Why Did the Bankers Behave So Badly?[1]

Hamid Sabourian, University of Cambridge[2]

Anne C. Sibert, Birkbeck, University of London and CEPR[3]

25 May 2011

[1]We are grateful to Willem Buiter, Huw Dixon and Mike Oaksford for helpful comments.

[2]Address: Faculty of Economics, Sidgwick Avenue, Cambridge CB3 9DD, UK; Email: hamid.sabourian@econ.cam.ac.uk

[3]Address: School of Economics, Mathematics and Statistics, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK; Email: asibert@econ.bbk.ac.uk.

Abstract

It is widely believed that bankers played an important role in causing the financial crisis that began in August 2007. In this paper we demonstrate that the compensation system in the financial services industry which rewards perceived talents, rather than long-term performance, leads rational bankers to exhibit belief persistence, overconfidence and confirmation bias.

JEL classification: D81, D82, D83, G21
Key words: confirmation bias, belief persistence, overconfidence, signalling, financial crisis.

1 Introduction

There were many causes of the credit crisis that erupted in the summer of 2007, but it is widely believed that the behaviour of the employees of financial institutions – hereafter referred to as 'bankers' – played a prominent role. It is argued that these bankers gambled on a continuation of the US housing boom long after most economists predicted its demise. They were overly optimistic, allowing leverage ratios to reach unsustainable levels. They purchased collateralised debt obligations and asset-backed securities and appear to have deliberately avoided investigating the details of the underlying assets.[1]

Given that bankers are not perceived as particularly unintelligent, how was it that they engaged in such apparently irrational behaviour? One explanation is that humans are prone to cognitive errors involving biases toward their own prior beliefs. A vast social psychology literature documents that people tend to make the type of errors that the bankers made: they fail to put sufficient weight on evidence that contradicts their own initial hypotheses, they are overconfident in their own ideas and they have a tendency to avoid searching for evidence that would disprove their own theories. These types of errors are known as *confirmation bias* and a number of reasons have been advanced for why they are made. One explanation is emotional factors: it appears that many politicians and much of the public believe that greed played a role in the case of the bankers. It is also argued that the evolutionary development of the human brain has facilitated the ability to use experience-based techniques which provide good judgements quickly, but which can also lead to systematic biases, and some analysts have suggested that these 'fast and frugal' heuristics were to blame for the bankers misbehaviour.[2] Coates and Herbert (2008) advance the notion that steroid feedback loops may explain why male bankers exhibited overconfidence when caught up in a bull market.

The rewards system in the financial services industry has also been blamed. Pay is based on a bonus system that depends on perceived talents, rather than on long-term

[1] See Rajan (2008) for a discussion of this.

[2] See, for example, Vanguard Research (2011).

results. Bankers who are viewed as exceptionally talented receive vast rewards, lest they be snatched away by competitors; those viewed as less able quickly find themselves unemployed. Apparently, as a consequence of this system, bankers have an incentive to distort their behaviour and to act in a way that - somehow - makes them look competent, even though it leads to bad results in the long run. Discussing bankers' avid participation in the subprime mortgage market, Allan Meltzer remarked, 'These are my MBA students, not just mine but MBAs from Harvard, Stanford, Pennsylvania. They were buying and selling this garbage. Are they so stupid? They got compensated for doing it. If they didn't do it they'd lose their jobs.'[3]

The purpose of this paper is to demonstrate that in many instances what seems to be irrational behaviour in the form of confirmation bias may instead be rational behaviour in response to the compensation system in the financial services sector. To show this, we consider a scenario where an individual takes an action, such as making a prediction, and the consequences of this action are not known until sometime in the future. The individual cares, not just about making the best choice, but also about how competent he is perceived to be in the period between when he acts and when the consequences of his action are revealed. It is demonstrated that the individual's incentive to manipulate beliefs about his ability leads him to distort his actions in a way that is observationally equivalent to confirmation bias.

We present three variants of a simple model where an expert, who we think of as a banker, takes an observable action. Experts differ in their ability to take the correct action and this ability is their private information. We model this by assuming that prior to selecting an action the expert receives a noisy signal indicating which action is likely to be best. The probability that the signal is correct is viewed as the expert's competency and it and the signal are known only to the expert. In the long run, it is learned whether the action taken by the expert is the best one or not and at this later time the expert receives a payoff that is higher if he chose correctly than if he did not. In the short run,

[3] Quoted in Samuelson (2008).

however, it is not known if the expert made the right choice and his reward depends instead upon how competent he is perceived to be.

In the first variant of the model, presented in Section 2, the expert chooses an action, which we view as making a forecast, and is then confronted with publicly observable conflicting information of known quality. He must then choose whether or not to change his forecast. We show that relatively able experts, those whose signals are of better quality than the public signal, maintain their original prediction as do some or all of the experts whose signals are of lower quality than the public signal. The payoff to masquerading as a more competent expert exceeds the benefit of making a choice that is more likely to be correct. Strikingly, we find that even when the public signal is almost certainly correct, it is possible for all experts to persist in forecasting an event that will almost certainly not occur. Thus, the payoff structure leads to behaviour which looks like the type of confirmation bias that is known as *belief persistence*. The model in this section is related to those in the anti-herding literature, such as Avery and Chevalier (1999) and Levy (2004) and we compare these models to ours.

In the second variant of the model, presented in Section 3, the expert receives his private information and chooses an action. He is then asked how likely he thinks it is that he has chosen the best action. In the long run, if his action turns out to be wrong, then he bears a cost that is increasing in the likelihood that he said that he had chosen correctly. Even though it is potentially costly to the expert to claim that it is likely that he made the best choice and there is no intrinsic benefit to doing so, if the payoff to being viewed as competent is high enough, experts will claim to be more certain than they are. Relatively competent experts will pool, all claiming to be correct with certainty. Less competent experts will separate, but they too all overstate their ability. This closely resembles the type of confirmation bias known as *overconfidence*. Alternative views of overconfidence are presented by Brocas and Carrillo (2002) and Van den Steen (2004), discussed in Section 3.

In the third variant of the model, presented in Section 4, the expert receives his private information and chooses an action. He is then given the opportunity to acquire costly additional information which, if his initial choice is incorrect, might confirm this. The expert can then choose whether or not to pursue his initial action. In this scenario, relatively competent experts value the additional information less than less competent experts because they are unlikely to be wrong and, hence, unlikely to learn anything. If the most competent expert chooses not to acquire the additional information, then, whether or not acquiring the additional information is observable, a range of less competent experts will pool with the most competent experts and also choose not to acquire more information. This is similar to the classical form of confirmation bias: a tendency to fail to search for disconfirming evidence.

Our paper concludes with Section 5. There we consider some alternative applications of our framework.

2 Belief Persistence

> 'And I think 2000 will be a good year as well.' Abby Cohen, famously bullish partner at Goldman Sachs, 1999[4]

Suppose that you are a banker who has invested in a particular asset and that new data has emerged making it apparent that the price of this asset is unsustainably high. You realise that your investment decision was likely the wrong one. Do you reverse your position in response to the new information? Or, would you reason that your boss would view you as incompetent – first buying and now sell selling: that changing your position would threaten a sizable bonus payment and perhaps your job. The incentive to remain long and collect a large short-term payment may outweigh the likely eventual cost to your reputation when it is revealed that you made the wrong choice.

In this section, we present a model of optimising experts who cling to their beliefs in the face of contradictory evidence. We suppose that an expert makes a forecast and is

[4] Quoted in Gilpin (1999)

then presented with conflicting evidence, after which he has the opportunity to continue with his initial forecast or to change it. In particular, we have in mind a banker who makes a forecast so that his customers or his employer can make the best investment decision. Eventually, it will be revealed whether or not the banker is correct. While the banker would prefer to be later proven right than wrong, his bonus in the meantime depends upon how competent he is perceived to be. We demonstrate that even if new evidence makes an expert believe that his initial forecast is likely to be incorrect, the desire to be seen as competent may prevent him from revising his prediction. Thus, although he is entirely rational, he exhibits behaviour that is observationally equivalent to the cognitive error of belief persistence.

Formally, we assume that one of two events will occur and that *ex ante*, each is equally likely. Initially, the expert receives a signal indicating which of the events will occur. Both this signal and the probability that it is correct, denoted by π, are the expert's private information. We refer to π as the expert's *competency* and it is common knowledge that it is drawn from a uniform distribution on $[1/2, 1]$.

After receiving his signal the expert publicly forecasts which of the two events will occur. Then, a publicly observed noisy signal either agrees or disagrees with the expert's forecast. This public signal is correct with commonly known probability $\pi^p \in [1/2, 1)$. After observing the public signal, the expert makes a second forecast, either persisting with his original forecast or changing it. After observing his decision, the market updates its beliefs about the expert's competency. Sometime in the future, the event occurs and is observed.

The expert's (discounted) payoff is $\chi\Pi + P$, where Π is the market's assessment of the expert's competency after he has made his second forecast, but before the event occurs, and P is a variable that equals one if the expert's forecast later turns out to be correct and zero otherwise. We refer to Π as the expert's (short-term) *reputation*. The strictly positive parameter χ is the weight that the expert puts on his reputation relative to his desire to forecast correctly.

