The Culture of Overconfidence*

V Bhaskar† Caroline Thomas‡

July 29, 2018

Abstract

Perceptions of overconfidence can exacerbate the tendency of reputationally concerned leaders to continue bad projects. Reputation concerns alone induce a bias towards inefficient continuation in a leader receiving information privately. When she is overconfident — holds a more favorable prior than observers — this tendency is aggravated. This remains true even when she is not really overconfident, but merely perceived to be so. Higher-order beliefs regarding overconfidence induce inefficient equilibrium selection even when there is "almost common knowledge" that the leader is not overconfident. This provides a novel perspective on how culture selects among equilibria: via higher-order 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, Navin Kartik, Stephen Morris, Muhamet Yildiz for helpful comments. We also thank Larry Samuelson (co-editor), and three anonymous referees. 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.

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

Why do corporate and political leaders persist with their pet projects despite adverse information? Margaret Thatcher was "not for turning": not when UK unemployment hit two million, not when there was widespread opposition to the poll tax. Mao intensified the Great Leap Forward, despite reports of widespread starvation in rural China. Over 30 million people died in the consequent famine (Dikötter (2010)). When Texas Instruments abandoned its foray into the home computer market, its stock price rose 22%. Statman and Sepe (1989) find that a company's share price rises, on average, after a project termination preceded by public bad news, a finding inconsistent with efficient termination decisions.

To the ancient Greeks, the answer was hubris. In Aeschylus' *The Persians* and Sophocles' *Antigone*, the protagonist kings persist in their chosen path, ignoring advice and omens, precipitating their demise. Historians invoke hubris to explain Napoleon's doomed march on Moscow (Kroll, Toombs, and Wright (2000)) and Hitler's expansion (Kershaw (2000)), and the mistakes of many US presidents and UK prime ministers (Owen and Davidson (2009)). Roll (1986), Morck, Shleifer, and Vishny (1990), and Malmendier and Tate (2008) argue that managerial overconfidence explains takeovers that reduce firm value.

Reputational concerns can induce a leader to continue an ineffectual project when she has correct beliefs regarding the quality of the project. This paper shows that overconfidence, perceptions of overconfidence, or higher-order beliefs regarding overconfidence, magnify the resulting inefficiencies. Thus, a culture where leaders are expected to be overconfident can have adverse consequences even when the leader in question is not overconfident.

Consider a leader who receives inconclusive *private* information regarding her project, and must decide whether to continue it to conclusion, or abandon it. She cares both about the social value of the project, and about an observer's evaluation of the project's quality, as this is correlated with her ability. Abandoning the project stops the arrival of further information; continuing reveals the underlying state publicly, resolving all uncertainty. This gives rise to a *disclosure value of continuation* — continuing the project publicly verifies the leader's private belief about its quality. Terminating it prevents any further learning, and only reveals that she deemed the project bad enough to warrant termination. The reputation cost faced by a marginally pessimistic leader incudes her to continue socially wasteful projects. Strategic complementarities link the leader's continuation decision and the outside observer's inference: if the leader is more stubborn and continues the project at worse beliefs, the observer's inference upon project termination is more adverse, increasing the reputational penalty for stopping. Consequently, there can be multiple equilibria, and an equilibrium where the project is continued more often (at higher beliefs) is more inefficient.

Suppose now that the leader is overconfident: she has a more optimistic prior on the project's quality than the observer. Because continuing the project reveals more information, it becomes even more appealing. In addition, an observer believing the leader to be overconfident draws even more negative inference about her private information if she terminates the project. Indeed, this more negative inference arises even when the leader is not overconfident, but is merely *perceived* to be so.

Taking this argument further, suppose that the leader is not overconfident, and not perceived to be overconfident. But she believes that she is perceived to be overconfident. More generally, there may be mutual knowledge that the leader is *not* overconfident up to many levels, but not common knowledge: at some level, there is a perception that the leader is (believed to be) overconfident. Equilibria in such a game are anchored to equilibria in the game where the leader is *in fact* overconfident, exacerbating the leader's tendency to continue projects. This is true also in the limit, under"almost common knowledge" that the leader is not overconfident. If potential overconfidence is large enough, then there is a unique rationalizable outcome: the most inefficient equilibrium of the game with common knowledge of a common prior.

These results show that a culture where corporate 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 culture ensures that the most inefficient equilibrium is selected.

Our theory highlights a novel channel though which cultural stereotypes might determine outcomes: via higher-order beliefs. 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.[1]

One possible normative implication of our results is that individuals who are not stereo-

[1]For instance, in a society where women defer to men, it is focal to coordinate on the men-preferred equilibrium in the battle of the sexes. Holm (2000) presents experimental evidence supporting this.

typically expected to be overconfident — such as women — might prove better leaders, as they are under less pressure to persist with unprofitable projects.[2]

Related Literature: Dur (2001) and Majumdar and Mukand (2004) argue that leaders are reluctant to abandon failing projects because this signals incompetence. They assume that competent leaders always select good projects; incompetent leaders sometimes select bad ones, and receive subsequent information on project quality. Cancelling the project reveals that a leader has changed her mind, and that she must be unskilled. Majumdar and Mukand find that there is a unique equilibrium, with excessive continuation by the unskilled leader.[3]

The key difference is our ("Keynesian") assumption that no leader is infallible, and it is therefore wise to sometimes reverse one's decisions. Consequently, the inference drawn by the observer when the leader cancels the project is not invariant, but depends upon the leader's strategy. This may result in multiple equilibria, and higher-order beliefs can play a role in selecting equilibria. This property is robust to assuming the leader has private information on her ability, provided she is not infallible — see Bhaskar and Thomas (2018). Majumdar and Mukand argue that if the incompetent leader is overconfident, policy persistence is reduced. Since overconfidence affects the decisions of both ability-types in our setting, the findings are the opposite.[4]

Models with heterogeneous priors are used to examine a range of issues — examples include speculative trade (Harrison and Kreps (1978)) or the exchange of opinions (Sethi and Yildiz (2016)). The strategic implications of mis-specified beliefs are explored by Chen, Di Tillio, Faingold, and Xiong (2017). Dekel, Fudenberg, and Levine (2004) argue against the use of Nash equilibrium as a solution concept in games with heterogeneous priors, and advocate the use of non-equilibrium notions, such as rationalizability. Our analysis takes on board this criticism, and our main results hold for rationalizable strategies. Dekel, Fudenberg, and Morris (2007) provide a connection between Bayes Nash equilibria in games without a common prior and interim correlated rationalizability.

