Self-Esteem, Shame and Personal Motivation[*]

Roberta Dessí[†] Xiaojian Zhao[‡]

2013 December

Abstract

The available evidence from numerous studies in psychology suggests that overconfidence is a more important phenomenon in North America than in Japan. Relatedly, North Americans appear to view high self-esteem more positively than Japanese. The pattern is reversed when it comes to shame, a social emotion which appears to play a more important role among Japanese than North Americans. We develop an economic model that endogenizes these observed differences. A crucial tradeoff arises in the model between the benefits of encouraging self-improvement and the benefits of promoting initiative and new investments. In this context, self-esteem maintenance (self-enhancement) and high sensitivity to shame emerge as substitute mechanisms to induce efficient effort and investment decisions, generating a "North American" equilibrium with overconfidence and low sensitivity to shame, and a "Japanese" equilibrium with high sensitivity to shame and no overconfidence. The analysis identifies the key equilibrium costs as well as the benefits of reliance on each mechanism, and the implications for welfare.

Keywords: Overconfidence, shame, cultural transmission.
JEL Classification: Z1; D03; D83

[*] For helpful comments and discussions, we would like to thank Olivier Armantier, Douglas Bernheim, Alberto Bisin, Antonio Cabrales, Stephen Chiu, Armin Falk, Raquel Fernandez, Rupert Gatti, Sanjeev Goyal, Johannes Koenen, Gilat Levy, Andrew Postlewaite, Drazen Prelec, Gérard Roland, Paul Seabright, Jean Tirole, Elu von Thadden, Andrea Vindigni, Jidong Zhou, and participants at the CEPR Conference on the Economics of Interactions and Culture (EIEF, Rome), the BEE Workshop on Neuroeconomics and Psychology (TSE, Toulouse), the Maastricht Behavioral and Experimental Economics Symposium, the Econometric Society World Congress (Shanghai), the first conference of the Chinese Game Theory and Experimental Economics Association (Beijing), 6th Biennial Conference of Hong Kong Economic Association (Tianjin), 6th Pan Pacific Conference on Game Theory (Tokyo), North American Summer Meeting of the Econometric Society (St. Louis), Asian Meeting of the Econometric Society (Seoul & Singapore), 4th World Congress of the Game Theory Society (Istanbul), China Meeting of the Econometric Society (Beijing) and seminars in Cambridge, Chengdu (SWUFE), Hangzhou (Zhejiang University), Hong Kong (HKUST & HKU), Madrid (Carlos III), Mannheim, Rome (Tor Vergata), Shenzhen (Peking University HSBC School of Business).

[†] Toulouse School of Economics (GREMAQ and IDEI) and CEPR. Correspondence address: IDEI, Toulouse School of Economics, Manufacture des Tabacs, Aile Jean-Jacques Laffont, 21 Allée de Brienne, 31000 Toulouse, France (roberta.dessi@tse-fr.eu).

[‡] Department of Economics, Hong Kong University of Science and Technology, Hong Kong (xjzhao@ust.hk).

“Pride hurts, modesty benefits.”

The Counsels of Great Yu in the Document of Shangshu, 6th century BC

“All you need in this life is ignorance and confidence; then success is sure.”

Mark Twain (1835 - 1910), *Letter to Mrs Foote, Dec. 2, 1887*

1 Introduction

A large literature in psychology documents people’s need for a positive view of themselves. One aspect of this is individuals’ tendency to selectively focus attention, interpret and remember events so as to maintain or enhance *confidence in their ability*, and thereby maintain self-esteem[1]. Relatedly, recent work in economics shows that individuals tend to update their beliefs differently in response to good news or bad news about their ability.[2]

Interestingly, though, the importance of (over)confidence in one’s ability varies significantly across cultures. Heine et al. (1999) find striking differences in their review of the evidence on North America and Japan (discussed in detail in section 2). In a nutshell, while the distributions of *self-esteem scores* for North Americans are heavily skewed towards high self-esteem, this is *not* the case for the Japanese, whose self-evaluations are lower and approximately normally distributed. Relatedly, the *false uniqueness bias* (the tendency to see oneself as better than most others) has been found in a number of studies of North Americans, but is absent from similar studies of Japanese. Moreover, the *self-serving biases* documented in the North American attribution literature, showing that individuals tend to attribute their successes to their intrinsic characteristics (e.g. talent), while attributing their failures to bad luck or other external factors, do not appear in analogous studies of the Japanese, who tend to attribute failures as much as successes to their own (in)abilities.