The form of the objective function, in particular, the short-run payoff – the bonus – for perceived competency, is exogenous to our model and chosen because it appears to mimic real-world objective functions in the financial services industry. It is possible that this is a consequence of bankers being unable to commit themselves to long-run employment in a firm and a lack of firm-specific human capital in the industry. It is assumed that the expert's competency matters to the firm in other ways than his ability to forecast. This is because, as will be seen in this section, more competent experts do not necessarily make better second-round forecasts.

The equilibrium concept used throughout the paper is the natural one for signalling games: the perfect Bayesian equilibrium concept. In signalling games, the first player is the sender of a signal. He has private information about his type and he chooses a strategy. Here, player one is the expert who has private information about his competency. His strategy is a probability distribution over his two possible actions: to persist with his original forecast in the face of conflicting public information, or to change it. Player two, here the market, has prior beliefs about the sender's type and these prior beliefs are common knowledge. Player two makes a conjecture about how player one's strategy depends upon player one's type. Then, after observing player one's action, player two updates his beliefs using Bayes' rule. It is required that player one chooses a strategy that maximises his welfare, taking into account player two's conjecture about his strategy and how his action will affect player two's posterior beliefs. Player two's conjecture about player one's strategy must turn out to be correct.[5]

As he has no other information and as his priors are flat, the expert initially forecasts the event that his signal favours. If the public signal favours the same event, then he has no reason to change his forecast. We consider the more interesting case where the public signal does *not* favour the same event. Given that the public information favours an event at odds with the expert's original forecast, the market conjectures that the probability that an expert with competency π does *not* change his forecast is $\Psi_c(\pi)$. The market

[5] See Fudenberg and Tirole (1992).

observes the action A of the expert: either he does not change his forecast ($A = N$) or he does change it ($A = C$). Then the market updates its beliefs about the expert's competency. The market's conjectured joint probability density function of π and A is denoted by $h(\pi, A)$. The marginal density of A is denoted by $h(A)$. Thus, in accordance with Bayes Rule, the conditional probability density function of π given A and is

$$\begin{aligned} h(\pi|A) &= h(\pi, A)/h(A) = h(\pi, A) \Big/ \int_{1/2}^{1} h(p, A)\, dp \\ &= \begin{cases} \frac{\Psi_c(\pi) g(\pi, \pi^p)}{\int_{1/2}^{1} \Psi_c(p) g(p, \pi^p) dp} \text{ if } A = N \text{ and } \int_{1/2}^{1} \Psi_c(p)\, g(p, \pi^p)\, dp > 0 \\ \frac{[1-\Psi_c(\pi)] g(\pi, \pi^p)}{\int_{1/2}^{1} [1-\Psi_c(p)] g(p, \pi^p) dp} \text{ if } A = C \text{ and } \int_{1/2}^{1} [1 - \Psi_c(p)]\, g(p, \pi^p)\, dp > 0, \end{cases} \end{aligned} \tag{1}$$

where $g(\pi, \pi^p)$ is the *ex ante* probability that the public signal differs from the expert's signal when the expert has competency π. We have

$$g(\pi, \pi^p) = \pi (1 - \pi^p) + (1 - \pi)\, \pi^p. \tag{2}$$

Given that the public information differs from the expert's original forecast, if the market observes action A then its expectation of the expert's competency is

$$\Pi^A = \int_{1/2}^{1} p h(p|A)\, dp. \tag{3}$$

Using Bayes Rule, after observing both his private signal and the conflicting public signal, the expert believes that his original forecast is correct with probability

$$\theta(\pi, \pi^p) \equiv \frac{\pi(1 - \pi^p)}{\pi(1 - \pi^p) + (1 - \pi)\, \pi^p}. \tag{4}$$

Thus, the expected payoff to an expert with competency π of choosing action A is

$$\begin{cases} \chi \Pi^N + \theta(\pi, \pi^p) \text{ if } A = N \\ \chi \Pi^C + 1 - \theta(\pi, \pi^p) \text{ if } A = C. \end{cases} \tag{5}$$

By equation (5), the expert maximises his payoff if and only if

$$\chi\Pi^{N}+\theta\left(\pi,\pi^{p}\right)\left\{\begin{array}{c}>\\=\\<\end{array}\right\}\chi\Pi^{C}+1-\theta\left(\pi,\pi^{p}\right)\text{ and }\Psi\left(\pi\right)\left\{\begin{array}{c}=1\\\in[0,1]\\=0\end{array}\right\}, \tag{6}$$

where $\Psi(\pi)$ is the actual probability that an expert with competency π does not change his forecast. By equation (4), $1-2\theta(\pi,\pi^p)$ is strictly decreasing in π on $\left[\frac{1}{2},1\right]$, with $1-2\theta(\pi^p,\pi^p)=0$. Thus, as the expert takes $\chi\left(\Pi^N-\Pi^C\right)>0$ as given, his optimisation problem has a threshold solution. If $\chi\left(\Pi^N-\Pi^C\right)>1-2\theta\left(\frac{1}{2},\pi^p\right)$, then he chooses strategy N. If there exists a π^* such that $\chi\left(\Pi^N-\Pi^C\right)=1-2\theta(\pi^*,\pi^p)$, then he chooses N if $\pi>\pi^*$ and C if $\pi<\pi^*$. If $\pi=\pi^*$ then he is indifferent between randomisations over N and C.

In equilibrium the market's conjecture must be correct: $\Psi_c(\pi)=\Psi(\pi)$. Thus, the equilibrium is a threshold equilibrium characterised by a π^* such the expert does not change his forecast if his competency is greater than π^* and he does change it if his competency is less than π^*. We have

$$\Psi\left(\pi\right)\left\{\begin{array}{c}=1\\\in[0,1]\\=0\end{array}\right\}\Leftrightarrow\pi\left\{\begin{array}{c}>\\=\\<\end{array}\right\}\pi^{*}. \tag{7}$$

Substituting equation (7) into equation (1) and the result into equation (3) yields

$$\Pi^{A}=\Pi^{A}\left(\pi^{*},\pi^{p}\right)=\left\{\begin{array}{l}\int_{\pi^*}^{1}pg\left(p,\pi^{p}\right)dp/\int_{\pi^*}^{1}g\left(p,\pi^{p}\right)dp\text{ if }A=N\text{ and }\pi^*<1\\\int_{1/2}^{\pi^*}pg\left(p,\pi^{p}\right)dp/\int_{1/2}^{\pi^*}g\left(p,\pi^{p}\right)dp\text{ if }A=C\text{ and }\pi^*>\frac{1}{2}.\end{array}\right. \tag{8}$$

As seen in equation (1), if a particular action is never chosen in equilibrium, then Bayes' rule cannot be used to form the posterior distribution if such an action were to be observed. Thus, $\Pi^A(\pi^*,\pi^p)$ is not defined in equation (8) if $A=N$ and $\pi^*=1$ or if $A = C$ and $\pi^* = \frac{1}{2}$. That is, if no expert ever sticks with his original forecast, then Bayes' rule cannot be used to specify the market's beliefs, were the market to observe the out-of-equilibrium, or probability zero, phenomenon of an expert persisting with his original forecast and if all experts stick with their original forecast, it cannot be used to specify the market's beliefs if the market were to observe an expert changing his forecast. As *any* beliefs are admissible, we make the following intuitively appealing assumption.

Assumption 1. $\Pi^N(1, \pi^p) = \lim_{\pi^* \to 1} \Pi^N(\pi^*, \pi^p) = 1$ *and* $\Pi^C\left(\frac{1}{2}, \pi^p\right) = \lim_{\pi^* \to 1/2} \Pi^C(\pi^*, \pi^p) = \frac{1}{2}$.

Assumption 1 says that if all types of experts change their forecast and an expert were seen not changing his forecast, then the market would believe that the expert is the most competent type. Likewise, if no type of expert changes his forecast and an expert were seen changing his forecast, then the market would believe that the expert was the least competent type. We discuss the implications of Assumption 1 later in this section.

Using equations (7) and (8), we have the following definition

Definition 1. *An **equilibrium** is a* $\pi^* \in \left[\frac{1}{2}, 1\right]$ *such that*

$$\chi \Pi^N(\pi^*, \pi^p) + \theta(\pi, \pi^p) \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} \chi \Pi^C(\pi^*, \pi^p) + 1 - \theta(\pi, \pi^p) \text{ and } \pi^* \left\{ \begin{array}{c} = \frac{1}{2} \\ \in \left[\frac{1}{2}, 1\right] \\ = 1 \end{array} \right\}. \quad (9)$$

Let $\tilde{\chi} \equiv 6(2\pi^p - 1)(3 - 2\pi^p) / (5 - 4\pi^p)$. Then we have the following result.

Proposition 1. *If* $\chi < \tilde{\chi}$ *then there exists a unique equilibrium* π^* *and it has the property that* $\pi^* < \pi^p$. *Furthermore, if* $\chi \geq \tilde{\chi}$ *then there is a unique equilibrium where no expert changes his forecast.*

Proof. The proofs of all of the Propositions are in the Appendix.