[2]Barber and Odean (2001) examine common stock investments, and find that men trade 45 percent more than women and earn lower returns, a difference they attribute to men's overconfidence. Gneezy, Niederle, and Rustichini (2003) show that girls perform worse than boys in a competitive contest, especially when they compete against boys. Their performance is no different in a non-competitive setup. If girls are less confident in their abilities, they will exert less effort in a winner-take-all contest.

[3]Unlike our model, the leader knows her ability when making the continuation decision.

[4]Holmström (1999), Section 3.2, is less related, since project quality is *uncorrelated* with managerial ability, and the manager cares *only* about her reputation.

Our results on the effects of higher-order beliefs are reminiscent of the electronic mail game (Rubinstein (1989)) and global games (Carlsson and van Damme (1993), Morris and Shin (2003)). The literature on global games has pursued a universal theory of equilibrium selection. Weinstein and Yildiz (2007) stymie this agenda, by showing that any rationalizable action becomes uniquely rationalizable with suitable higher-order beliefs. Our paper offers a context-specific theory of selection, based on the culture surrounding leadership.

1 A Model with Common Priors

We study the interaction between a leader — a manager or a politician — who undertakes a project, and an outside observer — the firm's shareholders/potential employers, or voters — who evaluates the leader's ability. In period zero, nature chooses the ability $\tau \in \{H, L\}$ of the leader and the quality $\omega \in \{G, B\}$ of her project. An able leader is more likely to have a good project. Neither player observes the leader's ability or the project's quality, and they share a common prior p that the project is good.

In period one, the leader privately observes a signal of project quality, and decides whether to continue or terminate the project. If she continues (action Y), a cost c is incurred, and the project's outcome is publicly realized in period two: it succeeds if good, yielding a return v, and fails if bad, yielding nothing. If the leader terminates (action N), no payoff accrues, and there is no further learning about the project's quality.

At the end of period two, the observer chooses an action in $[0, 1]$. His optimal action equals his posterior belief about the project's quality, which we interpret as the leader's *reputation*. Appendix A.1 shows that if the observer's optimal action equals his posterior belief about the leader's ability, as in Holmström (1999), the analysis is identical.

The leader maximizes

$$\theta \times \text{project payoff} + \text{observer's action},$$

where $\theta > 0$ measures the intensity of her social concerns relative to reputational ones.

The private signal the leader observes at date one induces a cumulative distribution F over her posterior belief μ that the project is good. The socially optimal decision is to continue with the project if and only if $\mu > \mu^{**} := \frac{c}{v}$. For simplicity, assume:

Assumption 1 *F has an atomless distribution, and* $F(\mu^{**}) \in (0, 1)$.

Our results do not depend on F being atomless — see Bhaskar and Thomas (2018).

We analyze Perfect Bayesian Equilibria. A pure strategy σ for the leader maps her private posterior belief $\mu \in [0,1]$ to $\{Y, N\}$. A strategy for the observer specifies an action in $[0,1]$ at date two for each public event: success, failure, and project termination. Since the observer must take action 1 following a success and 0 following a failure, it suffices to specify his action when the project is cancelled, and we identify it with his strategy ρ. Sequential rationality implies that ρ must equal the observer's belief about project quality upon cancellation.

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

$$U(Y, \mu) := \mu\,[\theta v + 1] - \theta c. \tag{1}$$

With probability μ, the project succeeds, so that the social value rises by v and the observer's belief jumps to 1. With complementary probability the project fails and the observer's belief falls to 0. Thus the payoff from continuation is an increasing affine function of μ, and is *independent* of σ. Since project continuation fully reveals the project's quality, it is analogous to *costly disclosure with a verifiable type* (Jovanovic (1982), Verrecchia (1983)) with the difference that the cost depends on the belief.

However, since no information is revealed when the leader terminates the project, her reputation is independent of her private belief, as is the social payoff (which is zero). Therefore, any equilibrium is characterized by a threshold, x. The leader cancels the project if $\mu < x$, and continues if $\mu > x$. The leader's payoff from cancelling equals:

$$U(N, x) := \mathbb{E}(\mu | \mu < x). \tag{2}$$

An equilibrium is a pair (μ^*, ρ^*), where

$$\rho^* = \mathbb{E}(\mu | \mu < \mu^*), \tag{3}$$

$$U(Y, \mu^*) = \rho^*. \tag{4}$$

From (3), the observer's optimal action is increasing in the leader's threshold μ^*, while (4) implies that the leader's optimal stopping threshold is increasing in the observer's action ρ^*. Thus, our game is supermodular, and an equilibrium exists (Milgrom and Roberts (1990)).

Multiple equilibria arise naturally due to supermodularity. Figure 1 illustrates: the payoffs $U(Y, x)$ and $U(N, x)$ intersect in three points, at $\underline{\mu}^*, \mu_2^*$ and $\bar{\mu}^*$, corresponding to three distinct equilibrium thresholds. We will use $\underline{\mu}^*$ to denote the smallest equilibrium threshold, and $\bar{\mu}^*$ to denote the largest.

