Perceived Overconfidence, Private Information and Career Concerns

V Bhaskar and Caroline Thomas

December 8, 2017

Abstract

Why do political leaders or managers persist with their pet projects/policies even after privately observing adverse information? Since project continuation is a more informative experiment than project termination, a reputationally concerned leader is biased towards continuation for this reason. Overconfidence on the part of the leader aggravates this tendency. We show that perceived overconfidence has similar adverse effects on efficiency, and that higher order beliefs regarding overconfidence can induce inefficency even when the leader is not in fact overconfident.

.

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

In 1327, Mohammad bin Tughlaq, having recently murdered his father and become Sultan over the vast expanse that included much of modern day India and Pakistan, resolved to relocate the capital of his kingdom from Delhi to the city of Daulatabad, over 1200 km to the south. Tughlaq had good reasons for his move – he wanted to conquer territory in the south, and also wanted his capital secure from invasions from the north. Nonetheless, an upheaval of this magnitude, in an era when his populance was restricted to bullock carts and walking, caused enormous difficulties and resistance. Daulatabad proved to be plagued by water shortages, and revolts erupted in parts of his far-flung empire that were harder to control from his southern capital. Finally, eight years later, the Sultan decided to move his capital back to Delhi.

Why do managers and political leaders persist with pet projects, when faced with evidence that should call for their re-evaluation? Examples are abundant. Margaret Thatcher's riposte to critics of her monetarist policy,[1] which had raised unemployment in Britain to over 2 million, is a part of her legend. Later, Thatcher persisted with the unpopular poll tax when confronted by widespread opposition, a stubbornness that ultimately led to her being deposed as leader. Mao Tse-Tung persisted and intensified the Great Leap Forward, despite many reports of widespread starvation in the countryside, with over 30 million people dying in the consequent famine (see Dikötter (2010)). A popular explanation, going back at least to the ancient Greeks, is hubris. In Aeschylus's *The Persians*, King Xerxes is consumed by ambition. He seeks to surpass his father Darius and become the greatest of kings. Having conquered Egypt, he turns his ambition to Greece, not heeding the phantom of his father, the warnings of his wife, or the chorus of elders who foresee a calamity. He does not offer a sacrifice to the gods before the final battle. Blinded by hubris, he makes strategic errors and loses the naval battle of Salamis against the Greeks, precipitating his downfall. In Sophocles's *Antigone*, Creon, maddened by absolute power, decrees that no-one

[1]Quoted at the head of this paper.

may bury Polynices, ignoring the advice of others and the many omens that portend disaster. The modern literature on management has taken up this argument (see, for example, Roll (1986), or Malmendier and Tate (2005)). Leaders, be they CEOs, elected politicians or dictators who have seized power, are convinced that they are exceptional talents. The skills that enable them to gain positions of power may be unrelated to those required for exercising it. No matter; they are convinced of their singularity. Adverse information is filtered though the lens of prior conviction.

As economists, we seek, at least initially, more rational explanations for a leader's intransigence. In keeping with this requirement, we shall insist that not even the best leaders are infallible. Since even the wisdom of a Keynes does not preclude making a choice that turns out badly, changing your mind is not a sign of incompetence. We consider a decision maker – the manager of a firm, or a political leader – who receives information regarding the viability of a project, and who must decide, whether to continue with the project, or to abandon it. The information received is private, and so unobserved by outsiders who evaluate the decision maker's performance. Our explanation hinges on the reputational concern of a leader and highlights a critical informational difference between continuing with a project, and terminating it. By continuing a project to its conclusion, a leader verifies publicly her private belief regarding its quality. Thus, if she believes that a project has a quarter chance of success, she believes that the project will be revealed to have been a good one with probability 0.25 and a failure with probability 0.75. By terminating the project, she prevents any further learning, and only reveals that she thought the project bad enough not to continue. Thus, a leader who learns privately regarding the quality of the project faces a reputational cost when she is moderately pessimistic about the project. At a belief where she is indifferent between continuing the project and terminating it, she will usually incur a reputational cost from termination, and will therefore continue projects that are socially wasteful or destroy firm value. We also find that often, there will be multiple equilibria, with equilibria being ranked in efficiency in decreasing order of the continuation belief threshold. 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 decision maker's overconfidence, in particular the interaction between private information and DM overconfidence. We assume that the DM and outside observer have different prior beliefs on the project's chances of success; the DM has a prior q that is strictly larger than p, the prior of the observer.Since continuing the project results in more information revelation, there is an additional value to continuation, from the DM's point of view. An overconfident DM continues the project after some signals where he would not, were he to have the same prior belief as the outside observer. However, 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 the DM's private information when the DM does in fact terminate 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 *perceived* as being so, since the negative inference from project termination is identical to that when he is in fact overconfident. In other words, perceived overconfidence, in conjunction with the DM's private information, induces a greater tendency to stubbornness.

One may take this argument further. Suppose that the DM is not overconfident, and is also not perceived to be overconfident. In other words, one has neither overconfidence, nor perceived overconfidence, but perceived perceived overconfidence. Similarly, one may have mutual knowledge that the DM is not overconfident to a very high level, but not common knowledge, so that at some level, there is a perception that the DM is overconfident. Suppose that there are multiple equilibria in the game with a common prior, and suppose also that parameters are generic. We show that as one has higher and higher levels of mutual knowledge, the limit equilibrium set converges to a strict subset of the equilibria of the game with a common prior. If the difference in priors $q - p$ is large enough, then there is unique limit equilibrium, which is the most inefficient one.

1.1 Related literature

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

Hubris, or overconfidence is a leading explanation of why political leaders fail to reverse bad decisions despite bad news. Hubris has been used to explain Napoleon's doomed march on Moscow (Kroll, Toombs, and Wright (2000)), Hitler's over-ambitious expansion (Kershaw (2000)), Margaret Thatcher's failure to reverse the poll tax despite evidence of its unpopularity (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 successful politicians. Dictators brook no contrary opinion and end up believing their own propaganda, while success convinces democratically elected leaders that they are unusually talented and one of a kind. They provide a detailed discussion of several US presidents and British prime ministers to support their claim that hubris syndrome is pervasive at the highest levels of leadership.

Hubris has similarly been used to explain the behavior of CEO's in firms and their tendency to persist with bad decisions. Since mergers/acquisitions is one area where the effects on shareholder value can be quantified, this is an important area of application. Roll (1986) argues that overconfidence in the CEO's ability is a major factor behind corporate takeovers, and explains the adverse effects of takeovers on the acquiring firm's value. Indeed, the well documented phenomenon that takeovers often lose value for shareholders of acquiring firms has been explained by managerial overconfidence by Morck, Shleifer, and Vishny (1990) and Malmendier and Tate (2008). Malmendier and Tate (2005) argue that CEO overconfidence leads to over-investment in their pet projects, and also leads them to view external funds as unduly costly, leading to a strong link between cash flows and corporate investment. Although systematic empirical evidence on managers' reluctance to abandon pet projects is harder to get, it is straightforward that an excessively optimistic prior belief will lead a manager to discount adverse information and continue a failing project.

A leading rational choice explanation for the persistence of bad policies is the phenomenon of gambling for resurrection. Decision makers who face convex incentives will naturally prefer risky actions. The term appears to have been coined by Downs and Rocke (1994) to explain the incentives of a political leader who is likely to be voted out of office and who may go to war in order to gamble on a major, if unlikely, change in fortune. Similarly, banks and corporations are protected by limited liability, and Freixas, Parigi, and Rochet (2000) argue that "moral hazard and gambling for resurrection are typical behaviors for banks experiencing financial distress." Convex incentives are a straightforward explanation, but also limited in scope – the phrase gambling for resurrection is suggestive of desperate times. A leader who is popular and expects to be re-elected has concave preferences and should minimize risks, and should therefore cancel a dubious project. Managers are also likely to be risk averse, and this may lead them to be cautious.

An alternative explanation, that does not rely on convex preferences, is based on signaling incentives, and has been proposed by Dur (2001) and Majumdar and Mukand (2004). The underlying idea is that good leaders always select good projects, while bad leaders sometimes select bad ones, and may receive subsequent information about the state of the project. Thus cancelling a project is a sign that a leader has changed her mind, and is unambiguously bad news regarding her capability. In Majumdar and Mukand (2004), a leader can be of either high or low ability, and privately knows her ability level. A high ability leader is endowed with a good quality project, while a low ability leader's project may be either good or bad. Once he initiates the project, the leader receives private signals regarding project quality, and must then decide whether to continue with it or abandon it. Since an able leader knows for sure that her project is good, she never revises her beliefs, and never abandons the project. On the other hand, the low ability leader will realize that her project is most likely bad after some signals, but is reluctant to abandon it, since doing so is a sure sign that she is of low ability. In other words, persisting with a chosen project is a characteristic of every able leader, and thus low ability leaders have an incentive to be persistent.