If experts care enough about their reputation, then there is a pooling equilibrium where no expert ever changes his forecast when faced with conflicting public information.[6]

[6] This pooling equilibrium is similar to the one in Cho and Kreps (1987). The senders of the better-quality signals are not able to separate themselves from the senders of the poorer-quality signals because

Otherwise, there is an equilibrium where highly competent experts (those with $\pi \in [\pi^p, 1]$) do not change their mind in the face of conflicting information because their own information is better than the public information. Experts of intermediate competency (those with $\pi \in (\pi^*, \pi^p)$) also do not change their forecast. Their private information is worse than the public information but the greater-than-even probability of predicting the wrong outcome if they do not change their forecast is worth the reputational gain from pooling with more competent experts. Relatively incompetent experts (those with $\pi \in \left[\frac{1}{2}, \pi^*\right)$) change their forecast. Their private information is sufficiently worse than the public information that the reputational gain from masquerading as a more competent expert is not worth the expected cost of an incorrect forecast.

The following intuition is useful in understanding the result. Clearly, an equilibrium cannot have $\pi^* > \pi^p$. If there were such an equilibrium, then any expert with competency $\pi \in (\pi^p, \pi^*)$ would find it preferable – both in terms of making the best forecast and in terms of his reputation – to defect from the equilibrium and not change his forecast. There also can be no equilibrium with $\pi^* = \pi^p$. If there were, all experts with π below, but sufficiently close, to π^p would defect. The expected increased cost of making the incorrect forecast would be negligible compared to the jump in their reputation.

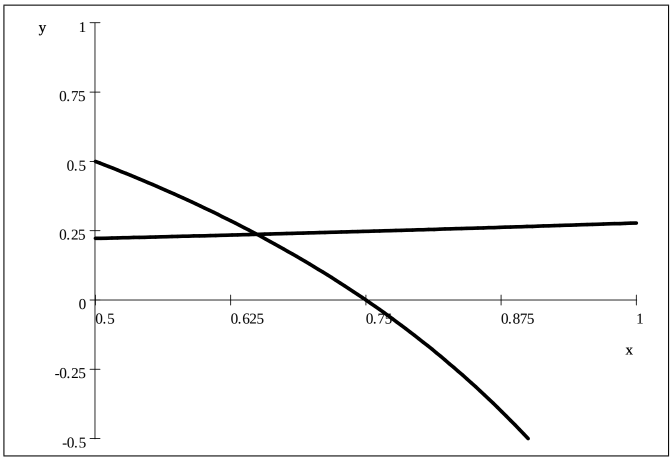

Fig 1. Belief Persistence Equilibrium

the action space is limited. It differs from the pooling equilibrium in Kreps and Wilson (1982) where the senders of the better quality signals not only have limited ability to separate themselves, but – as they are assumed to be mechanistic – also have no incentive to do so.

A formal proof of the proposition is found in the Appendix, but a sketch is as follows. By equation (9), an equilibrium with $\pi^* \in \left(\frac{1}{2}, 1\right)$ satisfies $\chi\left[\Pi^N(\pi^*) - \Pi^C(\pi^*)\right] = 1 - 2\theta(\pi^*, \pi^p)$. The right-hand side of this equation is the expert's expected cost if he does not change his forecast in the face of conflicting information and it equals the likelihood that he is correct if he does not change his forecast minus the likelihood that he is correct if he does change his forecast. It is decreasing in π^*, going to $\frac{1}{2}$ as π^* goes to $\frac{1}{2}$ and to zero as π^* goes to π^p. This is shown in Figure 1, drawn for $\pi^p = .75$. The left-hand side of the equation is the expert's reputational gain if he does not change his forecast. It is strictly positive and is demonstrated in the formal proof to be strictly increasing. The curve representing the left-hand side of the equation shifts up as χ increases and is shown in Figure 1 for $\chi = 1$. From the geometry, it is clear that as long as χ is not too large, the curves representing the left- and right-hand sides of equation (9) cross exactly once at some $\pi^* \in \left(\frac{1}{2}, \pi^p\right)$. If χ is sufficiently large, the curve representing the right-hand side lies above the curve representing the left-hand side on $\left(\frac{1}{2}, 1\right)$ and the equilibrium has $\pi^* = \frac{1}{2}$.

An implication of Proposition 1 is that an increase in the quality of the public signal can increase the size of the set of experts who do *not* change their forecast when faced with conflicting and better quality public information. To see this, suppose that χ is sufficiently large that, given π^p, $\chi > \hat{\chi}$. Then experts pool: no expert changes his mind and the set of experts who continue to forecast an event that they believe is the less likely is $[\pi^*, \pi^p)$, where $\pi^* = \frac{1}{2}$. A marginal increase in π^p has no effect on π^* ($\hat{\chi}$ is continuous in π^p); hence, the set $[\pi^*, \pi^p)$ is enlarged.

Another – and striking – implication is that even when the public information is almost perfect, it is possible for all experts to continue to predict an event that they know is virtually certain *not* to occur. To see this, suppose that $\pi^p \to 1$. If $\pi^* = \frac{1}{2}$, then an expert who changes his forecast is believed to have competency $\frac{1}{2}$ and his revised forecast is correct with probability one. His payoff is thus $\frac{\chi}{2} + 1$. An expert who does not change his forecast is believed to have the average competency, conditional on initially forecasting incorrectly, of $\int_{1/2}^{1} p(1-p)\,dp / \int_{1/2}^{1} (1-p)\,dp = \frac{2}{3}$. His forecast is incorrect with probability one; hence his payoff is $\frac{2\chi}{3}$. As long as $\chi \geq 6$, it is an equilibrium for all experts to continue to forecast an event that will almost certainly not occur.

In this section, attention was restricted to equilibria where out-of-equilibrium beliefs were specified as the limits of equilibrium beliefs. However, other specifications of beliefs can result in other equilibria. In particular, there may be pooling equilibria where all experts change their forecast and an expert who changes his forecast is believed to have the average competency, conditional on initially forecasting incorrectly, of $\frac{2}{3}$. Such an equilibrium might be supported by the out-of-equilibrium belief that an expert who does not change his forecast is the *worst* possible type: his competency is $\frac{1}{2}$. To demonstrate that this is an equilibrium, it is sufficient to demonstrate that no expert would deviate from it. The expert with the most incentive to deviate from it is the most competent expert. Therefore, it is sufficient to show that an expert with $\pi = 1$ would not deviate. Such an expert would receive a payoff of $\frac{2\chi}{3}$ from following the equilibrium strategy and a payoff of $\frac{\chi}{2} + 1$ from deviating. Hence, if $\chi \geq 6$ such an equilibrium exists.

This type of pooling equilibrium is unappealing as the out-of-equilibrium beliefs are not sensible. Why would the market believe that an expert who deviates is the type of expert who has the *least* incentive to deviate? The problem, as previously noted, is that the perfect Bayesian equilibrium concept does not place any restrictions on out-of-equilibrium beliefs, other than that they support the equilibrium. It is typical to rule out pooling equilibria in signalling models that are supported by implausible beliefs by requiring that equilibria satisfy the D1 criterion.[7]

This is not the first paper to demonstrate that reputational or career concerns can distort decision making. The results in this section are related to the literature on anti-herding. In Avery and Chevalier (1999), two experts who care about being perceived as competent and who may have private information about their ability make forecasts in succession. If the second expert has sufficiently precise private information that he is of low competency, he may contradict the first expert's forecast with positive probability,

[7]See Ramey (1996). We discuss the D1 criterion in more detail in the next section.

even though he believes it likely that the first expert is correct. In Levy (2004), experts who care about their reputations for competency and who have private information about their competency make a single forecast after observing public information. Experts of intermediate ability signal their competency by departing from the forecast favoured by the public information, even though their own information supports it. In these models, as in ours, anti-herding is a result of experts wanting to signal their private information.[8] This contrasts with the herding that results in models where agents do not know their own private information.[9]

This section has demonstrated that optimising experts appear to ignore information that conflicts with their beliefs, even though they would make better decisions by considering it. An interesting consequence of this is that more competent experts do not necessarily make better predictions. Although they make better first-round forecasts, by clinging to their original forecast in the face of superior contradictory evidence, experts with $\pi \in (\pi^*, \pi^p)$ make worse second-round forecasts than less competent experts.

3 Overconfidence

> 'We are hitting on all 99 cylinders, so you have to ask yourself, What can we do better? And I just can't decide what that might be... Everyone says that when the markets turn around, we will suffer. But let me tell you, we are going to surprise some people this time around. Bear Stearns is a great place to be.' James E. Cayne,

[8]Some other theoretical models of reputation in which experts know their ability and take actions to signal their types are Trueman (1994) Prendergast and Stole (1996) and Ottaviani and Sorensen (2006).
Ehrbeck and Waldman (1996) consider the empirical predictions of models in which forecasters know their ability and rationally bias their forecasts for reputational reasons. These models have some similarity to our set-up in that the forecasters are required to make a sequence of forecasts. However, in contrast to our paper in which the expert observes a public signal after a first-round action, forecasters in Ehrbeck and Waldman, only observe a private signal before each round of forecasting. The empirical findings of Ehrbeck and Waldman (1996) are rather negative for their reputational model. Since the low-ability types in their models mimic high-ability types' behaviour for some parameter values, the predicted forecast bias is in the direction of forecasts typical of high-ability types; this prediction is empirically rejected by their data on T-bill forecasters where the bias is in the direction of those with large mean-squared forecast errors. They argue that their empirical observation is consistent with behavioural explanations for forecast bias.

[9]See, for example, Dewatripont, Jewitt and Tirole (1999), Hirshleifer and Teoh (2003) and Chamley (2004).