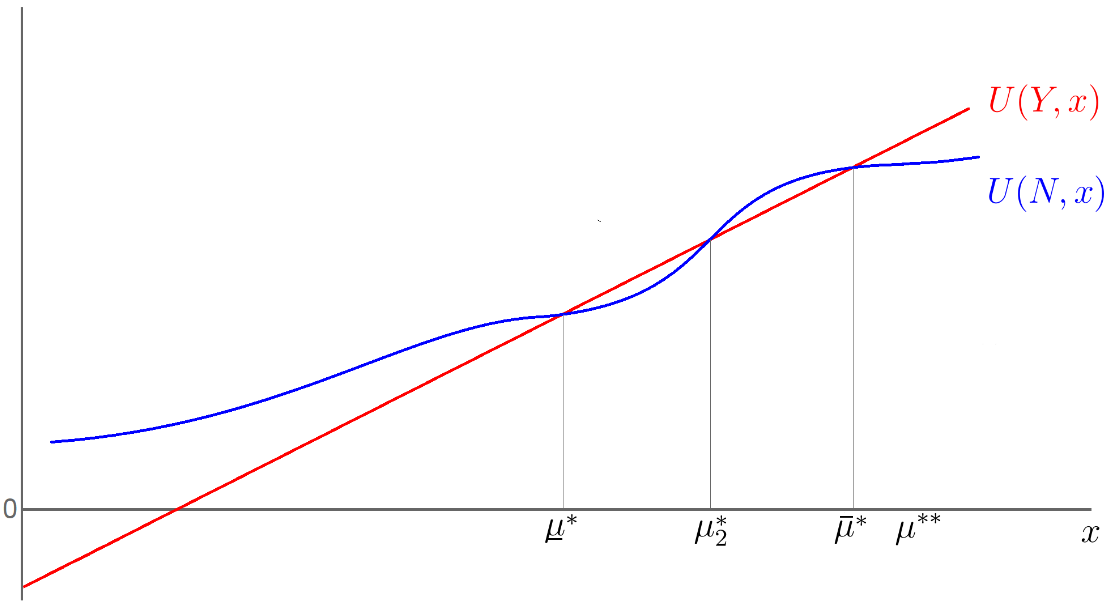

Figure 1: The leader's payoff $U(Y,x)$ from continuing when her private belief μ equals the threshold x, and her payoff $U(N,x)$ from stopping under the strategy with threshold belief x.

To examine efficiency and multiplicity, let $\phi(x) := x - \mathbb{E}(\mu|\mu < x)$ measure the reputational cost of stopping when the leader's belief equals the threshold x. Let $\mathcal{C}(F)$ denote the support of F, and let $\underline{\mu}$ be its smallest element. The reputational cost $\phi(x)$ is strictly positive for every $x > \underline{\mu}$. At any equilibrium threshold μ^*,

$$\theta\,(c - \mu^* v) = \phi(\mu^*) > 0,$$

which implies that $\mu^* < \mu^{**}$ (the efficient threshold) for any θ.

If ϕ is monotone increasing, equilibrium is unique. If $\phi(\mu)$ is not monotone, nor is $\phi(\mu) + \theta v \mu$, for θv small enough, and there exists a value of θc which intersects it more than once. Thus, there are parameter values, θ, c, v, for which multiple equilibria exist. Non-monotonicity of ϕ is equivalent to failure of log-concavity of the left-hand integral, $G(x) := \int_{\underline{\mu}}^{x} F(\mu) d\mu$.[5] This occurs when there are "large" news events, which give rise to mass points or sharp changes in the probability density function.

The next definition will be useful when we analyze overconfidence.

[5]Bagnoli and Bergstrom (2005), Lemma 1.

Definition 2 *An equilibrium with threshold μ^* is stable (unstable) if there exists an open interval containing μ^* such that $[U(N,x) - U(Y,x)][\mu^* - x] > 0$ (< 0) for all $x \neq \mu^*$ in this interval.*

Graphically, stability of a threshold μ^* corresponds to the line depicting the payoff from continuing cutting the graph of the payoff from stopping from below. Non-generically, the two graphs might be tangent at an equilibrium threshold, or coincide over an interval. To simplify exposition, we assume:

Assumption 3 *All equilibria are either stable or unstable.*

In Figure 1, the largest and smallest equilibria are stable; the middle equilibrium is unstable. If an equilibrium is stable, comparative statics are intuitive — for example, a small decrease in θ increases the equilibrium threshold.

Proposition 1 *1. When the distribution of beliefs has no gap immediately below μ^{**}, there is inefficient continuation in every equilibrium, no matter how small reputational concerns are.*

2. There is inefficient continuation in every equilibrium if reputational concerns are sufficiently large, i.e. if θ is small enough.

3. Equilibria are welfare-ranked in the order of their thresholds. The most inefficient equilibrium, with the smallest threshold, $\underline{\mu}^*$, is always stable.

Proof. 1. Let $\hat{\mu} := \sup\{\mathcal{C}(F) \cap [0, \mu^{**})\}$. If $\hat{\mu} = \mu^{**}$, then $[\mu^*, \mu^{**}]$ has positive F-measure, and there is inefficient continuation.

2. If $\hat{\mu} < \mu^{**}$, recall that $\phi(x) > 0$ for every $x > \underline{\mu}$, and choose $\hat{\theta}$ sufficiently small that $\hat{\theta}(c - \hat{\mu}v) < \min_{\mu \in [\hat{\mu}, \mu^*]} \phi(\hat{\mu})$. Then, no $\mu \geq \hat{\mu}$ can be an equilibrium threshold. Thus, if $\theta \leq \hat{\theta}$, any equilibrium threshold has $\mu^* < \hat{\mu}$, so $[\mu^*, \mu^{**}]$ has positive F-measure.