Dur (2001) analyzes a model with a policy maker, who cares both about her reputation for competence (which determines her re-election probability) and social welfare. The policy maker receives a perfectly informative signal, that reveals whether the policy is a success or a failure, and must decide whether or not to continue the policy. The voters do not observe the project's outcome, and so draw inferences only from the policy maker's decision. Dur shows that if the policy maker cares sufficiently about re-election, then she never repeals a policy, since doing so reveals her incompetence, while continuing a failed policy allows her to hide 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 – even one as competent as Keynes will sometimes choose a bad project, and therefore, it is wise to sometimes reverse one's decisions. Second, information is revealed to the public when a project or policy is continued. Thus continuing the policy is an informative experiment, in contrast with aborting it, and the policy maker's incentives to continue are tempered by the fact that she cannot hide the truth. We show that these two factors result in a tendency towards excessive continuation, but this inefficiency is tempered. Even the most reputationally concerned decision maker (DM) will cancel a policy if the news is bad enough. Finally, we examine the interaction between overconfidence and the private information of the DM. Overconfidence aggravates the tendency to persist with bad projects, but more subtly, even the perception of overconfidence has a deleterious impact. Indeed, the culture of hubris may infect decisions even when the DM is not overconfident, and even when there are high degrees of mutual knowledge that the DM is not overconfident.

Our model is also related to the literature on overconfidence, and to models where agents fail to have common priors. Dekel, Fudenberg, and Morris (2007) provide a connection between Bayes Nash equilibria in games without a common prior, and correlated rationalizability. They show that an action is interim correlated rationalizable if and only if it may be played in a Bayes Nash equilibrium on a non-common prior type space Our substantive results are reminiscent of the electronic mail game (Rubinstein (1989)), with the difference that we have to analyze a sequence of games.

2 The Model

Our model captures the interaction between a decision maker (DM henceforth) who undertakes a project, and is concerned with its profitability and with an outside observer's perception of her ability. The DM is either the manager of a firm or a political leader. The observers are, in reality, multiple, but are modelled as a single agent, since they share common beliefs. 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.

The game begins in period zero with nature choosing the ability of the DM, $\tau \in \{H, L\}$, choosing high ability, H, with probability λ, and the DM initiates a project – we do not model this decision explicitly, and simply assume that the project is worthwhile ex ante. The ability of the DM stochastically determines the quality ω of her project, where $\omega \in \{G, B\}$. A DM of type τ is endowed with a good project with probability p_τ, with $1 > p_H > p_L > 0$. The DM does not know her own ability or the type of the project, so both DM and observer share a common prior $p := \lambda p_H + (1 - \lambda) p_L$ that the project is a good one. As we shall see later, the assumption that the DM does not observe her type is not crucial, as long as even able DMs are fallible and sometimes choose bad projects, while incompetent ones may sometimes hit upon a good project.

In period one, the DM privately observes a signal that is informative about the quality of the project, and decides whether to continue to the project or to terminate it. If she continues the project (action Y), this incurs a flow cost c, and the project's outcome is publicly realized in period 2. The outcome is a success if the project is a good one, yielding a return v, and failure if the project is a bad one, yielding zero returns.[2] If she terminates the project (action N) this is publicly observed, and there is no further learning about the project's quality.

At the end of period 2, the observer chooses an action in $[0, 1]$, where his optimal action equals the posterior belief regarding the project's quality. Consequently, we will speak

[2]In reality, failure is often disastrous, but we do not need to assume this, since continuing the project is costly.

interchangeably about the observer's action and his belief. Our reason for introducing the observer's action is simply to make clear that we have a conventional game where DM's payoffs depend upon his actions and the observer's action (rather than a psychological game).

We assume that DM's payoff is a linear function of the net return from the project and the observer's action. The interpretation is that observer and the DM care about the observer's posterior belief regarding the DM's ability. Since the structure of our model implies a one-to-one affine map from beliefs about project quality to beliefs about the DM's ability, we may as well assume that the DM's payoff is a linear function of the observer's action, which equals his posterior belief regarding the quality of the chosen project. Consequently, we assume that the DM maximizes the sum of the social payoff from the project and α times the observer's action – we will often refer to this as the DM's reputation. Since α represents the *relative* weight the DM places on her reputation, this will be smaller:

- The greater the weight of the social payoff, either due to social concerns or explicit performance pay (for a DM).

- The smaller the correlation between DM ability and project quality, i.e. the smaller the difference between p_H and p_L.

That managers of firms have career concerns is well established, and the analysis of its implications has been pioneered by Holmström (1999). The interaction of explicit and implicit incentives is explored by Gibbons and Murphy (1992) and Meyer and Vickers (1997), among others. Similarly, many politicians enter politics motivated by social concerns, and these concerns may remain even if moderated by re-election pre-occupations. The assumption that the DM's payoff is linear in both dimensions, the social payoff and reputation, is a strong one and merits discussion. Our main reason for making this assumption is for analytical clarity. If the DM's evaluation of either the social payoff or her reputation was 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. Of course, one must bear in mind that firm DMs are often rewarded with performance bonuses while simultaneously being protected from pay-cuts due to limited liability, and this may also explain excessive risk taking, although this may also be mitigated by managerial risk aversion. Similarly, different politicians may face different incentives: a popular politician may be inclined to be risk averse about her reputation, while one who is behind in the polls may take risks to rescue her popularity. These caveats need to be borne in mind while using our model to explain specific situations.

The costs and returns should be interpreted as either accruing to the firm (in the DM example) or to society as large (in the politician context). Assume that $p >> \frac{c}{v}$, so that it is profitable to start the project at date 1.

Recall that prior to taking the decision on whether to continue the project, in period one, the DM privately observes a signal that is informative about the quality of the project. The signal induces a cumulative distribution F over the DM's posterior belief μ that the project is good, with $F : [0,1] \to [0,1]$. For any $z \in (0,1]$, let $F(z^-) := \lim_{x \uparrow z} F(x)$ denote the left-hand limit of F at z, 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 set of posterior beliefs induced by the signal. The signal satisfies the criterion of Bayes-plausibility, i.e.

$$\int_0^1 \mu \, dF = p. \tag{1}$$

Let $\mu^{**} = \frac{c}{v}$. The optimal decision, from a social point of view, is to continue with the project at the end of period 1 if $\mu > \mu^{**}$, and to abandon it if the inequality is reversed. Observe that since the project is initially profitable, $p > \frac{c}{v} = \mu^{**}$. Thus, for any Bayes-plausible information structure, there must be some posterior beliefs such that it is socially optimal to continue the project. We say that 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 will turn out to be 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 not degenerate, i.e. it does not assign probability one to a single value of μ.*

The following examples show the rich variety of belief distributions that can arise under some simple signal structures.

- Suppose that news arrives at the jump times of a standard Poisson process with state-dependent intensity, and the information of the DM is summarized by the number of signals she sees in a unit interval of time. Suppose also that news is *bad*, in the sense that it causes the agent to assign a greater likelihood to the bad state of the world. That is, suppose the arrival rate of news in the bad state, λ_B, is greater than the arrival rate in the good state, λ_G. In this case, there will be critical number, k^*, such that if the number of news events observed between dates 1 and 2 is $k \geq k^*$, then $\mu < \mu^{**}$.

Since there is no upper bound on the number of possible news events, the richness assumption is automatically satisfied.

- Modify the first example, so that the news process is a good news one, with $\lambda_B < \lambda_G$. Now, the socially optimal policy implies a threshold k^*, such that if $k \leq k^*$, then $\mu < \mu^{**}$. The signal is decision-relevant if $k^* \geq 1$, and it satisfies the richness assumption if $k^* > 2$.
- Suppose that the news process is given by a Brownian motion, whose drift differs depending on the state, but whose variance is state independent. The signal is therefore the position of the Brownian motion at the end of the unit interval, and are therefore normally distributed. The set of posterior beliefs is now given by a continuous distribution, and the richness assumption is satisfied.
- Suppose that the agent observes a binary signal, so that the value after one realization is below μ^{**}, while that after the other realization exceeds μ^{**}. Here the richness assumption is not satisfied.

One can of course combine these information structures, so that the DM could observe two news processes, such as a Brownian one and a Poisson. In consequence, there are many plausible belief distributions that the DM could have. In particular, the distribution of beliefs F can be continuous in part, but can also have mass points, and gaps in the support.

2.1 Equilibrium analysis

We analyze Perfect Bayesian Equilibria. A pure strategy for the DM is a function $\sigma : [0, 1] \to \{N, Y\}$, where Y denotes continuing the project at date 2. A mixed strategy maps her beliefs in $[0, 1]$ to a probability of continuation. A strategy for the observer, ρ, specifies an action in $[0, 1]$ for each public event: success, failure, and N (project termination). Since the observer must take action 1 in the case of success, and 0 in the case of failure, it suffices to specify the observer's action when the project is cancelled, and we will denote this by ρ. Sequential rationality implies that it equals the observer's belief when he observes cancellation.