Chairman and CEO of Bear Stearns, 2003.[10]

A popular explanation for why the bankers made so many disastrous decisions in the run up to the financial crisis is that they were overconfident.[11] A sizable literature demonstrates that overconfidence can damage financial markets. Papers by Odean (1998), Barber and Odean (2001) and Biais et al (2005) demonstrate that it leads to excess trading and lower profits. Papers by Daniel et al (1998), Scheinkman and Xiong (2003) and Burnside at al (2011) show that it can lead to a asset-price anomalies, such as overreactions, excess volatility and bubbles.[12] In this section we present a model where rational bankers exhibit overconfidence.

Here we suppose that there is no publicly observed information, as there was in the previous section, but that after observing his signal the expert announces the likelihood, ρ, that his forecast is correct. In terms of our banking story, ρ can be thought of as the vigour with which a banker attempts to sell his forecast to his employer and his clients. We assume that there is no direct benefit to an expert of announcing that he is correct with high probability and that there is a cost: if the expert turns out to be incorrect, he later suffers a loss that has a discounted present value of $c(\rho)$, where $c : \left[\frac{1}{2}, 1\right] \rightarrow \mathbb{R}_+$ is strictly increasing, concave, twice differentiable and has $c\left(\frac{1}{2}\right) = 0$.[13] We assume that the weight put on reputation is sufficiently high: $\chi \geq c'(1/2)$.

The payoff to the expert if he announces that he correct with probability ρ is then

$$\chi \Pi + P - c(\rho), \tag{10}$$

where χ, Π and P are defined as in the previous section.

We consider perfect Bayesian equilibria. As in the previous section, we make an assumption that rules out pooling equilibria based on implausible out-of-equilibrium beliefs.

[10] Quoted in Thomas (2003).

[11] See, for example, Kohn (2008) and Gladwell (2009).

[12] See Glaser (2004) for a survey.

[13] Concavity will turn out to be sufficient, but not necessary, for the second-order condition of the expert's problem to be satisfied.

Assumption 2. *Equilibria must satisfy the D1 criterion.*[14]

Intuitively, imposing the D1 criterion implies that following an observation of an out–of-equilibrium announcement, the public must put zero posterior weight on the expert being type π if there is another expert of type π' who has a greater incentive to deviate from the equilibrium, in the sense that type π' would strictly prefer to deviate for any resulting market belief Π that would make π weakly prefer deviating to not deviating.

If the D1 criterion holds, then the equilibrium must be separating, except possibly for an interval of the most competent experts who claim to be right with probability one.

Proposition 2. *The equilibrium must have the following form: there is a* $\pi^* \in \left[\frac{1}{2}, 1\right]$ *such that experts in* $(\pi^*, 1]$ *say that they are certain that they are right. Experts with* $\pi \in \left[\frac{1}{2}, \pi^*\right)$ *separate: they each announce that they are correct with some probability in* $\left[\frac{1}{2}, 1\right)$ *and their announcement reveals their type.*

Note that the proposition does not rule out $\pi^* = \frac{1}{2}$ or $\pi^* = 1$. The formal proof, in the Appendix, borrows from Ramey (1996). The strategy is to demonstrate that if the D1 criterion holds and if any two experts of different types pool at any announcement other than $\rho = 1$, then the more competent expert has an incentive to deviate. Pooling with $\rho = 1$ is not ruled out by the D1 criterion, but there is no equilibrium where a less competent expert chooses $\rho = 1$ and a more competent expert chooses $\rho < 1$. To see this, suppose that the market believes that an expert who chooses $\rho = 1$ is more competent than an expert who chooses $\rho < 1$. Then if the less competent expert is willing to choose $\rho = 1$ to be thought more competent, the more competent expert must also be willing.

The market conjectures that a policy maker of type $\pi \in \left[\frac{1}{2}, \pi^*\right)$ announces that he is correct with probability $\rho_c(\pi) < 1$. Separability implies that $\rho_c : \left[\frac{1}{2}, \pi^*\right] \to \left[\frac{1}{2}, 1\right)$ is one

[14] Let $\{\rho(\pi), \Pi(\rho), \pi\}$ be a perfect Bayesian equilibrium. Let $\hat{\rho}$ be an out-of-equilibrium action and suppose that $\hat{\Pi} \in \left[\frac{1}{2}, 1\right]$ is the market's assessment of the expert's type if it observes such an action. Suppose that there is a non-empty set of expert types $S' \subset \left[\frac{1}{2}, 1\right]$ such that for every expert of type $\pi \notin S'$ who weakly prefers following the out-of-equilibrium strategy $\hat{\rho}$ and being thought to be type $\hat{\Pi}$ to following his equilibrium strategy $\rho(\pi)$ and being thought to be type $\Pi(\rho(\pi))$ there exists an expert of type $\pi' \in S'$ who strictly prefers following the out-of-equilibrium strategy $\hat{\rho}$ and being thought to be type $\hat{\Pi}$ to following his equilibrium strategy $\rho(\pi')$ and being thought to be type $\Pi(\rho(\pi'))$. Then the equilibrium violates the D1 criterion unless, upon observing $\hat{\rho}$, the market infers that the expert's type $\pi \in S'$.

to one. Hence, upon observing $\rho < 1$, the market infers that the expert is type $\rho_c^{-1}(\rho)$. Thus

$$\Pi = \begin{cases} \frac{\pi^*+1}{2} \text{ if } \rho = 1 \\ \rho_c^{-1}(\rho) \text{ otherwise.} \end{cases} \tag{11}$$

In equilibrium, the market's conjecture must be correct and $\rho_c(\pi^*) = \rho(\pi^*)$. Suppose that there is an interior threshold $\pi^* \in \left(\frac{1}{2}, 1\right)$. Then the threshold expert, the one with competency $\pi = \pi^*$, must be indifferent between announcing that he is correct with probability one and announcing that he is correct with probability $\rho(\pi^*)$. If he claims to be correct with probability one, then by equation (11), the market's assessment of his competency is $\frac{\pi^*+1}{2}$. If his forecast turns out to be incorrect, then he suffers a loss of $c(1)$. Thus, by equation (10), his expected payoff is $\chi\frac{\pi^*+1}{2} + \pi^* - (1-\pi^*)\, c(1)$. If, instead, he claims to be correct with probability $\rho(\pi^*)$, then he is thought to have competency π^*. If his forecast turns out to be incorrect, then he incurs a loss of $c(\rho(\pi^*))$. Thus, by equation (10), his expected payoff is $\chi\pi^* + \pi^* - (1-\pi^*)\, c(\rho(\pi^*))$. Equating the expected payoff from claiming to be correct with probability one to the expected payoff from claiming to be correct with probability $\rho(\pi^*)$ yields

$$\pi^* = \begin{cases} \rho^{-1}\left(c^{-1}\left(c(1) - \frac{\chi}{2}\right)\right) \in \left(\frac{1}{2}, 1\right) \text{ if } \chi < 2c(1) \\ \frac{1}{2} \text{ otherwise.} \end{cases} \tag{12}$$

If $\pi^* \in \left(\frac{1}{2}, 1\right]$ and $\pi < \pi^*$, then by equations (10) and (12), the expert maximises

$$\chi\rho_c^{-1}(\rho) - (1-\pi)\, c(\rho). \tag{13}$$

We conjecture that $\rho_c(\pi)$ is twice differentiable and it will later be clear that this is the case.[15] The first- and second-order conditions for a solution to the expert's problem are:

$$\chi\rho_c^{-1\prime}(\rho) - (1-\pi)\,c'(\rho) = 0 \quad (14)$$

$$\chi\rho_c^{-1\prime\prime}(\rho) - (1-\pi)\,c''(\rho) = 0 \quad (15)$$

Using the rules $f^{-1\prime}(x) = 1/f'(x)$ and $f^{-1\prime\prime}(x) = -f''(x)/f'(x)^3$ and imposing $\rho_c(\pi) = \rho(\pi)$, equations (14) and (15) yield

$$\frac{\chi}{(1-\pi)\,c'(\rho(\pi))} = \rho'(\pi) \quad (16)$$

$$-\frac{\chi\rho''(\pi)}{\rho'(\pi)^2} - (1-\pi)\,c''(\rho(\pi)) < 0. \quad (17)$$

Equation (16) is a first-order differential equation with no boundary condition. Following Riley (1979), it is conventional in signalling models to generate a boundary condition by assuming that the agent with the lowest-quality private information (here, an expert with $\pi = \frac{1}{2}$) would not send a costly signal. The logic is that, in a separating equilibrium, expectations can be no worse; hence there is no point to costly signalling.[16] Thus

$$\rho(1/2) = 1/2. \quad (18)$$

Definition 2. *An equilibrium is a* $\pi^* \in \left[\frac{1}{2}, 1\right]$ *and a twice-differentiable function* $\rho(\pi) : \left[\frac{1}{2}, \pi^*\right] \to \left[\frac{1}{2}, 1\right]$ *such that equations (12) and (16) - (18) are satisfied.*

An equilibrium is a perfect Bayesian equilibrium. The expert is maximising his payoff while taking into account the effect of his action on the beliefs of the market. The market's beliefs are (trivially) consistent with Bayes rule and are formed using the correct conjecture about the equilibrium strategies and the observation of ρ.

Proposition 3. *If* $\chi \geq 2c(1)$ *then all experts claim that they are correct with probability one. Furthermore, if* $\chi < 2c(1)$ *then all experts with competencies in* $[\pi^*, 1]$ *claim that*

[15] There are no separating equililbria that are not differentiable. See Mailath (1987).