3. Social welfare is increasing in the threshold μ^*, since $\mu^* < \mu^{**}$, establishing the first claim. At $\underline{\mu}$, $\phi(\underline{\mu}) = 0$ and the social payoff is negative. Therefore, $U(Y, \underline{\mu}) < U(N, \underline{\mu})$. Consequently, $U(Y, \mu) < U(N, \mu)$ for all $\mu < \underline{\mu}^*$: if this inequality was violated for some $\mu < \underline{\mu}^*$, since both functions are continuous, there must be another equilibrium to the left of $\underline{\mu}^*$, a contradiction. Since the stability condition holds on an interval to the left of $\underline{\mu}^*$, Assumption 3 implies that it also holds on an interval to the right of $\underline{\mu}^*$. ■

In any equilibrium, the observer's posterior belief is a martingale, and its expectation at date zero equals the prior p. Therefore, the leader's expected reputational payoff does not vary across equilibria, and equilibria with higher social payoffs are also better for her.

Explicitly modelling the project initiation decision shows that it is conditionally efficient: any initiated project is ex ante profitable. Consequently, terminating it must be, on average, bad news. However, *public* bad news can depress the average public belief below μ^{**}, and in this case, project terminations will increase the stock price, as found by Statman and Sepe (1989). Bhaskar and Thomas (2018) show that a modification of our model explains why new managers scrap the projects of their predecessors. If a manager's contribution to firm value is persistent, a new manager will seek to minimize the reputation of her predecessor, thereby maximizing the observer's evaluation of her own contribution to firm value. She will terminate her predecessor's projects inefficiently often, and these terminations must reduce the share price. Our analysis suggests that empirical work on the announcement effects of project terminations ought to condition on recent changes in management, and on the nature of preceding public news.

2 Overconfident Leader

Suppose that leader and observer have different priors on the competence of the leader, and therefore, on the quality of the project. The leader's prior on project quality, q, is strictly greater than p, the observer's prior. Assume that this is common knowledge. This defines Γ^0, a game with non-common priors. Define γ, a measure of the overconfidence of the leader relative to the observer:

$$\gamma := \frac{q}{1-q}\frac{1-p}{p} - 1 \ > 0.$$

If the leader observes a signal s with likelihood ratio ℓ,[6] her posterior belief equals

$$\pi(\ell) := \frac{q\,\ell}{q\,\ell + 1 - q}. \tag{5}$$

The posterior belief of the observer, were he to observe the same signal, would be

$$\mu(\ell) := \frac{p\,\ell}{p\,\ell + 1 - p}. \tag{6}$$

We henceforth identify a signal with its associated likelihood ratio and define the above belief-updating functions for all $\ell \in [0, \infty)$, with $\pi(\infty) = \mu(\infty) = 1$. Define $\pi^{\dagger}(\mu)$ to be the posterior belief of the leader after a signal that would induce in the observer the posterior

[6] $\ell = \frac{h(s|G)}{h(s|B)}$, where $h(s|\omega)$ denotes the probability density at s when project quality is ω.

belief μ. That is, $\pi^\dagger : [0,1] \to [0,1]$ is obtained by using the inverse of (6) as the argument in (5):

$$\pi^\dagger(\mu) = \frac{(\gamma + 1)\,\mu}{\gamma\,\mu + 1}. \tag{7}$$

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

Proof. See appendix. ■

Arguing as before, in any equilibrium, the leader must follow a threshold strategy. It is analytically convenient to define the leader's threshold, x, in terms of the observer's belief: the leader continues the project if and only if $\pi^\dagger(\mu) > \pi^\dagger(x)$, or, equivalently, $\mu > x$.

If the leader terminates the project, the observer updates his belief regarding the project quality to $\mathbb{E}_F(\mu|\mu < x)$, where the expectation is taken with respect to F, the distribution of beliefs under the prior p.[7] Consequently, the leader's payoff from termination under the strategy x is given by

$$U(N, x) := \mathbb{E}_F(\mu|\mu < x). \tag{8}$$

The leader's private belief at the threshold is $\pi^\dagger(x)$, and her payoff from continuing the project is

$$U(Y, \pi^\dagger(x)) = \pi^\dagger(x)\,[\theta v + 1] - \theta c. \tag{9}$$

The threshold μ_0 is an equilibrium strategy for the leader if

$$U(Y, \pi^\dagger(\mu_0)) = U(N, \mu_0) = \rho_0. \tag{10}$$

where ρ_0 denotes the observer's equilibrium action, and the second equality follows from (8).

The game is supermodular: the leader's optimal threshold is increasing in the observer's action ρ, and ρ is increasing in the leader's threshold. Thus equilibrium exists. Let $\underline{\mu}_0$ ($\bar{\mu}_0$) denote the smallest (largest) equilibrium threshold, and $\underline{\rho}_0$ ($\bar{\rho}_0$) the corresponding observer action. From Milgrom and Roberts (1990), the sets of rationalizable strategies are $[\underline{\mu}_0, \bar{\mu}_0]$ and $[\underline{\rho}_0, \bar{\rho}_0]$.

Figure 2 illustrates the equilibrium of the game Γ^0. The straight line, $U(Y, x)$, shows the payoff from continuation without overconfidence, i.e. when the posterior belief under

[7]Let $H(s|\omega)$ denote the cumulative distribution function of the leader's private signals in the state ω. From the point of view of the observer, whose prior belief that $\omega = G$ is p, $H_p := p\,H(s|G) + (1-p)\,H(s|B)$ is the distribution over the signals that the leader observes. We let F denote the associated distribution over the resulting posterior beliefs, $\mu(\ell(s))$.