It could be conjectured that these differences do not reflect genuine differences in beliefs, but only in the *appearance* of beliefs: the Japanese may wish to appear more modest, while North Americans may wish to appear more confident, than they really are. This, it could be argued, might be a rational response to different social norms, with modest self-presentation gaining greater social approval in Japan, and more confident self-presentation securing greater social approval in North America. However, there is evidence of greater self-enhancement by North Americans and greater self-criticism by Japanese even with complete anonymity of responses, and when individuals are unaware that their behaviors are being observed (see Heine et

[1] We discuss this literature in section 2.

[2] Eil and Rao (2011) find that updating following good news adheres quite closely to the Bayesian benchmark, while updating following bad news produces posterior beliefs nearly uncorrelated with Bayesian inference. Möbius, Niederle, Niehaus and Rosenblat (2013) find that subjects substantially over-weight good news relative to bad news.

al. (1999) for a review). Our paper takes a first step towards understanding these observations, and their economic implications. In our model, differences in beliefs about the self arise endogenously, generating a "North American" equilibrium in which overconfidence is more valuable than in another ("Japanese") equilibrium. Our theory suggests an interpretation of the empirical evidence on cultural differences as reflecting differences in *actual* beliefs *and* in the functional value of those beliefs. Traits such as self-confidence or modesty can then be expected to be viewed as more desirable where their functional value is greater.

Beliefs about the self interact with the social and economic environment. We focus on one main characteristic of the environment that has attracted considerable attention in comparisons of Japan and the United States during the postwar period[3]: the degree of *stability*, or conversely the degree of *mobility*. What we have in mind are the following well-documented[4] differences between the two countries: job mobility has been considerably lower in Japan, largely because of institutions such as "lifetime" employment and late promotions in larger firms, which also offer more attractive conditions than smaller firms; unemployment rates have been lower in Japan; takeovers (with all the changes in management and strategy that they often entail) have been far less frequent in Japan; business start-up rates have been lower in Japan; investments in companies have tended to be more long-term in Japan (partly because of greater reliance on relationship banking); divorce rates have been much lower in Japan... While these observations may have a variety of different causes, the important point for our purposes is that they tend to go in the same direction: i.e. from the point of view of a single individual who takes the environment as given, life in Japan would be expected to entail greater stability (lower mobility) than in the United States. This was clearly the case up to the 1990s; since then the gap seems to have narrowed but certainly not disappeared (see, for example, Moriguchi and Ono (2004), Ono (2010)).

In our model, we capture this characteristic of the environment through the probability, π, that an individual will continue an existing "project" (activity, task, relationship) in the long term. If instead the project comes to an end earlier, the individual has to decide whether to invest in a new project. We can also think of π as representing the expected fraction of projects that are continued in the long run. Obviously there are many situations in which individuals can choose whether to continue an existing project or invest in a new one; however, a salient difference between Japan and the United States seems to be the higher probability of continuing projects in the long term in Japan for institutional and other reasons that are largely exogenous from an individual perspective. Thus, we treat π as a characteristic of the environment. Specifically, we begin by assuming that π is given in a society, reflecting existing institutions and other exogenous characteristics. We then consider the implications of endogenizing π by allowing individuals to vote on the institutions in their society.

The key idea we explore in the first part of the paper is the following. There is

[3] While the quotes presented at the beginning of the paper are intriguing, data on measures of self-esteem and experimental evidence on self-confidence are only available for the postwar period. We therefore focus attention on this period.

[4] See, among others, Hashimoto and Raisian (1985), Imai and Kawagoe (2000), Moriguchi and Ono (2004), Ono (2006, 2010).

a potential tradeoff between the benefits and the costs of overconfidence. In many circumstances, an individual who is overconfident about his talent/skills will overestimate his probability of success if he undertakes a new project ("I am talented, I will succeed"). He will also underestimate the benefits from exerting effort to identify ways of improving his performance on an existing, continuing project ("I am doing fine"). Overconfidence will then increase the likelihood of investing in new projects, while reducing effort to scrutinize performance on existing projects, pay attention to criticism and negative feedback, and seek better ways of doing things. Evidence on this is available for a variety of contexts[5]. The first,"*initiative*" effect can be beneficial when individuals have time-inconsistent preferences, by helping them to undertake new worthwhile projects that would otherwise be forsaken because of a bias in favor of immediate gratification. The second,"*complacency*" effect can inhibit valuable learning and self-improvement.