[16] This is also the only equilibrium that satisfies the D1 criterion.

they are the correct with probability one and experts with $\pi \in \left[\frac{1}{2}, \pi^*\right)$ *claim that they are correct with probability* $\rho(\pi)$, *where*

$$\rho(\pi) = c^{-1}\left(-\chi \ln\left(2\left(1-\pi\right)\right)\right) \text{ and } \rho(\pi) > \pi, \ \pi \in \left[\frac{1}{2}, \pi^*\right) \tag{19}$$

$$\pi^* = 1 - \frac{1}{2}\exp\left(\frac{1}{2} - \frac{c(1)}{\chi}\right) \in \left(\frac{1}{2}, 1\right). \tag{20}$$

Proposition 3 demonstrates that experts with competencies in the separating region $\left[\frac{1}{2}, \pi^*\right)$ are overconfident, as well as experts in the pooling region $[\pi^*, 1]$. It also ensures that $\pi^* < 1$ and equilibria with complete separation do not exist. The intuition is that, in the separating region, experts with good quality private information separate themselves from senders of poorer quality information by saying that they are more confident than they actually are. However, there is an upper bound on how overconfident an expert can be: ρ can be no greater than one. Thus, experts with very good quality information are unable to separate themselves.[17]

The overconfidence in our model occurs when people are rewarded based on their perceived abilities; it does not exist if people are rewarded solely on their performance (that is, when $\chi = 0$). There is some empirical evidence that is consistent our result. While overconfidence is widespread, a few types of experts appear to exhibit little or no overconfidence. Examples are bridge players, oddsmakers and weather forecasters.[18] For all of these people, the success or failure of their conjectures is immediately and publicly observable. Hence, it is likely that they perceive their reward to be based on their performance rather than their perceived competency. In their study of financial services professionals, Gloede and Menkhoff (2009) found that fund managers were the least overconfident employees. They note that they were also the employees with the most direct feedback and whose salary was most closely linked to performance.

[17] Cho and Sobel (1990) consider a general game where the sender of the signal's action space is bounded above and find that a possible outcome is a set of types pooling at the highest possible action.

[18] See Plous (1993) for a survey of this literature.

We are not the first to explain overconfidence in an optimising model; alternative frameworks are offered by Van den Steen (2004) and Brocas and Carrillo (2002). Van den Steen (2004) explains overconfidence by supposing that individuals have noisy idiosyncratic information. Thus, agents who select an option are more likely to be optimistic about their choice than other agents. Brocas and Carrillo (2002) suppose that agents choose between a riskless activity and an activity that can yield either a high or a low payoff, depending upon their competency. If agents are uncertain about their abilities and information acquisition is costly they choose the risky activity if preliminary evidence about their competency is positive; they do not choose the safe activity without substantial information that they are incompetent. Thus, it is more likely that incompetent people will engage in the risky activity than it is that competent people will engage in the safe activity.

It appears to be widely believed that the overconfidence of bankers played a key role in the financial crisis. The model we present here demonstrates that bankers may rationally display overconfidence, even if they are not actually overconfident. An obvious alternative explanation is that bankers are genuinely overconfident. Overconfidence is pervasive. Most of us display it our own lives, in our certainty, for example, that we are better drivers than average.[19] A vast social psychology literature documents its existence. It is possible that the bankers' overconfidence was real. However, the empirical evidence on this is mixed. Glaser et al's (2010) study of 100 professional traders and investment bankers finds some evidence of overconfidence. Gloede and Menkhoff (2009), however, asked 105 forecasters who contributed to the ZEW, Manheim financial market survey to rank their own ability compared to the average ability on a scale of 1 - 21; the mean answer of 11.9 was only slightly above the average of 11.0.

[19] This is shown in numerous studies. Svenson (1981), for example, found that eighty percent of survey respondents claimed to be in the top thirty percent of all drivers.

4 Disconfirming Evidence

> 'But how exactly does a bank, such as UBS, conjure up losses larger than the gross domestic product of many countries... ?' Gillian Tett, [20]

Since the beginning of the financial crisis, the Swiss bank UBS has been forced to write down about $50 billion of mortgage-related assets and, as of end-2007, 'super senior' tranches contributed to about half of this loss.[21] The super senior tranche is the most protected tranche of a collateralised debt obligation (CDO), receiving payment before all other tranches, but if the issuer of the CDO borrowed to purchase assets for the CDO there is a potential risk. UBS's investment banking unit did consider the riskiness of the super senior tranches when it did its value at risk analysis, but it used the AAA ratings assigned to these tranches by Moody's and S & P's. It appears that the unit made no attempt whatsoever to investigate the fundamentals of the US housing market.[22] In this section we present a model where bankers rationally fail to look for evidence that would disconfirm that their theories.

In this section we consider a scenario where an expert must predict which one of three or more events will occur. He receives a signal that tells him that one of the events will occur with probability $\pi \in \left[\frac{1}{2}, 1\right]$. As in the previous sections, this likelihood that his signal is correct is referred to as his *competency* and it is his private information. After receiving the signal the expert forecasts the event that his signal favours. He then has the opportunity to invest in the possibility of finding disconfirming information. Specifically, if the expert pays a cost q, then if his signal was incorrect, with probability π_d he receives private information that (with probability one) will confirm this. The expert can then continue with his original prediction or he can withdraw his forecast. Later, the event is observed and it is learned if the expert was correct or not. Because there are more than two possible events, disconfirming evidence does not resolve the uncertainty. Upon

[20]Tett (2008).
[21]See Salmon (2008).
[22]See Shefrin (2009).

receiving proof that his original forecast was wrong, the expert no longer has information that is useful to the market. Thus, changing his forecast is not an option in this scenario.

The expert's payoff is

$$\chi\Pi + P^d - \delta q, \tag{21}$$

where the variable P^d equals one if he persists with his original forecast and it is correct, zero if he withdraws his original forecast and minus one if he persists in his original forecast and it turns out to be incorrect. The variable δ equals one if he invests in additional information and zero otherwise. The parameter χ and the variable Π are as defined in the previous sections.

As a benchmark, we first consider the case where $\chi = 0$. Suppose that an expert invests in the possibility of finding disconfirming evidence. With probability π, his original forecast is correct and he finds no disconfirming information. Thus, he continues with his original choice, which he knows to be correct with probability greater than one half, and is later proved to be correct. Hence, $P^d = 1$. With probability $(1-\pi)\,\pi_d$, his original forecast is wrong and he receives confirmation of this. He withdraws his original forecast and $P^d = 0$. With probability $(1-\pi)(1-\pi_d)$ his original forecast is wrong, but he does not receive disconfirming information. He continues with his original choice, which he believes is correct with probability greater than one half, and $P^d = -1$. Thus, the expected value of P^d is $\pi - (1-\pi)(1-\pi_d)$ and the expert's payoff when he invests in the possibility of finding disconfirming evidence is $\pi - (1-\pi)(1-\pi_d) - q$.

If the expert does not invest in the possibility of finding disconfirming evidence, then he continues with his original forecast and his payoff is equal to the expected value of P^d, which is equal to the probability his choice is correct minus the probability it is not, or $\pi - (1-\pi)$. The expert will choose to invest in the possibility of receiving disconfirming information if the expected payoff from doing so exceeds the expected payoff from not doing so. This is the case when $1 - q/\pi_d > \pi$. Thus, if there are no reputational considerations, it is the *less* competent experts who invest in acquiring disconfirming information; relatively competent experts do not. This is because, as their own signal is more likely to be correct, relatively competent experts find that a search for evidence proving otherwise is less likely to be informative.[23] We assume that the cost of acquiring information is sufficiently low that, in the absence of reputational concerns, some experts would acquire it: $\pi_d > 2q$.

We now suppose that reputational concerns matter, that is $\chi > 0$, and we initially suppose that the investment in information is observable, although the result is not. We look for a threshold equilibrium where experts with $\pi \in \left[\frac{1}{2}, \pi^*\right)$ invest in information acquisition and experts with $\pi \in [\pi^*, 1]$ do not.

Let Π^D be the expert's reputation if he does not invest in additional information, Π^N be his reputation if he does invest and does not withdraw his original forecast and Π^W be his reputation if he invests and then withdraws his forecast. We consider equilibria where $\Pi^D > \Pi^N > \Pi^W$.[24] In such equilibria an expert who invests in additional information and does not receive disconfirming evidence continues with his original forecast: this is best both in terms of maximising his payoff from making the best forecast and enhancing his reputation. An expert who invests in additional information and receives disconfirming evidence withdraws his forecast. This is because an equilibrium strategy of investing in additional information and persisting with his original forecast with strictly positive probability in the face of disconfirming evidence must have the same payoff as the strategy of investing in additional information and always persisting with his original forecast. This latter strategy is dominated by the strategy of not investing in additional information.

The payoff to the strategy of not investing in additional information is $\chi\Pi^D + \pi - (1 - \pi)$. The payoff to the strategy of investing in additional information and withdrawing one's forecast if and only if disconfirming evidence is found is $[1 - \pi(1 - \pi_d)]\,\Pi^N +$

[23] The argument that, even without reputational concerns, competent experts are unlikely to look for disconfirming evidence because they are unlikely to find it is related to Oaksford and Chater's (1994) argument that a failure to focus solely on evidence that might disprove a hyphothesis may be a result of the properties that figure in a causal relationship being rare.

