image in subjective assessments,” *American Economic Review*, 95(5), 1386–1402.
- SCHEINKMAN, J. A., AND W. XIONG (2003): “Overconfidence and speculative bubbles,” *Journal of Political Economy*, 111(6), 1183–1220.
- SETHI, R., AND M. YILDIZ (2016): “Communication with unknown perspectives,” *Econometrica*, 84(6), 2029–2069.
- VAN DEN STEEN, E. (2004): “Rational overoptimism (and other biases),” *The American Economic Review*, 94(4), 1141–1151.
- WEINSTEIN, J., AND M. YILDIZ (2007): “A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements,” *Econometrica*, 75(2), 365–400.
- YILDIZ, M. (2003): “Bargaining Without a Common Prior—An Immediate Agreement Theorem,” *Econometrica*, 71(3), 793–811.