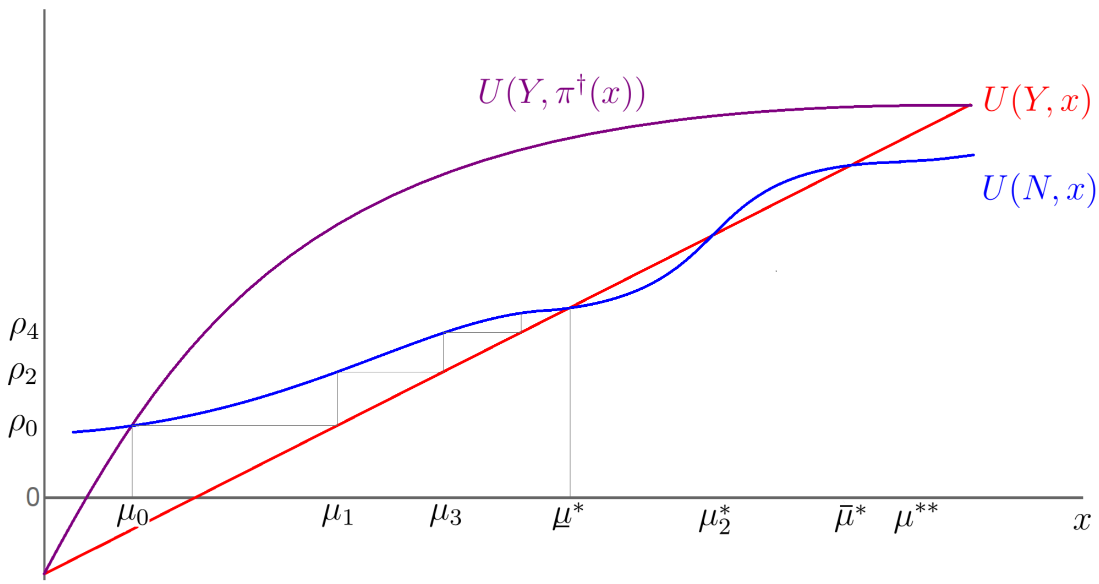

Figure 2: The threshold μ_0 and the observer action ρ_0 constitute an equilibrium of the game Γ^0. The threshold $\underline{\mu}^*$ is the smallest equilibrium threshold in the game with common knowledge of a common prior.

the prior p equals the threshold x. The concave function $U(Y, \pi^\dagger(x))$ shows the leader's payoff from continuation with overconfidence, i.e. when the leader's posterior belief is $\pi^\dagger(x)$. By Lemma 1, this lies above $U(Y, x)$. The equilibrium threshold with overconfidence, μ_0, satisfies (10). In Figure 2, overconfidence is large enough that μ_0 is strictly below $\underline{\mu}^*$, the smallest equilibrium threshold under a common prior. In general, overconfidence leads to excessive continuation, over and above that arising under a common prior, for two reasons. The first is straightforward: since the leader's belief is $\pi^\dagger(\mu_0) > \mu_0$, she continues because she has more optimistic beliefs in Γ^0 than in the game with a common prior. The second reason is more subtle. Since in Γ^0 the observer knows that the leader is overconfident (as compared to his own prior), his inference upon project termination is more adverse. Thus the leader knows that she will be penalized more for terminating the project in Γ^0 than in the game with a common prior.

Proposition 2 *The smallest (largest) equilibrium under overconfidence (game Γ^0) is smaller than the smallest (largest) equilibrium under common priors: $\underline{\mu}_0 < \underline{\mu}^*$ and $\bar{\mu}_0 < \bar{\mu}^*$. If overconfidence, γ, is sufficiently large, then every equilibrium of Γ^0 is smaller than the smallest equilibrium under common priors: $\bar{\mu}_0 < \underline{\mu}^*$.*

Proof. Since $\pi^\dagger(\mu) > \mu$ for $\mu \in (0, 1)$, and $U(Y, \mu) \geq U(N, \mu)$ for $\mu \geq \bar{\mu}^*$, $U(Y, \pi^\dagger(\mu)) >$

$U(N,\mu)$ for $\mu \geq \bar{\mu}^*$, implying $\bar{\mu}_0 < \bar{\mu}^*$. Similarly, $U(Y, \pi^\dagger(\underline{\mu}^*)) > U(N, \underline{\mu}^*)$. If $U(Y, \pi^\dagger(\underline{\mu})) < U(N, \underline{\mu})$, then, since both functions are continuous, Γ^0 must have an equilibrium with $\mu_0 \in (\underline{\mu}, \underline{\mu}^*)$; otherwise, there exists an equilibrium where the leader always continues, i.e. $\underline{\mu}_0 \leq \underline{\mu}$.

From (7), for any interior value of μ, $\lim_{\gamma\to\infty} \pi^\dagger(\mu) = 1$. Consequently, for γ sufficiently large, every equilibrium threshold in Γ^0 is strictly below $\underline{\mu}^*$, implying $\bar{\mu}_0 < \underline{\mu}^*$. ■

The effect of a small increase in γ will be to increase $U(Y, \pi^\dagger(x))$ for every x. This lowers the equilibrium threshold if the equilibrium is stable (i.e. $[U(N,x) - U(Y, \pi^\dagger(x))][\mu_0 - x] > 0$ for all $x \neq \mu_0$ in an open interval containing μ_0), yielding the intuitive comparative statics that overconfidence makes project continuation more appealing.