We require σ to be a best response to ρ for every belief in $[0, 1]$. If a belief μ is not in the support of F (so that $\mu \notin \mathcal{C}(F)$), $\sigma(\mu)$ could be sub-optimal. However, it is without loss to require σ to be optimal for all μ values, and we do this to simplify exposition.

Fix an equilibrium, (σ, ρ). If the DM continues the project at some belief μ, then her expected payoff is

$$U(Y, \mu) = \mu[v + \alpha] - c. \tag{2}$$

With probability μ, the project succeeds, and social value rises by v, and the observer's belief jumps to one, since success is perfectly informative of project quality (though not of the DM's ability). But 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, that is, it does not depend upon the observer's beliefs about the strategy played by the DM. In other words, since continuing the project resolves the uncertainty, the DM can fully reveal her type by doing so. Thus, project continuation is analogous to *disclosure with a verifiable type*, as in Grossman (1981) or Milgrom (1981).

However, the DM cannot verifiably disclose her type when he terminates the project, and thus her payoff when he cancels the project does depend upon σ. The net effect on the firm's value is zero, since the cost is not incurred, and nor is there any return. Suppose that, in equilibrium, the DM cancels the project on the set of beliefs Ω be the set of beliefs when she plays σ. If Ω has positive $F-$ measure, $\mathbb{E}(\mu|\Omega)$ is well defined, and will be the observer's belief about the quality of the project when the DM cancels the project. Thus, the DM's payoff from cancelling the project equals $\alpha\rho = \alpha\mathbb{E}(\mu|\Omega)$, and is independent of her belief, μ. If Ω is empty, it is still the case that the beliefs of the observer depend only on the observed cancellation, and are therefore independent of the DM's own belief. Thus the payoff of the DM from stopping is independent of her beliefs in any equilibrium. 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; there exists a threshold x, that depends upon on the equilibrium σ, where he cancels the project if $\mu < x$ and continues if $\mu > x$. The DM may possibly randomize at the belief x, and this may be of consequence when F has an atom at x.

Fix an equilibrium with threshold $x \in (\underline{\mu}, \bar{\mu}]$. Let $\theta \in [0, 1]$ denote the probability with which the DM cancels the project at belief x. If F does not a have a mass point at x, then the payoff from cancelling the project at x is independent of θ and equals

$$U(N, x, \theta) = \alpha\mathbb{E}(\mu|\mu < x).$$

If F has a mass point at x, the payoff from stopping does depend upon the equilibrium probability with which the DM stops at x, and is given by

$$U(N,x;\theta)=\alpha\frac{\int_0^{x^-}\mu dF+\theta\Delta(x)x}{F(x^-)+\theta\Delta(x)},\theta\in[0,1]. \tag{3}$$

It is straightforward to verify that $U(N,x,\theta)$ is continuous and strictly increasing in θ; by choosing θ appropriately, any value between $U(N,x,0)$ and $U(N,x,1)$ can be achieved.

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

$$g(x)=U(Y,x)-U(N,x,\theta),\theta\in[0,1]. \tag{4}$$

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 firm 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 $0\in g(\mu^*)$.[3] We summarize our result in the following proposition.

Proposition 2 *There exists a pure strategy equilibrium with a threshold belief $\mu^*\in(\underline{\mu},\mu^{**})$, where the DM cancels the project if and only if $\mu\leq\mu^*$.*

Equilibrium need not be unique. We examine the implications of equilibrium multiplicity in subsection 2.3.

Henceforth, we may identify an equilibrium by the pair (μ^*,ρ^*). The DM's threshold is denoted by μ^*, and the observer's action/belief when the DM stops is ρ^*, and takes into account any randomization by the DM at the threshold.

We now identify the conditions under which the DM incurs a reputational loss from cancelling the project in an equilibrium with threshold μ^*. Let $\hat{\mu}=\max\{\mu\in\mathcal{C}(F):\mu\leq\mu^*\}$. Thus $\hat{\mu}$ is the largest belief at which the DM cancels the project – it equals μ^* if μ^* belongs to the support of F.

Clearly, $\mathbb{E}(\mu|\mu\leq\hat{\mu})\leq\hat{\mu}$. Observe that if the DM continues the project at $\hat{\mu}$, then she believes that this induces a two-point distribution over observer beliefs over 0 and 1, where the probability of 1 equals $\hat{\mu}$. Thus the DM can never make a reputational gain by stopping – our goal is to identify the circumstances under which she makes a loss.

[3]Note that μ^* need not be in the support of F, since there may be gaps in distribution of beliefs.

2.2 Efficiency

A social planner who seeks to maximize the value of the project would have a threshold μ^{**}. We now turn to the efficiency of equilibria. In any equilibrium, the observer's posterior belief is a martingale, and thus its expectation 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 project under this equilibrium, plus her reputational payoff, which equals αp. Thus the equilibrium that maximizes the project payoff is also the one that is best for the DM.

In any equilibrium where $\hat{\mu}$ is the largest belief at which the DM cancels the project, and where this cancellation takes place with probability θ the reputational cost of cancellation is proportional to

$$\hat{\mu} - \frac{\int_0^{\hat{\mu}^-} \mu dF + \theta\Delta(\hat{\mu})\hat{\mu}}{F(\hat{\mu}^-) + \theta\Delta(\hat{\mu})}.$$

The above expression is always non-negative, since $\hat{\mu}$ is the largest belief at which the DM cancels. This implies that $\hat{\mu} \leq \mu^{**}$, so that the DM never terminates a worthwhile project. If the above expression is strictly positive, then the DM incurs a reputational cost from cancelling the project at $\hat{\mu}$ – her reputational payoff from continuation is $\alpha\hat{\mu}$, while her payoff from stopping is strictly less. Thus it can optimal to stop only if the social cost of continuation is high enough.

Proposition 3 *If the distribution of the DM's interim beliefs is* rich*, then at the threshold belief μ^{**}, the DM suffers a reputational loss from cancelling the project. Thus there exists γ such that if $\alpha > \gamma$, then in any equilibrium the DM continues the project at beliefs when it is unprofitable to do so. If μ^{**} belongs to the support of F, and if there exists an interval (a, μ^{**}) such that F is strictly increasing on this interval, then $\gamma = 0$, so that the DM continues an unprofitable project at beliefs where it is unprofitable, no matter how small her reputational concerns. In any equilibrium, the DM terminates the project at $\underline{\mu}$, the lowest belief in $\mathcal{C}(F)$, no matter how large reputational concerns are. Thus, if the distribution of the DM's interim beliefs is not rich, then any equilibrium is efficient, no matter how large reputational concerns are.*

To summarize:

- When the distribution of beliefs is rich, then there will inefficient continuation in any equilibrium if reputational concerns are sufficiently large, i.e. if α is large enough.

- When the distribution of beliefs is continuous and without gaps at μ^{**}, there will be inefficient continuation no matter how small reputational concerns are.

- When the distribution of beliefs is not rich, so that there only a single belief below μ^{**}, then the unique equilibrium will be efficient.

- Inefficiency is bounded: no matter how important reputational concerns are, the DM never continues the project at the lowest belief, $\bar{\mu}$.

- Inefficiency is one-sided: the DM never terminates a profitable project.

- When there are multiple equilibria, the most efficient equilibrium is the one with the largest threshold.

2.3 Multiple equilibria and comparative statics

While the payoff from continuation is affine in the belief μ, the payoff from stopping at any threshold $\tilde{\mu}$ is not, in general, well behaved, since the conditional expectation $\mathbb{E}(\tilde{\mu}|\tilde{\mu} \leq \mu)$ has upward jumps at mass points of F, and this can give rise to multiple equilibria. As already noted, mass points arise naturally, e.g. with Poisson news. An example is depicted in Figure 1, which graphs the expectation $\mathbb{E}(\tilde{\mu}|\tilde{\mu} \leq \mu)$. Recall that this equals $\frac{1}{\alpha}U(N, \mu; \theta), \theta \in [0, 1]$. The threshold belief μ is on the horizontal axis and the DM's (scaled) payoff from stopping on the vertical axis. The payoff is increasing and jumps up discontinuously at any mass point of the distribution F, such as μ_2^*. The scaled payoff from continuing, $\tilde{U}(Y, \mu) := \frac{1}{\alpha}U(Y, \mu)$ is affine and the figure shows three intersections of the two curves, at μ_1^*, μ_2^* and μ_3^*, corresponding to three distinct equilibrium thresholds. At μ_2, the equilibrium requires that the DM randomize between stopping and continuing. Also, the payoff from stopping is flat in the interval $(\mu_2^*, \mu_3^*]$, corresponding to a gap in the distribution of F in this interval. Thus the equilibrium threshold μ_3^* is equivalent to the DM stopping with probability one at μ_2^*. Mass points are not necessary for multiplicity – it suffices that the conditional expectation $\mathbb{E}(\tilde{\mu}|\tilde{\mu} \leq \mu)$ increases rapidly enough, as depicted in Figure 2, where all three equilibria are in pure strategies.