[24] As in the previous sections, there may be equilibria supported by unappealing out-of-equilibrium beliefs that do not satisfy this condition.

$(1-\pi)\,\pi_d\Pi^W + \pi - (1-\pi)(1-\pi_d) - q$. By the same reasoning as in Section 2, the expert follows a threshold strategy. There is a π^* such that the expert invests in additional information if $\pi < \pi^*$, does not invest if $\pi > \pi^*$ and is indifferent over randomisations if $\pi = \pi^*$. Thus, in equilibrium the market conjectures that the expert follows such a threshold strategy. Thus, using Bayes' rule (as in equation (1)), we have

$$\Pi^A = \begin{cases} \Pi^D(\pi^*) = \frac{1+\pi^*}{2} \text{ if } A = D \text{ and } \pi^* < 1 \\ \Pi^N(\pi^*) = \frac{\int_{1/2}^{\pi^*} p[1-\pi_d(1-p)]dp}{\int_{1/2}^{\pi^*}[1-\pi_d(1-p)]dp} \text{ if } A = N \text{ and } \pi^* > \frac{1}{2} \\ \Pi^W(\pi^*) = \frac{\int_{1/2}^{\pi^*} p(1-p)dp}{\int_{1/2}^{\pi^*}(1-p)dp} \text{ if } A = W \text{ and } \pi^* > \frac{1}{2}. \end{cases} \tag{22}$$

We specify out-of-equilibrium beliefs as in Assumption 1.

Assumption 3. $\Pi^D(1) = \lim_{\pi^*\to 1}\Pi^D(\pi^*) = 1$ *and* $\Pi^A\left(\frac{1}{2}\right) = \lim_{\pi^*\to 1/2}\Pi^A(\pi^*) = \frac{1}{2}$, $A = N, W$.

We have the following definition.

Definition 3. *An* ***equilibrium*** *is a* π^* *such that*

$$\pi^* \left\{ \begin{array}{c} = \frac{1}{2} \\ \in [0,1] \\ = 1 \end{array} \right\} \text{ and } \chi\Pi^D(\pi^*) + q \left\{ \begin{array}{c} \geq \\ = \\ \leq \end{array} \right\}$$
$$[1-(1-\pi^*)\,\pi_d]\,\Pi^N(\pi^*) + (1-\pi^*)\,\pi_d\Pi^W(\pi^*) + (1-\pi^*)\,\pi_d. \tag{23}$$

Proposition 4. *If $\chi \geq 2(\pi_d - 2q)$ then there is a unique equilibrium where no expert invests in the possibility of finding disconfirming evidence. Furthermore, if $\chi < 2(\pi_d - 2q)$ then there is a unique $\pi^* < \tilde{\pi}$ such that experts with competency $\pi < \pi^*$ invest in the possibility of finding disconfirming evidence and experts with competency $\pi > \pi^*$ do not.*

Thus, we have that when the search for disconfirming evidence is observable, fewer experts will invest in the possibility of finding disconfirming evidence than they would if they did not have reputational concerns. If the reputational concerns are important enough, no expert will invest in the possibility of finding disconfirming evidence.

We now consider the case where an investment in the possibility of finding disconfirming evidence is unobservable. In this case, the market observes only whether the expert continues to maintain his original forecast (action N) or withdraws it (action W). If an expert does not withdraw his forecast then the market believes that either the expert did not invest in the possibility of finding disconfirming information and, hence, $\pi \in [\pi^*, 1]$ or that the expert did invest, and hence $\pi \in \left[\frac{1}{2}, \pi^*\right]$, but no disconfirming evidence was received. Thus, if an expert does not withdraw his original forecast, the market's assessment of his competency is

$$\Pi^A = \begin{cases} \Pi^N(\pi^*) = \frac{\int_{1/2}^1 p dp - \pi_d \int_{1/2}^{\pi^*} p(1-p) dp}{\int_{1/2}^1 dp - \pi_d \int_{1/2}^{\pi^*} (1-p) dp} \text{ if } A = N \\ \Pi^W(\pi^*) = \frac{\int_{1/2}^{\pi^*} p(1-p) dp}{\int_{1/2}^{\pi^*} (1-p) dp} \text{ if } A = W \text{ and } \pi^* > \frac{1}{2}. \end{cases} \tag{24}$$

As before, we have:

Assumption 4. $\Pi^W\left(\frac{1}{2}\right) = \lim_{\pi^* \to 1/2} \Pi^W(\pi^*) = \frac{1}{2}$.

If an expert invests in the possibility of finding disconfirming information, then his expected payoff is $\chi\left\{[\pi + (1-\pi)(1-\pi_d)] \Pi^N(\pi^*) + (1-\pi)\pi_d \Pi^W(\pi^*)\right\} + \pi - (1-\pi)(1-\pi_d)$ - q. If he does not invest in the possibility of finding disconfirming information, then his expected payoff is $\chi \Pi^N(\pi^*) + \pi - (1-\pi)$. The threshold expert is indifferent; hence, we have the following.

Definition 4. *An **equilibrium** is a π^* such that*

$$\pi^* \left\{ \begin{array}{c} = \frac{1}{2} \\ \in [0,1] \\ = 1 \end{array} \right\} \text{ and } (1-\pi^*)\pi_d - q \left\{ \begin{array}{c} \geq \\ = \\ \leq \end{array} \right\} \chi (1-\pi^*) \pi_d \left[\Pi^N(\pi^*) - \Pi^W(\pi^*)\right]. \tag{25}$$

Proposition 5. *If $\chi \geq 4(\pi_d - 2q)/\pi_d$ then there is an equilibrium where no expert invests in the possibility of finding disconfirming evidence. Furthermore, if $\chi < 4(\pi_d - 2q)/\pi_d$ then there is a $\pi^* < \tilde{\pi}$ such that experts with competency $\pi \leq \pi^*$ invest in the possibility of finding disconfirming evidence and experts with competency $\pi \leq \pi^*$*

do not.

The model of this section provides a rationale for the seemingly irrational refusal of bankers to consider disconfirming evidence. It does not, however, provide a full explanation for classical confirmation bias, which includes both a tendency to search too little for disconfirming evidence and a tendency to weight confirming evidence too heavily. In addition, while we have provided an explanation for not considering evidence that would disconfirm one's hypotheses that is consistent with rational optimising behaviour, there are undoubtedly other explanations as well. Westen at al (2006) provide a physiological one. They used neuroimaging to study the brains of party loyalists during the 2004 US Presidential election. Subjects were confronted with reasoning tasks involving information damaging to their candidate, the other candidate or some neutral control target. They found that when subjects had an emotional stake, there was neural activity in different parts of the brain than when they did not. This supports a belief that the brain seeks solutions that satisfy emotional, as well as cognitive, constraints.

In this and the previous two sections, the desire of experts to be seen as competent distorts their behaviour. In typical signalling models the harm caused by suboptimal behaviour may be mitigated be the increased information available to the market about the senders of the signals' types. Here, however, the signalling does not necessarily convey more information about the expert's competency. In the belief persistence model of Section 2 and in the model of this section, in the absence of reputational concerns, the experts would split into two pools, one consisting of more competent experts and one consisting of less competent experts. With relatively weak reputational concerns, the experts still split into two pools, although there would be more experts in the relatively competent pool and fewer in the less competent pool. With strong reputational concerns, however, there is complete pooling and the market has *less* information than it would have without reputational concerns. In the overconfidence model of Section 3, in the absence of reputational concerns the experts would pool With strong reputational concerns there would also be pooling: everyone would exhibit the same perfect confidence. However, with weak reputational concerns there is some separation and the market does gain some information relative to what it would have had with no reputational concerns.

5 Conclusion

In this paper, we demonstrate that a desire to appear competent may explain what appear to be the cognitive errors that are known as confirmation bias. Our motivation is an explanation of the actions of bankers. We chose bankers as the subject of our research because their apparent cognitive errors in the run up to the recent financial crisis have been so widely documented and because of the important policy implications of their behaviour. However, our framework can explain what appear to be cognitive errors in other contexts where people are rewarded in the short run for their perceived ability, as well as for their performance in the long run.

The research of Bogan and Just (2009) suggests another economic application. They note that the academic literature has demonstrated that most mergers either add no shareholder value for the acquiring firm or reduce it and they argue that once executives have identified a potential merger, they avoid seeking out information that would disconfirm the benefits of the merger. Tuchman (1984) suggests a political application. She describes several examples of governments continuing to justify policies that they had committed themselves to long after it became apparent that their cause was lost. Nickerson (1998) claims that history is full of scientists maintaining their beliefs, long after the available evidence had suggested they should be abandoned and he suggests that many theories could have been easily invalidated if scientists had made a serious effort to show that they were false. Supporters of Ptolemaic cosmology clung to their beliefs despite mounting evidence against it. The theory that heavier bodies fall faster than lighter ones could have been disproved long before Galileo. In all three of these cases, a desire to be seen as competent in the short run as well as ultimately right could make these apparent errors the result of rational behaviour.