2.1 Perceived Overconfidence

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

- (**S1**) The leader and the observer share the prior p.
- (**T1**) The observer believes that the leader's prior is $q > p$.
- The observer's second-order belief (T1) is common knowledge.

Since Γ^0 denotes the game with actual overconfidence, an alternative formalization of Γ^1 is:

- The leader and the observer share the prior p.
- The observer believes that the game Γ^0 is being played, and this belief is common knowledge.

In an equilibrium of the game Γ^1, the observer's strategy must be the same as in some equilibrium of the game Γ^0, since he believes that Γ^0 is being played. The leader's strategy differs across the two games. Thus, an equilibrium of Γ^1 is a triple $(\mu_0, \rho_0, \sigma_1)$ where (μ_0, ρ_0) is an equilibrium of the game Γ^0, and σ_1 is a best response to ρ_0. Arguing as before, σ_1 must be a threshold strategy. Let μ_1 denote the threshold. It satisfies

$$U(Y, \mu_1) = \rho_0. \tag{11}$$

Since the left-hand side is strictly increasing in μ_1, (11) has a unique solution. Equivalently, for any equilibrium (μ_0, ρ_0) of the game Γ^0 with overconfidence, there is a unique equilibrium (μ_0, ρ_0, μ_1) in the game Γ^1 with perceived overconfidence. This is illustrated in Figure 2. Here, the unique equilibrium of Γ^0 has threshold μ_0 satisfying (10). In Γ^1, the unique corresponding threshold, μ_1, satisfies (11), where ρ_0 is the observer's action in both equilibria. The equilibrium *outcome* in the game Γ^1 is (μ_1, ρ_0).

Conditions (10) and (11) imply $\mu_1 = \pi^\dagger(\mu_0)$, so the leader's equilibrium cut-off beliefs are identical in the two games. Lemma 1 implies $\mu_1 > \mu_0$, so a leader who really is overconfident requires more adverse news for termination than a leader who is merely perceived to be overconfident. Nonetheless $\mu_1 < \mu^*$ because, when compared to the game with a common prior, the perception of overconfidence penalizes the reputation of a leader who is not, in fact, overconfident.

2.2 Higher-Order Beliefs about Overconfidence

Suppose that the leader is not overconfident, the observer knows this, the leader knows that the observer knows that the leader is not overconfident, and so on, up to a finite order. However, it is not common knowledge that the leader is not overconfident. To analyze this situation, we define the game Γ^N. We defined the statements (S1) and (T1) when describing the game Γ^1. Now, define the following statements for every integer $K > 1$:

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

For every integer $N > 1$, Γ^N is defined by:

- The statements S1 to SN are true.
- The statement TN is true.
- TN is common knowledge.

Consider the sequence of games, (Γ^N), $N \in \{0\} \cup \mathbb{N}$. Let $\mathcal{E}$ denote the set of even numbers, and $\mathcal{O}$ denote the set of odd numbers. 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 Γ^N consists of the sequence truncated at N, with the property that:

- (μ_0, ρ_0) is an equilibrium of the game Γ^0.

- For any $n \in \mathcal{O}, n \leq N$, μ_n is a best response to ρ_{n-1}:

$$U(Y, \mu_n) = \rho_{n-1}. \tag{12}$$

- For any $n \in \mathcal{E}, n \leq N$, ρ_n is a best response to μ_{n-1}:

$$\rho_n = U(N, \mu_{n-1}). \tag{13}$$

These equilibria are illustrated in Figure 2. The pair (μ_0, ρ_0) represents the equilibrium in Γ^0, the game with overconfidence. In Γ^1, the game with perceived overconfidence, the observer believes that Γ^0 is being played, and therefore plays ρ_0. The leader's equilibrium strategy in Γ^1 is μ_1, her best response to ρ_0, satisfying (12). In Γ^2, the game with perceived perceived overconfidence, the leader believes that Γ^1 is being played, and therefore plays μ_1. The observer's equilibrium strategy is ρ_2, his best response to μ_1, satisfying (13). In game Γ^n the equilibrium outcome is (μ_n, ρ_{n-1}) if n is odd, and (μ_{n-1}, ρ_n) if n is even.

This defines a tâtonnement process that converges to $(\mu^*_+(\mu_0), \rho^*_+(\mu_0))$, an equilibrium the game with common knowledge of common priors, henceforth denoted Γ. More precisely, $\mu^*_+(\mu_0)$ denotes the smallest equilibrium threshold in Γ that is larger than μ_0:

$$\mu^*_+(\mu_0) = \min\{x > \mu_0 : U(Y, x) = U(N, x)\}.$$

Let $\rho^*_+(\mu_0)$ denote the observer's associated equilibrium strategy.

Proposition 3 *Fix an equilibrium (μ_0, ρ_0) of Γ^0, the game with overconfidence. This induces sequences $(\mu_n)_{n\in\mathcal{O}}$ and $(\rho_n)_{n\in\mathcal{E}}$, such that for any $n \in \mathcal{O}$, the equilibrium outcome is (μ_n, ρ_{n-1}) in game Γ^n, and (μ_n, ρ_{n+1}) in game Γ^{n+1}. The sequences (μ_n) and (ρ_n) are both increasing, and converge to $\mu^*_+(\mu_0)$ and $\rho^*_+(\mu_0)$ respectively, a stable equilibrium of Γ, the game with common priors.*

Proof. Fix an equilibrium (μ_0, ρ_0) in Γ^0, and let $\mu_{-1} := \mu_0$. Equations (12) and (13) define a unique pair of sequences $(\mu_n)_{n\in\mathcal{O}\cup\{-1\}}, (\rho_n)_{n\in\mathcal{E}\cup\{0\}}$ that satisfy:

$$\rho_{n+3} = U(N, \mu_{n+2}) \geq U(Y, \mu_{n+2}) = U(N, \mu_n) = \rho_{n+1} \geq U(Y, \mu_n), \quad n \in \mathcal{O} \cup \{-1\}. \tag{14}$$