The nature of equilibria has implications for robustness – is there a nearby equilibrium if there is a small change in parameter values? It also has implications for comparative statics. The following definition is relevant here.

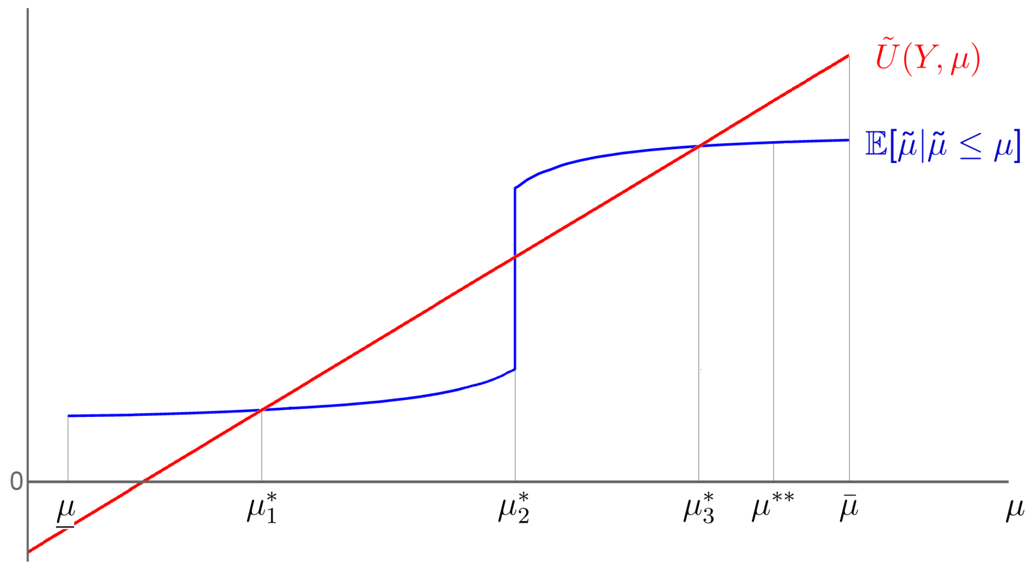

Figure 1:

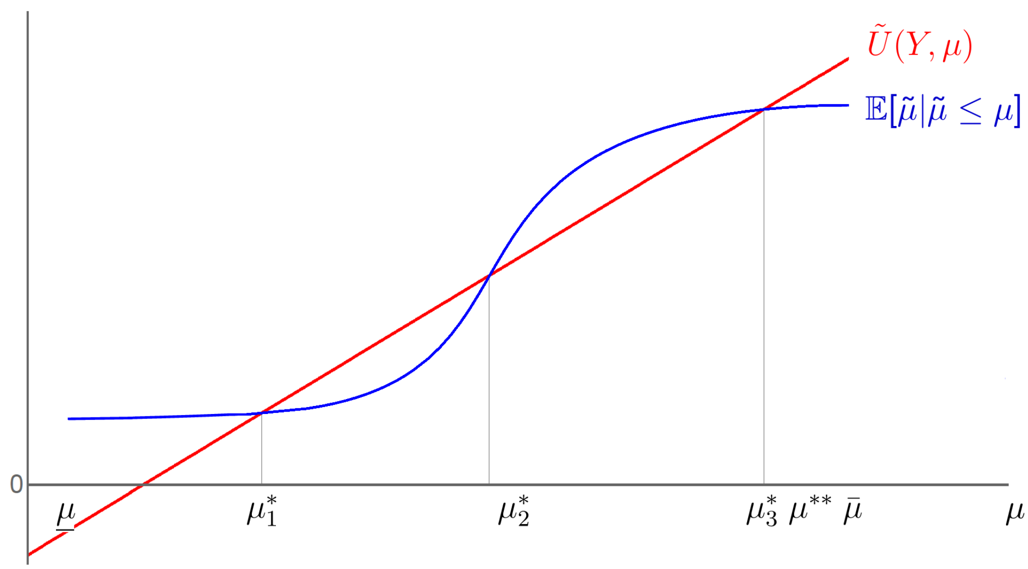

Figure 2:

Definition 4 *Let μ^* be an equilibrium threshold. μ^* is left-stable if there exists an open interval (μ^-, μ^*) such that $U(N, \mu, 0) > U(Y, \mu)$ for any $\mu \in (\mu^-, \mu^*)$. μ^* is right-stable if there exists an open interval (μ^*, μ^+) such that $U(N, \mu, 1) < U(Y, \mu)$ for any $\mu \in (\mu^*, \mu^+)$.*

μ^* is stable if is both left-stable and right-stable. μ^* is unstable if it neither left-stable nor right-stable.

There always exists a stable equilibrium. Generically, there will be an odd number of equilibria, with stable and unstable equilibria alternating. Any mixed strategy equilibrium will be unstable.

Can we provide an argument for selecting equilibria? For example, since the equilibrium with the smallest threshold is the most efficient one, would we expect the players to coordinate on this? Wilson (1980) studies a variant of the Akerlof (1970) lemons market, and shows that explicit price setting by the uninformed side of the market (the buyers) will imply, generically, that only the most efficient equilibrium survives (see also Mas-Colell, Whinston, and Green (1995)). Unfortunately, we cannot make a similar argument here, when the uninformed observer is passive. In Section 4.3, we provide an argument that eliminates unstable equilibria. Also, we show that under some conditions we can select a unique equilibrium, although this turns out to be the most inefficient one.

We now examine the effects of a greater reputational concerns, i.e. a larger value of α, upon equilibria. Since α parameterizes the *relative* importance of reputational concerns as compared to project value, changes in its value can also reflect a reduction in performance related incentives, e.g. for a firm DM.

An increase in α reduces the slope of $\tilde{U}(Y,\mu)$ and also increases its intercept. At μ^{**}, $\tilde{U}(Y,\mu^{**}) = \mu^{**}$, and thus the line swivels around the point (μ^{**},μ^{**}). This increases the equilibrium threshold at any stable equilibrium, μ^*, and reduces it at an unstable equilibrium. Thus if equilibrium is unique, it must be stable, and increased reputational concerns aggravate inefficiency.

2.4 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. Suppose now that when the project is continued, its quality is only revealed with positive probability. With probability λ, the outcome of the project is publicly realized, so that it succeeds with probability μ and fails with probability $1-\mu$. With probability $1-\lambda$, the outcome of the project is not realized. If this is the case, we assume, for simplicity, that the project is scrapped in the following period, so that its continuation social value is zero. Thus there is an additional public event where the quality of the project is not learnt: when the

DM continues the project and the outcome is not realized, and the observer's strategy also requires specifying his action in this event, which must equal his belief conditional on this event. Suppose that the DM follows a threshold strategy, continuing the project if $\mu \geq \mu^*$. Assuming, for simplicity, that there is no atom at the threshold, the payoff from continuation can be written as

$$\mu^* \lambda[v + \alpha] + (1 - \lambda)[\alpha \mathbb{E}(\mu|\mu \geq \mu^*) - c. \tag{5}$$

The payoff from terminating the project is, as before, $\alpha \mathbb{E}(\mu|\mu < \mu^*)$. Thus at the threshold, these two payoffs must be equal. 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 $\bar{\mu}$. The reputational benefit from continuation is now proportional to $\lambda \bar{\mu} + (1 - \lambda)\mathbb{E}(\mu|\mu \geq \bar{\mu})$, and this is decreasing in λ.

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 1, and who is up for re-election or re-appointment at be beginning of period T. In each period $\tau \in \{2, 3.., T-1\}$, 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, is terminated at date T. The flow cost of continuing the project is c per period. Suppose that the DM receives private information regarding the project at the 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 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 $s = 1, 2.., t - 1$, 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 the equilibrium, 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. Let mu_t^* be her equilibrium threshold. This is character

The equilibrium threshold μ_t^* is characterized by the condition,

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

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

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

Thus μ_t^* is decreasing in t: 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. It is as though the DM exhibits the "sunk-cost fallacy" in her behavior, although this is driven by reputational concerns rather than irrationality.