We analyze the interplay of these two effects in an intrapersonal game[6] between an individual's current self and his future self, where the current self can influence the future self's recall and interpretation of a (current) "bad" signal about his talent/skill. This allows us to capture parsimoniously the possibilities for memory management (e.g. through selective attention) and self-serving interpretations discussed more fully in section 2. At the same time, for much of the analysis we maintain the standard assumption in economics that individuals are rational and "Bayesian": in particular, the future self will update his beliefs taking into account the possibility that the current self might have "suppressed" the bad signal. Thus individuals cannot simply choose their future beliefs as they wish, but they can influence them to some extent.

Importantly, the current self takes into account the probability that the future

[5] Individuals' confidence in their ability has been found to be positively related to their intentions to start new businesses (Chen et al. (1998), De Noble et al. (1999)). Relatedly, patent inventors who chose to start a new business have been found to possess higher levels of self-confidence than patent inventors who chose not to start a new business (Markman et al. (2002)). On the other hand, higher self-confidence has been found to be correlated with persistence in unproductive activities in spite of negative feedback (Whyte and Saks (2007)). Vancouver and Kendall (2006) measured self-confidence and subsequent exam performance for the same individuals taking five different exams. They found a negative relationship at the within-person level of analysis. Leung (2002) examined data from the Third International Mathematics and Science Study (TIMSS), showing that Hong Kong, Japan, Korea and Singapore students outperformed their counterparts in other countries in mathematics achievement. He found that the most striking common factor in these four countries, different from the rest, was the relative low confidence in doing mathematics of the students. Wood and Lynch (2002) looked at the role of prior knowledge in learning about new products in situations where new information makes existing product knowledge obsolete. They found that, compared to consumers with lower prior knowledge, those with higher prior knowledge learn less about a new product, and this is due to inattention at encoding (rather than reconstructive errors at retrieval). Berner and Graber (2008) review the evidence on the link between physician overconfidence and errors in medical diagnosis. While a causal link in this context is particularly difficult to establish, there is some suggestive evidence. For example, in a study of radiologists given sets of "unknown" films to classify as normal or abnormal, the confidence level of the worst performers was higher than that of the top performers. Finally, although the economics literature has not focused on the implications of overconfidence for self-improvement, existing evidence concerning the impact of overconfidence on corporate investment decisions and acquisitions (Malmendier and Tate (2005, 2008)) seems consistent with the "complacency" effect.

[6] The model, with only minor modifications, also admits an interesting alternative interpretation as an interpersonal game between parent and child. We discuss this in section 1.1.

self will have to decide whether to invest in a new project, versus the probability that he will have to choose how much self-improvement effort to exert on an existing, continuing project. Our first main result is that *overconfidence* emerges in equilibrium when, and only when, the relative importance of undertaking new projects versus investing in improving one's performance on existing projects is sufficiently high. An immediate implication of this result is that we are less likely to observe overconfidence in "stable" societies, where the probability of continuing old projects is high, and more likely to observe overconfidence in "dynamic" societies, where this probability is lower. This may help to explain the differences in self-esteem and self-enhancement documented for Japan and North America.

Our model also suggests that perfectionism (meant here as a disposition to be self-critical and to persistently seek improvement) will be considered more valuable in the first type of society, while dynamism (showing initiative, being enterprising) will be considered more valuable in the second. Interestingly, a key business concept that has been very successful in Japan is the idea of *kaizen* ("continuous improvement"), which emphasizes the importance of gradual improvement at the corporate level.[7]

In order to investigate robustness and obtain additional empirical predictions, we extend our model in two ways in section 4. First, we consider a version of the model with a richer signal structure: an individual may receive not only a bad signal or no signal as before, he may also receive a good signal. Our previous result, that overconfidence emerges in equilibrium when the probability of facing new project investment decisions is sufficiently high, continues to hold in this version of the model. Similarly we find again the equilibrium with accurate beliefs for lower values of this probability. For lower values still, there is a new equilibrium, in which individuals suppress the good signal, thereby reducing their *ex-post* confidence. The first equilibrium, in which individuals suppress the bad signal, yields a distribution of beliefs that is skewed towards higher self-confidence (relative to accurate beliefs), since those who receive the bad signal "pool" with those who receive no signal. This pattern resembles the one documented for North America. The second equilibrium, where individuals do not suppress either signal, exhibits no such skewness, and is similar to the pattern observed for Japan. Finally the third equilibrium yields a distribution skewed towards lower self-confidence, different from North America and also from Japan.