The equalities follow from (12) and (13). The inequalities hold if the sequence $(\mu_n)_{n\in\mathcal{O}}$ is increasing, as we now prove. Section 2.1 shows that $\mu_1 > \mu_0$. Assume, by the induction hypothesis, that $\mu_n \geq \mu_{n-2}$. Then $U(N, \mu_n) \geq U(N, \mu_{n-2})$, since $U(N, .)$ is increasing. Let $U_Y^{-1}(.)$ denote the inverse of $U(Y, .)$. It is strictly increasing. Then, $\mu_{n+2} = U_Y^{-1}[U(N, \mu_n)]$ implies $\mu_{n+2} \geq \mu_n$. Thus, (μ_n) is increasing. Since it is bounded, it converges to some value, μ^∞. Observe that, by (13), (ρ_n) is increasing and converges to $\rho^\infty := U(N, \mu^\infty)$.

Since $U(N, \mu)$ and $U(Y, \mu)$ are continuous functions, $U(Y, \mu_n) \leq U(N, \mu_n)$ for all n implies $U(Y, \mu^\infty) \leq U(N, \mu^\infty)$. We now prove the reverse inequality. Since (μ_n) is Cauchy, and $U(Y, \mu)$ is an affine function, the sequence $(U(Y, \mu_n))$ is Cauchy. Thus, for any $\epsilon > 0$, there exists $\bar{n}$ such that $U(Y, \mu_{n+2}) - U(Y, \mu_n) < \epsilon$ if $n > \bar{n}$. By 14, $U(Y, \mu_{n+2}) = U(N, \mu_n)$. Therefore, $U(N, \mu_n) - U(Y, \mu_n) < \epsilon$ if $n > \bar{n}$. Since ϵ was arbitrary, $U(N, \mu^\infty) \leq U(Y, \mu^\infty)$. We conclude that $U(Y, \mu^\infty) = U(N, \mu^\infty)$, and μ^∞ is an equilibrium threshold of Γ. Assumption 3 and $U(Y, \mu_n) \leq U(N, \mu_n)$ for all n imply that $(\mu^\infty, \rho^\infty)$ is stable.

Next we show that $\mu^\infty = \mu_+^*(\mu_0)$, by showing, by induction, that $\mu_n \leq \mu_+^*(\mu_0)$ for all $n \in \mathcal{O}$. By (10), $\mu_{-1} := \mu_0 < \mu_+^*(\mu_0)$. Suppose $\mu_n \leq \mu_+^*(\mu_0)$. Then, $U(N, .)$ increasing implies $U(N, \mu_n) \leq U(N, \mu_+^*(\mu_0))$, and $U_Y^{-1}(.)$ increasing implies $\mu_{n+2} = U_Y^{-1}[U(N, \mu_n)] \leq U_Y^{-1}\left[U(N, \mu_+^*(\mu_0))\right] = \mu_+^*(\mu_0)$. ■

If we consider higher-order beliefs regarding underconfidence, i.e. the leader having a prior q that is strictly less than the prior p of the observer, it remains true that no unstable equilibrium can be approximated. Without common knowledge of common priors, regardless of how priors diverge, only stable equilibria can be selected.

2.2.1 Large Overconfidence

Hubris amongst leaders would not be a source of concern if it were not significant, and γ is therefore likely to be large. We focus on limit outcomes under large (potential) overconfidence, when it is "almost common knowledge" that the leader is not overconfident. Only the most inefficient equilibrium of Γ is selected. Furthermore, it is uniquely (limit) rationalizable. Thus, it is not essential to assume that players play a Nash equilibrium in any of the games that lack common knowledge of common priors. This addresses the concern, raised by Dekel, Fudenberg, and Levine (2004), that there is a conceptual inconsistency in imposing equilibrium requirements in a setting with heterogeneous priors.

Proposition 4 *If γ is sufficiently large, then any sequence of rationalizable leader 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 $\underline{\rho}^*$.*

Proof. By Proposition 2, for γ large enough, $\bar{\mu}_0 < \underline{\mu}^*$. Thus, $\bar{\rho}_0 \leq \underline{\rho}^*$, since $\bar{\rho}_0 = U(N, \bar{\mu}_0)$, $\underline{\rho}^* = U(N, \underline{\mu}^*)$ and $U(N, .)$ is increasing. A rationalizable strategy of the leader in Γ^1 must be a best response to a probability distribution over the rationalizable strategies of the observer in Γ^0, i.e. to a distribution over $[\underline{\rho}_0, \bar{\rho}_0]$. Since the leader's payoff from stopping is linear in ρ, her rationalizable thresholds μ_1 in Γ^1 satisfy

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

For $n \in \mathcal{O}$, let $\underline{\mu}_n$ denote the smallest equilibrium threshold for the leader in Γ^n, and $\bar{\mu}_n$ the largest. Any rationalizable threshold, μ_n, must belong to $[\underline{\mu}_n, \bar{\mu}_n]$. Similarly, for $n \in \mathcal{E}$, let $\underline{\rho}_n$ ($\bar{\rho}_n$) denote the smallest (largest) equilibrium action for the observer in Γ^n. Again, any rationalizable action ρ_n belongs to $[\underline{\rho}_n, \bar{\rho}_n]$. Consider the sequences (μ_n) and (ρ_n) induced by $(\underline{\mu}_0, \underline{\rho}_0)$. By Proposition 3, these 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. Finally, any rationalizable sequence (μ_n) (resp. (ρ_n)) is sandwiched between the two induced equilibrium sequences. ■

This result shows that a culture where corporate and political leaders are expected to be substantially overconfident may have pernicious effects even for leaders who are not overconfident. It may be mutual knowledge to a high degree that the leader has realistic beliefs. However, the lack of common knowledge entailed by the culture ensures that the most inefficient equilibrium is selected. Our theory highlights a novel and important channel though 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. Thus, a female CEO may benefit from stereotyping and may feel less pressure to pursue unprofitable projects than her male counterparts, since she will be penalized less for cancellations.