2.5 When the DM knows her ability

How would our analysis be affected if the DM knows her ability? We now consider this situation, where the DM's type belongs to $\{H, L\}$, and is known by the DM but not by the observer. Recall that both p_H and p_L are interior, so that either type of DM has positive probability of having either type of project, and so is uncertain about its quality. Assume that it is optimal for the low ability DM to start the project at date 1. Thus both types of DM will start the project. Now at date 2, after observing signal s, the two types of DM will update differently, with beliefs given by:

$$\mu_\tau(s) = \frac{p_\tau f(s|G)}{p_\tau f(s|G) + (1-p_\tau)f(s|B)}, \tau \in \{G, B\}.$$

Let F_H and F_L denote the distribution of beliefs for 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 2. Observe given any belief μ regarding project quality, the DM's own type is irrelevant in the continuation game. Thus both types of DM will have the same cut-off belief μ^*, that satisfies the same equilibrium condition with respect to F as in our previous analysis.

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 a good one, (i.e. $p_H = 1$), and thus does not update her belief on the likelihood of the project succeeding observing any signal. Thus a competent DM would never cancel a project. In other words, changing your mind regarding a project is a sure sign of incompetence. Our analysis is therefore pertinent in the case of a DM who is as smart of Keynes, but no smarter – there exist facts that would induce her to change her mind on a project that she has initiated.

3 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 we are in a common values environment. The current president of US is a possible case in point. Can our model be used to 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 reputational payoff. Assuming that the payoff is linear in the perceived project quality of the predecessor, the coefficient α is now negative.

Let us assume that the reputational concern is not too large, so that $v + \alpha > 0$. Then the payoff from continuing the project, $U(Y, \mu) = \mu(v+\alpha)$, is strictly increasing in μ. Fix an equilibrium σ so that the payoff from termination is a constant, that depends upon σ, since it must be measurable with respect to the stopping decision. 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. The payoff from terminating the project, $U(N, x, \theta)$, is now strictly decreasing in x at any x that is not a mass point of F. It is also decreasing in θ at any x that is a mass point of F. Thus the correspondence g is strictly increasing in x. Since it is negative at the lower bound of the support (given our assumption that the signal is decision relevant), there are two possibilities: either there is an interior equilibrium (μ^*, θ), or g is always negative, so that the DM always cancels the project. In either event, equilibrium is unique. Observe that in either case, the observer's beliefs are well defined, since the set of beliefs at which the DM cancels the project is non-empty. In an equilibrium where the DM always cancels, the observer's beliefs when she continues are well defined, since this resolves the uncertainty regarding project quality.

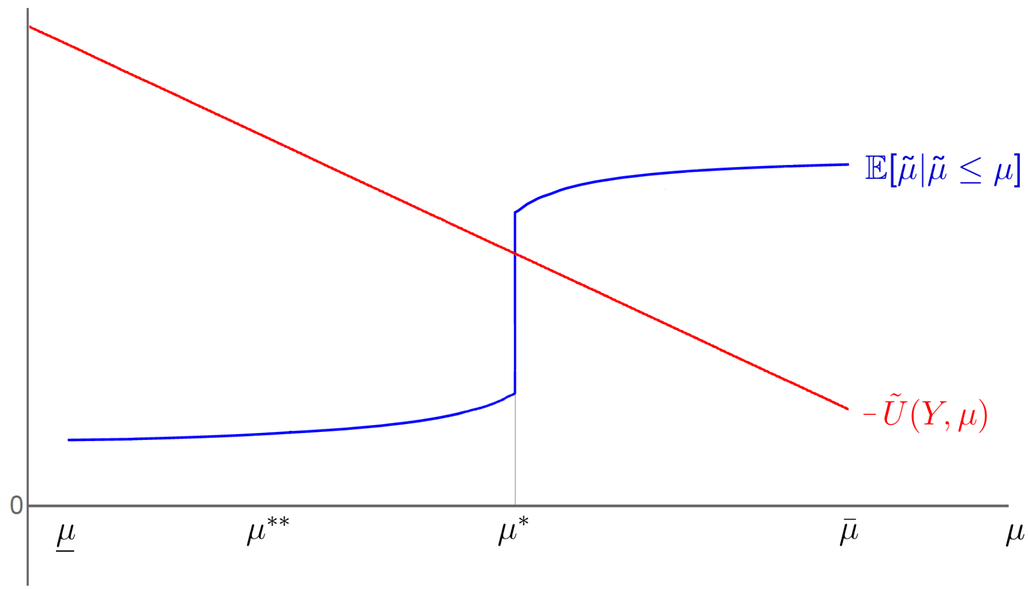

Figure 3:

Figure 3 depicts the equilibrium. The conditional expectation, $\mathbb{E}(\mu|\mu \leq \mu^*)$ now represents the (scaled) *negative* of the payoff from cancellation, while the straight line represents the scaled *negative* of the payoff from continuation. Since the first is upward sloping and the second is downward sloping, uniqueness follows – there is either one intersection point, or first is always below 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 $\mathbb{E}(\mu|\mu \leq \mu^*)$ is always weakly less than μ^*. That is, at the threshold, by cancelling the project, she permits a more adverse inference on her predecessor than she would be continuing the project. This is also true when the DM always cancels – at the highest possible belief, where the payoff from continuing is the greatest, the reputational premium is larger than the social cost of cancelling the project.

Since equiibrium is unique and regular (in the sense of Debreu (1970) or Harsanyi (1973)), the comparative statics properties are also intuitive. Suppose that the absolute value of α increases so that the reputational component has greater weight. The straight-line swivels around the point (μ^{**}, μ^{**}), becoming flatter. Thus the new equilibrium threshold is (weakly) larger, so that the inefficiency increases. So greater reputational concerns give rise to more destructive behavior towards the projects of one's predecessor.

However, there is a qualitative change in equilibrium once reputatational concerns become sufficiently large. Suppose now that α is sufficiently small, so that $v + \alpha < 0$. 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. We show that always continuing the project is an equilibrium (the proof of uniqueness is in the appendix). 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 is the highest belief type, $\bar{\mu}$ – 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 $\bar{\mu}$. $U(Y, \bar{\mu}) > U(N, \bar{\mu})$, since the project is profitable at $\bar{\mu}$, and the reputational payoff is the same under both policies. Since $U(Y, \mu)$ is decreasing in μ while the payoff from cancellation is constant, continuation is strictly better than termination for all other types. To see that the beliefs satisfy D1, consider a distribution over DM types (($\mu-$ values) such that type $\bar{\mu}$ is indifferent between continuing and terminating the project. Let $\tilde{\mu}$ be the expectation of this distribution – clearly $\tilde{\mu} < \bar{\mu}$ since at $\bar{\mu}$ continuation is strictly optimal. The payoff from stopping is the same for any belief type, and equals $-\alpha\tilde{\mu}$, but the payoff from continuing is strictly decreasing in μ. Thus every other type of DM strictly prefers to continue. This verifies that the beliefs satisfy D1.

This result is reminiscent of the unravelling result of Grossman (1981) and Milgrom (1981) in the context of the disclosure of verifiable information. When types are one-dimensional, when the sender has monotone preferences over the beliefs of the receiver, and the disclosure of verifiable information is costless, then the sender fully reveals his type. Here, disclosure is costly – the cost to the DM, of revealing her belief to the observer, is the social cost of continuing the project. (This is true in every version of the model, regardless of whether α is positive or negative.) This cost is highest for the lowest belief, $\bar{\mu}$. Thus if it is optimal for the lowest type for disclose here information, by continuing the project, so does every other type in the unique equilibrium.

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. Thus the theory is not probably the best explanation for President Trump's desire to rescind every initiative of his predecessor. We do not view this as a weakness of the theory, heeding instead Karl Popper's dictum, that "a theory that explains everything explains nothing". In other words, there are probably better alternative explanations of the President's behavior.