Appendix

Proof of Proposition 1. Define $L(\pi^*, \pi^p) \equiv \chi\left[\Pi^N(\pi^*, \pi^p) - \Pi^C(\pi^*, \pi^p)\right]$ and $R(\pi^*, \pi^p) \equiv 1 - 2\theta(\pi^*, \pi^p)$. The proposition follows from the following properties of L and R: (i) $\partial L/\partial\pi^* > 0$; (ii) $L\left(\frac{1}{2}, \pi^p\right) = \frac{\chi(5-4\pi^p)}{6(3-2\pi^p)} > 0$; (iii) $\partial R/\partial\pi^* < 0$; (iv) $R\left(\frac{1}{2}, \pi^p\right) = 2\pi^p - 1$; (v) $R(\pi^p, \pi^p) = 0$. Properties (ii) - (v) are straightforward. We show property (i). By equations (2) and (8)

$$\Pi^A(\pi^*, \pi^p) = \begin{cases} \frac{2\pi^*}{3} + \frac{1}{3(2\pi^p-1)}\left[\pi^p - \frac{2(1-\pi^p)^2}{1-(2\pi^p-1)\pi^*}\right] \text{ if } A = N \\ \frac{2\pi^*}{3} + \frac{1}{3(2\pi^p-1)}\left[\pi^p - \frac{\frac{1}{2}}{\frac{1}{2}+\pi^p-(2\pi^p-1)\pi^*}\right] \text{ if } A = C. \end{cases} \tag{A1}$$

By equation (A1), $\partial L/\partial\pi^* > 0$ iff $1/\left[\frac{1}{2} + \pi^p - (2\pi^p - 1)\pi^*\right]^2 > 4(1-\pi^p)^2/[1$ - $(2\pi^p$ - $1)\pi^*]^2$. This is true iff $[1 - (2\pi^p - 1)\pi^*] > 2(1-\pi^p)\left[\frac{1}{2} + \pi^p - (2\pi^p - 1)\pi^*\right]$. This follows because both sides are linear in π^p and it is true at the endpoints $\pi^p = \frac{1}{2}$ and $\pi^p = 1$.

Proof of Proposition 2. Let $\rho(\pi)$ be the equilibrium strategy of a type-π expert and let $\Pi(\rho)$ be the market's equilibrium assessment of π given an observation of ρ. We first demonstrate that there cannot be pooling at any $\rho < 1$. Suppose to the contrary that there exists a pooling equilibrium with $\rho(\pi) = \rho'$ for more than one $\pi \in \left[\frac{1}{2}, 1\right]$. Let $\pi' = \sup\{\pi | \rho(\pi) = \rho'\}$. Clearly $\Pi(\rho') < \pi'$. Choose $\pi'' \in (\Pi(\rho'), \pi')$.

By equation (10), if $\rho = \rho'$, then a marginal increase in ρ accompanied by an increase in Π raises the welfare of a type-π expert if and only if $d\Pi > \frac{1}{\chi}(1-\pi)c'(\rho')d\rho'$. Thus, since $\pi'' < \pi'$, for $\tilde{\rho}$ strictly greater than, but sufficiently close to, ρ' it is possible to find a $\tilde{\Pi} > \Pi(\rho')$ such that the type-π' expert strictly prefers choosing $\tilde{\rho}$ and being thought type $\tilde{\Pi}$ to choosing ρ' and being thought type $\Pi(\rho')$ and a type-π expert strictly prefers choosing ρ' and being thought type $\Pi(\rho')$ to choosing $\tilde{\rho}$ and being thought type $\tilde{\Pi}$, for every $\pi \leq \pi''$. Choose $\tilde{\rho}$ sufficiently close to ρ' that $\tilde{\Pi} < \pi''$.

Case 1. Suppose that $\tilde{\rho}$ is an out-of-equilibrium action. If type $\pi \leq \pi''$ weakly prefers prefers choosing $\tilde{\rho}$ and being thought type Π to choosing ρ' and being thought type $\Pi(\rho')$ then it must be that $\Pi > \tilde{\Pi}$. We also have that type π' strictly prefers choosing $\tilde{\rho}$ and being thought type Π to choosing ρ' and being thought type $\Pi(\rho')$. Thus, for the D1 criterion to hold it must be that if the public observes the out-of-equilibrium action $\tilde{\rho}$ it believes that $\pi > \pi''$. Thus, $\Pi(\tilde{\rho}) > \pi''$. Since type π' strictly prefers choosing $\tilde{\rho}$ and being thought type $\tilde{\Pi}$ to choosing ρ' and being thought type $\Pi(\rho')$, he strictly prefers choosing $\tilde{\rho}$ and being thought type $\Pi(\tilde{\rho}) > \tilde{\Pi}$ to choosing ρ' and being thought type $\Pi(\rho')$. Thus, type π' defects. This is a contradiction.

Case 2. Suppose that $\tilde{\rho}$ is an equilibrium action. We have that the type-π' expert strictly prefers choosing $\tilde{\rho}$ and being thought type $\tilde{\Pi}$ to choosing ρ' and being thought type $\Pi(\rho')$; hence, $\tilde{\Pi} > \Pi(\tilde{\rho})$. We have that a type-$\pi$ expert strictly prefers choosing ρ' and being thought type $\Pi(\rho')$ to choosing $\tilde{\rho}$ and being thought type $\tilde{\Pi}$, for every $\pi \leq \pi''$. Thus, no expert of type-π, $\pi \leq \pi''$, chooses $\tilde{\rho}$. This implies that $\Pi(\tilde{\rho}) \geq \pi'' > \tilde{\Pi}$: a contradiction.

This concludes the first part of the proof. We now show that if an equilibrium has a type-π' expert choosing $\rho = 1$ then each type-π expert, $\pi > \pi'$ also chooses $\rho = 1$. Suppose to the contrary that an equilibrium has a type-π' expert choosing $\rho = 1$ and a type-π, $\pi > \pi'$, expert choosing $\tilde{\rho} < 1$. By equation (10), for the experts not to defect we require $1 - \pi' \leq \chi [\Pi(1) - \Pi(\tilde{\rho})] / [c(1) - c(\tilde{\rho})] \leq 1 - \pi$: a contradiction.

Proof of Proposition 3. The differential equation (16) is separable and has solutions $\rho(\pi) = c^{-1}(k - \chi \ln(1-\pi))$, where k is a constant. Imposing the boundary condition (18) yields equation (19). Solving equations (12) and (19) yields equation (20). Differentiating equation (16) yields

$$-\frac{\rho''(\pi)}{\rho'(\pi)^2} = (1-\pi)\, c''(\rho(\pi))\, \rho'(\pi) - c'(\rho(\pi)). \tag{A2}$$

The concavity of c and $\chi \geq c'(1/2)$ ensures that $\rho'(\pi) > 1$. This and equation (A2) ensure that condition (17) is satisfied. Equation (19) and $\rho'(\pi) > 1$ imply $\rho(\pi) > \pi$ for $\pi \in \left(\frac{1}{2}, \pi^*\right]$.

Proof of Proposition 4. By Assumption 3 and equations (22) and (23), $\pi^* = \frac{1}{2}$ iff $\chi \geq 2(\pi_d - 2q)$ and $\pi^* \neq 1$. Suppose $\chi < 2(\pi_d - 2q)$ and let $R_0(\pi^*) \equiv \Pi^D(\pi^*) - \Pi^N(\pi^*)$

and $R_1(\pi^*) \equiv \pi_d(1-\pi^*)\left[\Pi^N(\pi^*) - \Pi^W(\pi^*)\right]$. Then by equation (22)

$$
\begin{aligned}
R_0(\pi^*) &= \frac{3-\pi^*}{6} - \frac{1}{3}\frac{\frac{3}{2}-\pi^*-\pi_d(1-\pi^*)}{2-\pi_d\left(\frac{3}{2}-\pi^*\right)} && \text{(A3)}\\
R_1(\pi^*) &= \frac{\pi_d(1-\pi^*)\left(\pi^*-\frac{1}{2}\right)^2}{3\left(\frac{3}{2}-\pi^*\right)\left[2-\pi_d\left(\frac{3}{2}-\pi^*\right)\right]}. && \text{(A4)}
\end{aligned}
$$

By equation (23), an interior threshold requires

$$(1-\pi^*)\pi_d - q = \chi\left[R_0(\pi^*) + R_1(\pi^*)\right]. \tag{A5}$$

The left-hand side of equation (A5) is strictly decreasing, equaling $\frac{\pi_d}{2}-q$ when $\pi^*=\frac{1}{2}$ and equaling zero as $\pi^*=\tilde{\pi}$. The right-hand side of equation (A5) is strictly positive, equalling $\frac{\chi}{4}<\frac{\pi_d}{2}-q$ when $\pi^*=\frac{1}{2}$ and equalling $\frac{\chi}{3}\frac{3-\pi_d}{4-\pi_d}>0$ when $\pi^*=1$. Thus, $\exists\pi^*\in\left[\frac{1}{2},\tilde{\pi}\right]$ such that equation (A5) holds. This π^* is unique if the right-hand side is increasing, or if it is decreasing and less steep than the left-hand side. This is the case if $\pi_d>\chi\left[R_0'(\pi^*)+R_1'(\pi^*)\right]$. We have $\frac{\chi}{2}<\pi_d-2q<\pi_d$; hence this is true if

$$\frac{1}{2}+R_0'(\pi^*)+R_1'(\pi^*)>0. \tag{A6}$$

By equations (A3) and (A4) We have

$$
\begin{aligned}
R_0'(\pi^*) &= \frac{\pi_d}{6}\frac{2+\pi_d-4\pi^*-\pi_d x^2}{(2-\pi_d x)^2} && \text{(A7)}\\
R_1'(\pi^*) &= -\frac{\pi_d}{6}\frac{x(1-x)(2-\pi_d x)-2(2x-1)x^2(1-\pi_d x)}{x^2(2-\pi_d x)^2} && \text{(A8)}
\end{aligned}
$$

where $x\equiv\frac{3}{2}-\pi^*\in\left(\frac{1}{2},1\right)$. Substituting equations (A7) and (A8) into inequality (A6) yields

$$
\begin{aligned}
&3x^2(2-\pi_d x)^2+\pi_d\left(\pi_d-4+4x-\pi_d x^2\right)\\
&-\pi_d x(1-x)(2-\pi_d x)+2\pi_d(2x-1)(1-x)^2(1-\pi_d x)>0 && \text{(A9)}
\end{aligned}
$$