The implications of large underconfidence (i.e. $\gamma < 0$) are in the opposite direction: the most efficient equilibrium in Γ is uniquely selected. Our results echo Weinstein and Yildiz (2007), who show that any rationalizable action in a complete information game can be made uniquely rationalizable by perturbing higher-order beliefs. But when it comes to *unique* selection, the class of perturbations we consider can only select extremal equilibria. It remains an open question whether and how one might uniquely select intermediate equilibria, especially unstable ones.

Weinstein and Yildiz's permissive results militate against a universal theory of equilibrium selection. One response is that the equilibrium which will be played depends on circumstances. Our focus on overconfidence (rather than underconfidence) to model divergences in higher-order beliefs is motivated by the context. Literature, politics and the world of business are replete with examples of leaders afflicted with hubris. Our result shows that this may influence outcomes even for the most clear-headed of leaders.

A Appendix: Proofs

A.1 When the leader is concerned about beliefs about her ability

In Holmström (1999) and much subsequent work on career concerns, the leader's reputational payoff is a linear function of her expected ability. We now show that this assumption yields a model formally equivalent to the one analyzed in the text of the paper, where the leader's payoff is linear in the observer's belief about project quality. The leader's ability is $\tau \in \{H, L\}$, and project quality is $\omega \in \{G, B\}$. Let λ denote the prior probability that the leader is of type H. Let $p_\tau := \Pr(\omega = G|\tau)$ denote the probability with which type τ has a good project, with $p_H > p_L$. The common prior on project quality is $p := \lambda\, p_H + (1-\lambda)\, p_L$. Let β be the observer's posterior belief, at date two, regarding project quality. Let $\nu(\beta)$ denote the observer's posterior belief at date two that the leader is of type H, given β. The relation between ν and β is as follows. When the project succeeds, $\beta = 1$ and

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

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

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

Both ν and β must satisfy the martingale property for a (hypothetical) experiment which perfectly reveals the project's quality, hence

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

Let $\delta := \nu(1) - \nu(0) < 1$. Then,

$$\nu(\beta) = \nu(0) + \delta\beta.$$

Suppose that the observer takes an action in $[0, 1]$ to match ν, and the leader's total payoff equals $\tilde{\theta}V + \nu$, where V denotes the social payoff from the project, and $\tilde{\theta} > 0$ is a constant parameter reflecting the intensity of the leader's social concerns. Thus the leader's payoff equals

$$\tilde{\theta}V + \nu(0) + \delta\beta.$$

If we let $\theta := \tilde{\theta}/\delta$, then the above payoff is identical to the one analyzed in the text, except for a constant term, $\nu(0)/\delta$, which accrues to both actions, stop and continue, and therefore does not affect the analysis.

A.2 Proof of Lemma 1

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

$$\frac{\gamma+1}{(\gamma\mu+1)^2} > 0. \tag{A.15}$$

The numerator in the above expression does not depend on μ, and the denominator is increasing in μ when $\gamma > 0$. Thus the derivative of $\pi^\dagger$ is strictly decreasing in μ. Since $\pi^\dagger$ is strictly concave in μ, with $\pi^\dagger(0) = 0$ and $\pi^\dagger(1) = 1$, it follows that $\pi^\dagger(\mu) > \mu$ for every $\mu \in (0,1)$.

In the case of underconfidence, i.e. when $\gamma < 0$, $\pi^\dagger$ continues to be increasing, since $\gamma > -1$, but the denominator in A.15 is strictly decreasing, so that $\pi^\dagger$ is strictly convex, and $\pi^\dagger(\mu) < \mu$ for every $\mu \in (0,1)$.

References

Bagnoli, M., and T. Bergstrom (2005): “Log-Concave Probability and Its Applications,” *Economic Theory*, 26(2), 445–469.

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.

Bhaskar, V., and C. Thomas (2018): “The Culture of Overconfidence,” Mimeo, https://sites.google.com/site/carotho14/research/stable/overconfidence.pdf.

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.

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.

Dur, R. A. (2001): "Why do policy makers stick to inefficient decisions?," *Public Choice*, 107(3), 221–234.

Gneezy, U., M. Niederle, and A. Rustichini (2003): "Performance in competitive environments: Gender differences," *The Quarterly Journal of Economics*, 118(3), 1049–1074.

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.

Jovanovic, B. (1982): "Truthful disclosure of information," *The Bell Journal of Economics*, pp. 36–44.

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 (2008): "Who makes acquisitions? CEO overconfidence and the market's reaction," *Journal of Financial Economics*, 89(1), 20–43.

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 (2003): "Global Games: Theory and Applications," in *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky. Cambridge University Press.

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.

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.

Sethi, R., and M. Yildiz (2016): "Communication with unknown perspectives," *Econometrica*, 84(6), 2029–2069.

Statman, M., and J. F. Sepe (1989): "Project termination announcements and the market value of the firm," *Financial Management*, pp. 74–81.

Verrecchia, R. E. (1983): "Discretionary disclosure," *Journal of accounting and economics*, 5, 179–194.

Weinstein, J., and M. Yildiz (2007): "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica*, 75(2), 365–400.