Proposition 5 *If $0 < \alpha < -v$, so that the DM has a moderate incentive to sabotage her predecessor's reputation, then there is a unique equilibrium, where the DM cancels some projects that are optimal to continue. If the DM has extreme incentives to sabotage her predecessor's reputation, so that $\alpha < -v$, then there a unique equilibrium where the DM always continues the project, at every belief. If the DM terminates the project, then the observer believes that the DM is the highest belief type, $\bar{\mu}$. These beliefs satisfy the D1 refinement.*

Proof. See appendix. ■

4 Over-confident decision makers

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, let us assume that the DM is over-confident, and has a prior q on the quality of the project, that is strictly greater than p, the prior of the observer. The belief of the DM, on observing signal s, with likelihood ratio $\ell(s)$,[4] is given by

$$\pi(\ell(s)) = \frac{q\ell(s)}{q\ell(s) + (1-q)}, \tag{6}$$

The belief of the observer, if he was to observe signal s, would be given by

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

In the interests of economy, we may as well identify signals with the associated likelihood ratio. It will be useful to expand, if necessary, the space of signals, so that every likelihood

[4] $\ell(s) := \frac{f(s|G)}{f(s|B)}$, where $f(s|\omega)$ denotes the value of the probability density function at s (or the atom at s) when the project quality is ω.

ratio in $[0,\infty]$ is possible, while allowing for some of these signals to have zero probability under both states. The payoff of the DM from continuing the project is an affine function of her belief, π. The payoff from stopping is measurable with respect to this decision and is therefore constant. Thus any equilibrium must be in threshold strategies: there exists some likelihood ratio ℓ_0 such that the DM continues the project for $\ell > \ell_0$ and terminates it for $\ell < \ell_0$, with the DM stopping with probability θ at ℓ_0. When the DM terminates the project, her payoff is given by the observer's belief regarding project quality. Let $\mu_0 := \mu(\ell_0)$. Thus the observer believes that the DM terminates the project if $\mu < \mu_0$, and that at μ_0, she terminates the project with probability θ. Let F denote the distribution μ, i.e. under the prior p. Thus the DM's payoff on termination is proportional to the observer's belief on project quality, and is given by

$$U(N,\mu_0,\theta) = \alpha \frac{\int_0^{\mu_0^-} \mu dF + \theta\Delta(\mu_0)\mu_0}{F(\mu_0^-) + \theta\Delta(\mu_0)}.$$

Let us define $\pi^\dagger(\mu)$ as the DM's belief when the DM's belief is μ. That is $\pi^\dagger : [0,1] \to [0,1]$ is defined for every $\ell \in [0,\infty)$ by 6 with argument μ defined by 7, and takes value 1 when $\mu = 1$. This has the explicit form:

$$\pi^\dagger(\mu) = \frac{q(1-p)\mu}{q(1-p)\mu + (1-q)p(1-\mu)}. \tag{8}$$

The following lemma identifies the critical properties of the function $\pi^\dagger$.

Lemma 6 *The function $\pi^\dagger$ is strictly increasing and strictly concave. For any $\mu \in (0,1)$, $\pi^\dagger(\mu) > \mu$.*

Proof. See appendix. ■

The DM's payoff from continuation, at the threshold belief μ_0 is given by

$$U(Y,\pi^\dagger(\mu_0)) = \pi^\dagger(\mu_0)[v+\alpha] - c.$$

Thus $(\mu_0,\theta_0)$is an equilibrium threshold if

$$\pi^\dagger(\mu_0)[v+\alpha] - c = U(N,\mu_0,\theta_0). \tag{9}$$

Existence of equilibrium follows the same arguments as before; however, if overconfidence

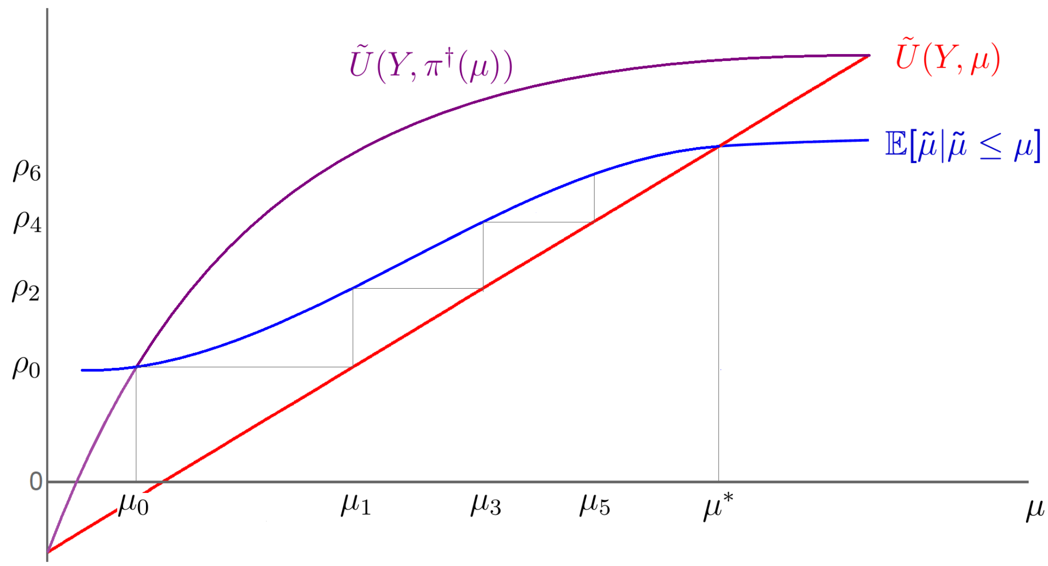

Figure 4:

is sufficiently large, the DM may continue the project even at the lowest possible belief.

Figure 4 depicts equilibrium under overconfidence.The straight line $\tilde{U}(Y, \mu)$, in red, shows the scaled payoff from continuation when the belief is μ, i.e. when there is no overconfidence. The concave function $\tilde{U}(Y, \pi^{\dagger}(\mu))$, in purple, shows the DM's payoff from continuation at an observer belief μ, when the DM's posterior belief is $\pi^{\dagger}(\mu)$. This lies above the straight-line. The equilibrium (μ_0, ρ_0) is where the concave function intersects the conditional expectation, $\mathbb{E}(\tilde{\mu}|\tilde{\mu} \leq \mu)$. This lies to the left of (μ^*, ρ^*), the unique equilibrium under a common prior, where the straight line intersects the conditional expectation. Thus overconfidence leads to excessive continuation, over and above that induced under a common prior.

Overconfidence aggravates the tendency to continue bad projects, for two reasons. The first is straight-forward: since the DM's belief is $\pi^{\dagger}(\mu_0) > \mu_0$, the DM continues since she has more optimistic beliefs. The second reason is more subtle – since the observers know that the DM is overconfident (as compared to their own prior), the inference from project termination is more adverse. Thus the DM knows that he will be penalized more for terminating the project, as compared to the situation where the observer and he share the common prior belief p.

More generally, when there are multiple equilibria under common priors, then if all the equilibria are regular, the effect of a small amount of overconfidence (i.e. of a q close to p)

will be so as to have move any equilibrium threshold slightly. The effect of small overconfidence is to raise the equilibrium threshold at any stable equilibrium and to reduce it at any unstable one. That is, at a stable equilibrium, one has the intuitive comparative statics, that overconfidence gives rise to greater project continuation. At an unstable equilibrium, overconfidence reduces project continuation.

4.1 Perceived Overconfidence

We now analyze the game G^1 where the DM and observer share the same prior, p regarding the project. However, the outside observer believes that the DM is overconfident, and has a prior $q > p$. We assume that the DM is aware of this belief. That is, we assume that following statement, **T1** is true:

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

Assume that 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 in the previous section, an alternative formalization of the game G^1 is as follows:

- The observer believes that the game G^0 is being played.
- The DM has a prior p and knows that the observer's believes that G^0 is being played.

Let (σ_0, ρ_0) be an equilibrium of the game G^0. That is, σ_0 is a strategy for the DM, and ρ_0 is the strategy of the observer. In our specific context, ρ_0 is an action that equals the belief of the observer when the DM terminates the project. An equilibrium of the game G^1 is a triple consisting of (σ_0, ρ_0) and σ_1, where:

- σ_0 and ρ_0 are mutual best responses.
- σ_1 is a best response to ρ_0 given the prior p.

Observe that in an equilibrium of the game G^1, the observer's strategy is the same as that in an equilibrium of the game G^0 – in both cases, it is ρ_0. The DM's strategy differs across the two games. This is a feature that will recur in our analysis later.

More specifically, fix an equilibrium threshold in the game G^0, (μ_0, θ_0). This determines the observer's belief on project quality in the game G^0, and his action ρ_0. Let ℓ_1 denote the equilibrium cut-off signal of the DM in the game G^1, and let and let $\mu_1 := \mu(\ell_1)$. This must satisfy the condition:

$$\mu_1[v+\alpha] - c = U(N, \mu_0, \theta_0) = \alpha\rho_0. \tag{10}$$

In addition, $(\mu_0, \theta_0; \rho_0)$ is given by 9. Since the left-hand side of the above equation is strictly increasing in μ_1 and the right-hand side is independent of μ_1, there is a unique solution. We now make the following assumption:

Assumption 7 *F does not a have a mass point at μ_1.*

Under the above assumption, the DM's behavior at his point of indifference, μ_1, is irrelevant. Thus equilibrium consists of a pair, μ_1 and $(\mu_0, \theta_0; \rho_0)$ that satisfy equations 10 and 9. In other words, for any equilibrium $(\mu_0, \theta_0; \rho_0)$ of the game G^0 with overconfidence, there is a unique equilibrium in the game G^1.