The left-hand side of inequality (A9) is decreasing in π_d if and only if

$$-8x^3 + 2\pi_d - 6 + 10x - 8x^2 - 16\pi_d x^2 + 18\pi_d x^3 - 2\pi_d x^4 + 4\pi_d x < 0. \tag{A10}$$

The left-hand side of inequality (A10) is linear in π_d and, hence, it can be verified to hold by checking the endpoints. We have $-8x^3 - 6 + 10x - 8x^2 < 0$ when $\pi_d = 0$ and $-4+14x-24x^2+10x^3-2x^4 < 0$ when $\pi_d = 1$. Thus, it is sufficient to show that inequality (A10) holds at $\pi_d = 1$. This is true if and only if $G(x) \equiv -x^4 + x^3 - 4x^2 + 12x - 5 > 0$. An interior minimum for G requires $F \equiv -4x^3 + 3x^2 - 8x + 12 = 0$. However, F has no roots in $\left(\frac{1}{2}, 1\right)$; hence it is sufficient to show that $G > 0$ at the endpoints. We have $G\left(\frac{1}{2}\right) = \frac{1}{16} > 0$ and $G(1) = 3 > 0$.

Proof of Proposition 5. We have that $(1-\pi^*)\pi_d - q$ is strictly decreasing in π^*, equalling $\pi_d/2 - q$ when $\pi^* = \frac{1}{2}$ and equalling zero when $\pi^* = \tilde{\pi}$. By equations (24) and (25), $\chi(1-\pi^*)\pi_d\left[\Pi^N(\pi^*) - \Pi^W(\pi^*)\right]$ is strictly positive and equals $\chi\pi_d/8$ when $\pi^* = \frac{1}{2}$. This yields the result.

References

Avery, C. N. and Chevalier, J. A. (1999). 'Herding over the career', *Economics Letters*, 63, 1999, 327-333.

Barber, B. M. and Odean, T. (2001). 'Boys will be boys: gender, overconfidence and common stock investments', *Quarterly Journal of Economics*, 116(1), 261-92.

Biais, B., Hilton, D., Mazurier, K., Pouget, S. (2005). 'Judgemental overconfidence, self-monitoring and trading performance in an experimental financial market', *Review of Economic Studies*, 90(2), 167-312.

Bogan, V. and Just, D. 'What drives merger decision behaviour? Don't seek, don't find and don't change your mind', *Journal of Economic Behavior and Organization*, 72(3), pp. 930-43.

Brocas, I. and Carillo, J. D. (2002). 'Are we all better drivers than average? Self perception and biased behaviour', CEPR Working Paper 3603.

Burnside, C., Han, B., Hirshleifer, D. and Wang, T. Y. (2011). 'Investor overconfidence and the forward premium puzzle', *Review of Economic Studies*, 78(2), pp. 523-58.

Chamley, C. P. (2004). *Economic Models of Social Learning*, Cambridge: Cambridge University Press.

Cho, I. and Kreps, D. M. (1987). 'Signalling games and stable equilibria', *Quarterly Journal of Economics*, 102(2), pp.179-221.

Cho, I. and Sobel, J. (1990). 'Strategic stability and uniqueness in signalling games', *Journal of Economic Theory*, 50(2), pp. 381-413.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998). 'Investor psychology and security market under- and overreactions', *Journal of Finance*, 53(6), pp. 1839-85.

Dewatripont, M., Jewitt, I. and Tirole, J. (1999). 'The economics of career concerns, part I: comparing information structures', *Review of Economic Studies*, 66(1), pp. 183-98.

Ehrbeck, T. and Waldemann, R. (1996). 'Why are professional forecasters biased? Agency versus behavioral explanations', *The Quarterly Journal of Economics*, 111(1), pp. 21-40.

Fudenberg, D. and Tirole, J. (1992). *Game Theory*, London: MIT Press.

Gilpin, K. N. (1999). 'Market insight; the numbers may change, but not the optimism', *New York Times*, 28 Mar.

Gladwell, M. (2009). 'Cocksure: banks, battle and the psychology of overconfidence', *New Yorker*, Jul.

Glaser, M., Nöth, M. and Weber, M. (2004). 'Behavioural finance', in (D. J. Koehler and N. Harvey, eds.), *Blackwell Handbook of Judgment and Decision Making*, pp. 527-46, Oxford: Blackwell Publishing.

Glaser, M., Weber, M. and Langer, T. (2010). 'Overconfidence of professionals and lay people: individual differences within and between tasks', Working Paper, http://ssrn.com/abstract=712583.

Gloede, O. and Menkoff, L. (2009). 'Financial professionals' overconfidence: is it experience, job or attitude', Working Paper, http://www.wiwi.uni-hannover.de/Forschung/

Diskussionspapiere/dp-428.pdf

Hirshleifer, D. and Teoh, S. H. (2003). 'Herd behavior and cascading in capital markets: a review and synthesis', *European Financial Management*, 9(1), pp. 25-66.

Kohn, D. (2008). Testimony before the Senate, in Editors, 'Fedspeak highlights: Kohn on risks to banking Industry', *Wall Street Journal*, 5 Jun.

Kreps, D. and Wilson, R. (1982). 'Reputation and imperfect information', *Journal of Economic Theory* 27(2), pp. 253-79.

Levy, G. (2004). 'Anti-herding and strategic consultation', *European Economic Review*, 48(3), pp. 503-25.

Mailath, George J. (1987). 'Incentive compatibility in signaling games with a continuum of types', *Econometrica*, 55(6), pp. 1349-65.

Nickerson, R. S. (1998). 'Confirmation bias: a ubiquitous phenomenon in many guises', *Review of General Psychology*, 2(2), pp. 175-220.

Oaksford, M. and Chater, N. (1994). 'A rational analysis of the selection task as optimal data selection', *Psychological Review*, 101(4), pp. 608-31.

Odean, T. (1998). 'Volume, volatility, price and profits when all traders are above average', *Journal of Finance*, 53(6), pp. 1887-934.

Ottaviani, M. and Sorenson P. N. (2006). 'Professional advice', *Journal of Economic Theory*, 126(1), pp. 120-142.

Plous, S. (1993). *The Psychology of Judgment and Decision Making*, London: McGraw-Hill, Inc.

Prendergast, C. and Stole, L. (1996). 'Impetuous youngsters and jaded old-timers: acquiring a reputation for learning', *The Journal of Political Economy*, 104(6), pp. 1105-34.

Rajan, R. (2008). 'A view of the liquidity crisis', speech, Chicago.

Ramey, G. (1996). 'D1 signaling equilibria with multiple signals and a continuum of types', *Journal of Economic Theory*, 69(2), pp. 796-821.

Riley, J. (1979). 'Informational equilibria', *Econometrica*, 47(2), pp. 331-359.

Salmon, F. (2008). 'How UBS lost money on super-senior bonds', *Seeking Alpha*, 23 Apr.

Samuelson, R. J. (2008). 'Who's to blame? Why capitalists are capitalism's most dangerous enemies', *Newsweek*, updated 23 Jan.

Scheinkman, J. and Xang, W. (2003). 'Overconfidence and speculative bubbles', *Journal of Political Economy*, 111(6), pp. 1183-219.

Shefrin, H. (2009). 'How psychological pitfalls generated the global financial crisis', Santa Clara University Leavey School of Business Research Paper No. 10-04.

Svenson, O. (1981). 'Are we all less risky and more skillful than our fellow drivers?' *Acta Psychologica*, 47(2), pp. 143–8.

Tett, G. (2008). 'Super-senior losses just a misplaced bet on a currency trade', *Financial Times*, 18 Apr.

Thomas, L. (2003). 'Distinct culture at Bear Stearns helps it surmount a grim market', *New York Times*, 28 Mar.

Trueman, B. (1994). 'Analyst forecasts and herding behavior', *Review of Financial Studies*, 7(1), pp. 97-124.

Tuchman, B. (1984). *The March of Folly: From Troy to Viet Nam*, New York: Random House.

Van den Steen, E. (2004). 'Rational overoptimism and (other biases)', *American Economic Review*, 94(4), pp. 1141-51.

Vanguard Research. (2011). 'Market bubbles and investor psychology', Vanguard.

Weston, D., Blagov, P.S., Harenski, K., Kilts, C. and Hamann, S. (2006). 'Neural bases of motivated reasoning: an fMRI study of emotional constraints on partisan political judgement in the 2004 U.S. presidential election', *Journal of Cognitive Neuroscience*, 18(11), pp. 1947-1958.