This is illustrated in Fig. 4. Corresponding to the unique equilibrium μ_0, there is the corresponding belief of the observer on termination, $\rho_0 = \mathbb{E}(\mu|\mu < \mu_0)$. The DM chooses the belief threshold μ_1 such that the payoff from continuation, given by the straight line, equals this conditional expectation. Thus $\mu_1 > \mu_0$, since $\mu_1 = \pi^\dagger(\mu_0) > \mu_0$. Thus the equilibrium *outcome* in the game G^1 is depicted by the pair, (ρ_0, μ_1). [5]

Observe that $\pi^\dagger(\mu_0) = \mu_1$, so the *equilibrium cut-off beliefs are identical* for the DM in the two cases, of actual overconfidence and perceived overconfidence. Of course, $\ell_1 > \ell_0$, and so the DM who is overconfident and perceived to be overconfident cancels the project less frequently than the DM who is only perceived to be overconfident. Nonetheless, the perception of overconfidence penalizes 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. Fix any such equilibrium, $(\mu_0, \theta_0; \rho_0)$. This induces a unique threshold $\mu_1 > \mu_0$, where $U(Y, \mu_1) = U(Y, \pi^\dagger(\mu_0))$. If assumption 7 is satisfied and there are no mass points at any of the induced thresholds, this implies that the behavior of the DM is uniquely determined, and one has the same number of equilibria in G^1 as in G^0, but each induced threshold is strictly greater than the one inducing it.

For generic payoffs, there will be finitely many equilibria in G^0, and therefore finitely many induced thresholds. Since the number of mass points of F must be at most countable, assumption 7 will be satisfied generically for every induced threshold. Nonetheless, it is instructive to consider what happens when the assumption is violated, and there is an atom

[5] Recall that the *outcome* is the distribution over terminal nodes, i.e. a joint distribution over states and player actions.

at the induced threshold, μ_1.

Note that the observer's beliefs and thus his best response do not depend upon the DM's actual behavior in G^1, and are determined in the game G^0. Thus the DM's decision at μ_1 does not affect the observer's best response. When there is a mass point at μ_1, this is of consequence: the DM's mixing probability at μ_1 can be arbitrary, since she is indifferent. Consequently, there is a continuum of equilibria $(\mu_1, \theta_1), \theta_1 \in [0, 1]$, corresponding to any $(\mu_0, \theta_0; \rho_0)$ in G^0. The possibility of a continuum of equilibria is illustrated in Figure 5 . The game with common priors has two equilibria, at thresholds μ_1^* and μ_2^*. There is a mass point μ_1^* and also a gap in the support of F immediately to the left, so that the conditional expectation is flat in a neighborhood to the left. In the game with overconfidence, where q is larger than p, there is a unique equilibrium, (μ_0, ρ_0). Now consider the behavior induced in G^1 by μ_0. The DM's optimal threshold is μ_1^*, but any mixing probability at the threshold is consistent with equilibrium. Thus one has a continuum of equilibria, and also equilibrium outcomes.

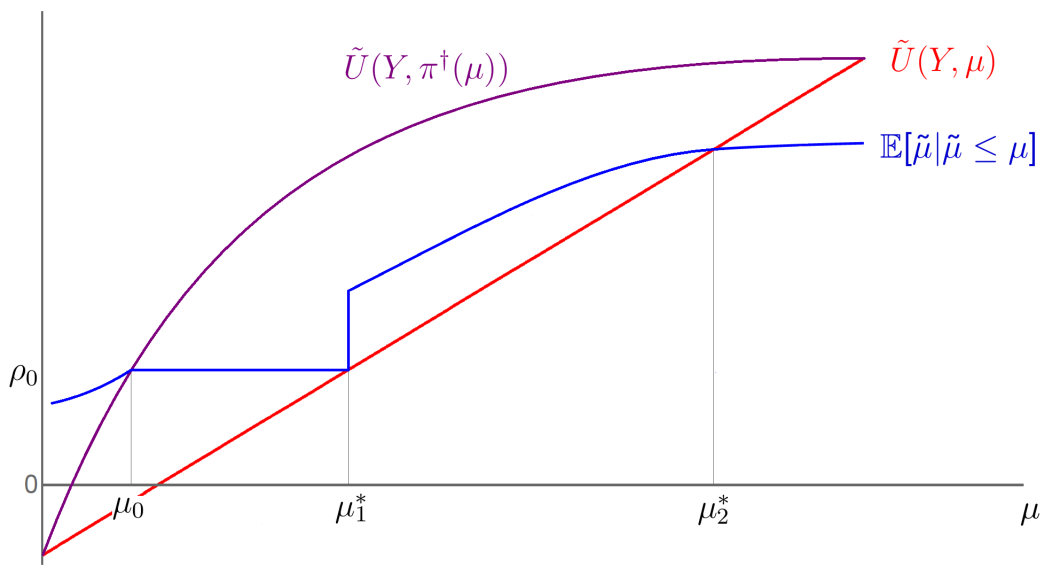

Figure 5:

4.2 Perceived perceived 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 other words, we have the DM's perception of perceived overconfidence.

Specifically, we assume that the following statements are true:

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

T2 The DM believes that T1 is true.

Recall that T1 stated that the observer believes that the DM's prior is q. That is, T2 says that the DM believes that the observer believes that the DM's prior is q. We assume that T2 is known by the observer.

In other word's the first order and second order beliefs agree but the third order belief does not.

Alternatively:

1. The observer believes that the DM's prior is p.
2. The DM believes that the game G^1 is being played.
3. The observer knows (2).

Fix an equilibrium of the game G^1, i.e. a triple $(\sigma_0, \rho_0, \sigma_1)$. Since the DM believes that the game G^1 is being played, σ_1 continues to be optimal in G^2, given (σ_0, ρ_0). Let ρ_2 be a best response by the observer to σ_1. Thus the quadruple, $(\sigma_0, \rho_0, \sigma_1, \rho_2)$ is an equilibrium of the game G^2.

Specializing to our specific context, and invoking assumption 7,[6] an equilibrium of the game G^1 consists of a triple $(\mu_0, \rho_0; \mu_1)$. Thus in the game G^2, the DM uses the same threshold (μ_1). However, the observer's evaluation of the DM is different in the games G^1 and G^1. Since the observer knows that the DM uses the threshold μ_1, the belief about the DM is given by $\mathbb{E}(\mu|\mu < \mu_1)$, which also equals his optimal action ρ_2. Returning to Fig. 4 , the equilibrium outcome in G^2 corresponds to the pair (μ_1, ρ_2).

4.3 Higher order beliefs about overconfidence

Generalizing this argument, let us consider a game where G^N, defined as follows. First, in every game in this class, the DM and observer have common prior q.

Define the following statements:

[6] I.e. there is no mass point at μ_1.

- **S1:** The observer believes that the DM's prior is p.
- If K is even, define **SK** by : **SK** The DM believes that **S(K-1)** is true.
- If K is odd, define **SK** by : **SK** The observer believes that **S(K-1)** is true.

We also define the statements **TK** as follows:

- **T1** The observer believes that the DM's prior is $q > p$.
- If K is odd, define **TK** by : **TK** The observer believes that **T(K-1)** is true.
- If K is even, define **TK** by : **TK** The DM believes that **T(K-1)** is true.

For every $N \in \mathbb{N}$, the game G^N is defined by statements **S1** to **S(N-1)** being true, and **T(N)** being true. Furthermore, **T(N)** is known by both players.

Consider the sequence of games, $G^N, N \in \{0\} \cup \mathbb{N}$. This defines a sequence of strategies $(\sigma_0, \rho_0), (\xi_N)_{N \in \mathbb{N}}$, where $\xi_N = \sigma_N$ if N is odd and $\xi_N = \rho_N$ if N is even. An equilibrium of the game G^N consists of the sequence truncated at ξ_N, with the property that:

- (σ_0, ρ_0) is an equilibrium of the game G^0.
- For any $n \leq N$, ξ_n is a best response to ξ_{n-1}.

Note that when n is an odd number, ξ_n can be written as (μ_n, θ_n), i.e. a threshold for the DM and a mixing probability at the threshold – the latter is relevant only if μ_n is a mass-point of F. If there is a mass point, then θ_n can be arbitrary, and its value is relevant for determining ρ_{n+1}. When n is an even number, ξ_n equals ρ_n, and is uniquely determined by $(\mu_{n-1}, \theta_{n-1})$.

The following proposition characterizes the limit behavior of the sequence of equilibria.

Proposition 8 *Fix an equilibrium of the game with overconfidence (μ_0, θ_0). This induces a sequence of equilibrium thresholds (μ_t, ρ_{t+1}) for every odd number t, such that for any game G^t, the equilibrium outcome is (μ_t, ρ_{t-1}), and for any game G^{t+1}, the equilibrium outcomes (μ_t, ρ_{t+1}). The subsequences $< \mu_t >$ and $\rho_t >$ are each increasing and converge to μ^* and ρ^* respectively, where (μ^*, ρ^*) is a left-stable pure strategy equilibrium of the game with common priors, G.*

Proof. See appendix. ■

When is the sequence induced by a given (μ_0, θ_0) unique? Intuitively, whenever there is no mass point at any of the induced values μ_t. We make this precise as follows.

Fix the first element of the sequence, $((\mu_0, \theta_0), \rho_0)$. ρ_0 induces a unique value μ_1. Assume 7 so that there is no mass point at μ_1. Then (μ_1) uniquely defines ρ_2, which in turn uniquely defines μ_3.

Assumption 9 *Let N be any even integer. Suppose that the sequence $((\mu_0, \theta_0), \rho_0; (\xi_t)_{t=1}^N)$ is defined, and that there is no mass point at μ_t for all $t < N$. Define μ_{N+1} to be the unique value of μ that solves $U(Y, \mu) = \alpha\rho_N$. Assume that F does not have a mass point at μ_{N+1}.*

Proposition 10 *Fix an equilibrium of the game with overconfidence (μ_0, θ_0), and suppose assumption 9 holds. The induced sequence of equilibrium thresholds ξ_t, where $\xi_t = \mu_t$ if t is odd and $\xi_t = \rho_t$ for t odd, is unique. Consequently, the limit points μ^* and ρ^* are also unique.*

Fig. 4 illustrates the uniqueness result. The values of μ_t for t odd and ρ_t for t even, are depicted, and are unique, and converge to the equilibrium of the game with common priors. Fig. 5 illustrates the possibility of non-uniqueness when assumption 9 is violated. Since there is a mass point at μ_1, the value of θ_1 matters. If $\theta_1 = 0$, and also $\theta_t = 0$ for all odd values of t, then we have convergence to (μ_1^*, ρ_1^*). But if we have $\theta_t > 0$ for any t, then the sequence converges to (μ_2^*, ρ_2^*).

Observe that in the game with common knowledge of common priors,, G,the equilibrium with the lowest threshold, μ_1^*, is always left-stable. Assume that there is no point of F at μ_0. The following proposition shows that if the difference $q - p$ is large enough, i.e. if the DM is overconfident enough in the game with different priors, G^1, then μ_1^* is uniquely selected as we allow for higher levels of knowledge that the DM is not overconfident.

Proposition 11 *Assume that there is no mass point at μ_1^*, the equilibrium with lowest threshold in the game G with common knowledge of common priors. If $q - p$ is large enough, then any sequence of equilibria μ_t in the games G^t converges to μ_1^*, the most inefficient equilibrium in G.*

Proof. From the expression for $\pi^\dagger(\mu)$, observe that for any interior value of μ $\pi^\dagger(\mu) \to 1$ as $q \to 1$. Thus for q sufficiently large, any equilibrium threshold in the game G^0 is less than μ_1^*, and thus the induced sequence from any initial value must converge to μ_1^*. ■

5 Appendix

Proof of Proposition 5

Proof. Since all other parts of the proposition have been proved in the text, now show that equilibrium is unique when $v < \alpha$. Consider a candidate equilibrium where the DM cancels the project with positive probability at some belief. The payoff from cancellation is constant in the DM's belief μ, while that from continuing is strictly decreasing in μ. Thus any equilibrium where the project is cancelled with positive probability is characterized by a threshold μ^*, where if $\mu > \mu^*$, the DM cancels the project, while he continues if $\mu < \mu^*$. Thus, if there is no mass point at μ^*, the DM's payoff when he cancels the project equals (modulo the scaling factor α), $\mathbb{E}(\mu|\mu > \mu^*)$, which lies in $(\mu^*, \bar{\mu}]$, the set of types that cancel. Thus the type $\mu = \mathbb{E}(\mu|\mu > \mu^*)$ must find it optimal to cancel. Since such a type gets the same reputational payoff from both actions, it follows that cancelling must be socially optimal, i.e. $\mathbb{E}(\mu|\mu > \mu^*) \leq \mu^{**}$. However, $\mathbb{E}(\mu|\mu > \mu^*) \geq \mathbb{E}(\mu) = p > \mu^{**}$, since the project is assumed to be ex ante profitable. If there is a mass point at μ^*, the only modification in the above argument is that the DM's reputational payoff is given by

$$\frac{\int_{\mu^{*+}}^{\bar{\mu}} \mu dF + \theta\Delta(\mu^*)\mu^*}{1 - F(\mu^*) + \theta\Delta(\mu^*)}, \tag{11}$$

for some $\theta \in [0, 1]$. Since this also greater than p, the same argument applies. ■

Proof of Lemma 6

Proof. 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 $\pi^\dagger$ is strictly decreasing. Since $\pi^\dagger$ is strictly concave, with $\pi^\dagger(0) = 0$ and $\pi^\dagger(1) = 1$, $\pi^\dagger(\mu) > \mu$ for any $\mu \in (0, 1)$. ■

Proof of Proposition 8

Proof. The payoff $U(Y, \mu)$ is affine with a positive slope, $v + \alpha$. The payoff $U(N, \mu)$ is singleton valued at any μ that is not mass point, is convex valued at any mass point, and is constant along any interval of μ values that has $F-$measure zero. It is strictly

increasing on any interval that has positive $F-$measure. Let (μ_0, θ_0) be an initial value, corresponding to an equilibrium of the game G^0, [7] so that $U(N, \mu_0, \theta_0) = U(Y, \hat{\pi}(\mu_0))$. Since $\mu_0 < \hat{\pi}(\mu_0)$, it follows that $U(N, \mu_0, \theta_0) > U(Y, \mu_0)$. Define μ_1 to be the value of μ such that $U(Y, \mu) = U(N, \mu_0)$. μ_1 is uniquely defined since $U(Y, \mu)$ is a strictly increasing function, and since $U(N, \mu_0, \theta_0) > U(Y, \mu_0)$, $\mu_1 > \mu_0$. Let θ_1 be arbitrary – its value is irrelevant unless F has a mass point at μ_1. Thus (μ_1, θ_1) is an equilibrium threshold of the game G^1. The values μ_t for integers t greater than 1 are defined recursively. If $U(Y, \mu_{t-1}) = U(N, \mu_{t-1}, \theta_{t-1})$, then $\mu_t = \mu_{t-1}$, and $\theta_t = \theta_{t-1}$ and the sequence becomes constant. If this equality is not satisfied, then μ_t is the value of μ such that $U(Y, \mu) = U(N, \mu_{t-1}, \theta_{t-1})$, and θ_t is arbitrary. We show now that the sequence $< \mu_t >$ is an increasing sequence. We have established that $\mu_1 > \mu_0$. Assume, by the induction hypothesis, that the sequence μ_τ is weakly increasing for all integers $\tau < t$. Now $\mu_{t-1} \geq \mu_{t-2}$. If this inequality is strict, this implies that $U(N, \mu_{t-1}, \theta_{t-1}) > U(N, \mu_{t-2}, \theta_{t-2})$. Since $U(Y, \mu)$ is strictly increasing, and since $U(Y, \mu_t)$ must equal $U(N, \mu_{t-1}, \theta_{t-1})$, this implies that $\mu_t > \mu_{t-1}$. If $\mu_{t-1} = \mu_{t-2}$, then $\theta_{t-1} = \theta_{t-2}$ and the sequence becomes constant. Thus $< \mu_t >$ is an increasing sequence, and since it is bounded, it must converge to some value μ^∞. Thus $U(Y, \mu^\infty) = U(N, \mu^\infty, 1)$, so that μ^∞ must be an equilibrium value of the original game. To show that the equilibrium μ^∞ must be left-stable, observe that along the increasing convergent sequence $< \mu_t >, U(N, \mu_t, \theta_t) \geq U(Y, \mu_t)$.
■

References

Akerlof, G. A. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84, 488–500.

Cho, I.-K., and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2), 179–221.

Debreu, G. (1970): "Economies with a finite set of equilibria," *Econometrica*, pp. 387–392.

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.

[7] If μ_0 is not a mass point of F, θ_0 is irrelevant and can be arbitrary.

Downs, 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.

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.

Gibbons, R., and K. J. Murphy (1992): "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 100(3), 468–505.

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.

Harsanyi, J. C. (1973): "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points," *International Journal of Game Theory*, 2(1), 1–23.

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.

Mas-Colell, A., M. D. Whinston, and J. Green (1995): *Microeconomic Theory*. Oxford University Press, New York, NY.

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.

MORCK, R., A. SHLEIFER, AND R. W. VISHNY (1990): "Do managerial objectives drive bad acquisitions?," *The Journal of Finance*, 45(1), 31–48.

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.

WILSON, C. (1980): "The nature of equilibrium in markets with adverse selection," *The Bell Journal of Economics*, pp. 108–130.