We then extend the model in another direction, by considering the role of "naive" agents. These may be individuals who suppress bad signals without being aware of it. Alternatively, they may be individuals who are aware of their biases in processing and recalling information, but lack the cognitive skills required for full Bayesian updating of beliefs *ex post*. Unaware agents do not act strategically; cognitively-constrained agents do, taking into account their cognitive constraints. We show that in very "dynamic" societies, all naive as well as sophisticated agents suppress the bad signal in equilibrium. *Ex post*, naive agents have higher self-confidence than sophisticated agents. On the other hand, in very "stable" societies only unaware agents suppress the bad signal; ex post, sophisticated and cognitively-constrained agents have the

[7] For a short description and commentary, see *The Economist*, 14 October 2009. Just as interesting is their explanation for why *kaizen* has "lost some of its shine" more recently: "Influential in the decline of the idea was the new-found emphasis on the speed of change and on the need for firms to "morph" in double-quick time to seize the opportunities presented by e-commerce and other developments in information technology".

same beliefs. Thus if the population consists of a mixture of sophisticated and naive agents, *average self-confidence will be higher in the more "dynamic" societies.* This is consistent with the evidence reviewed in section 2.

So far, our results have been obtained assuming that π, the degree of stability/dynamism, is given for a particular society. We then ask the question: what happens if citizens can vote over institutions and thereby choose π? In the last part of section 4, we show that for some parameter values this leads to the interesting possibility of *multiple equilibria with endogenous* π. In particular, we may observe two *ex-ante* identical societies in quite different equilibria: in one, individuals suppress bad signals and choose a low value of π, while in the other, individuals do not suppress bad signals, and they choose a high value of π. Thus *overconfidence and dynamism reinforce each other in one equilibrium, realistic self-assessment and stability in the other.*

Our analysis in section 4 focuses entirely on self-confidence as a motivational mechanism. However, shame, and the desire to avoid it, can also be a powerful motivational mechanism. Recent research in social anthropology shows that the capacity to feel shame is pervasive across cultures, but cultures differ significantly in their reliance on the emotion of shame as a motivational mechanism (see Fessler (2007)). This is particularly interesting for our purposes in light of the evidence, discussed in section 2, that shame plays a more important role in Japan than in North America, and that Japanese parenting practices tend to foster sensitivity to shame more than American ones. How do these motivational mechanisms interact?

We explore this question in section 5, where we modify the model by introducing a cost of shame, S, associated with social disapproval. Since decisions to undertake new projects (or not) are, at least imperfectly, observable by others (e.g. starting a new business, taking up new activities, finding a new job, learning new skills, moving to a different location, starting new relationships), they can be subject to social approval (disapproval). In contrast, self-improvement effort is not observable by others. We therefore assume that an individual will experience the cost S if he is faced with the choice to undertake a new project or not, and chooses not to go ahead[8]. We investigate two questions. First, for a given cost of shame S, i.e. taking existing social and cultural norms as given, what is the set of equilibria of the intra-personal game we studied in section 4 without allowing for shame? Second, what would be the socially optimal value of S?

Intuitively, shame can provide incentives to undertake new projects, obviating the need for overconfidence, and thereby improving incentives to invest in self-improvement. However, in the presence of unobservable individual heterogeneity, an equilibrium with shame might imply that some individuals efficiently refrain from undertaking new projects, and are inefficiently penalized for this, or that some individuals inefficiently undertake new projects. In fact, as we show, it is not efficient to have an intermediate cost of shame S, such that some individuals refrain from investment and

[8] We have also investigated a version of the model where the cost of shame is incurred when the original, continuing project fails. However, shame from failure is not efficient in our setting, essentially because failure occurs with some probability even when high effort is provided, so that reliance on this as a motivational mechanism would be too costly. On the other hand, failure will obviously lead to unfavorable updating of beliefs about the individual's ability, which can be interpreted as "stigma" from failure.

incur the cost of shame in equilibrium. The only efficient equilibria involving shame emerge for values of S which induce all individuals to invest rather than incur the cost of shame. Efficient equilibria with *shame*, therefore, entail a form of *conformism.*

We find that reliance on shame can be efficient in dynamic societies as well as in stable societies, depending on parameter values. For very "stable" societies, the efficient equilibrium never entails overconfidence, but may entail an important role for shame. For very "dynamic" societies, on the other hand, the efficient equilibrium will entail either overconfidence and no role for shame, or no overconfidence and an important role for shame. Thus *shame and overconfidence emerge as substitute mechanisms in dynamic societies, while overconfidence plays no role in very stable societies.* This is consistent with the evidence discussed below.

The paper is organized as follows. The remainder of this section relates our work to the existing literature in economics. Section 2 reviews the evidence from psychology, anthropology and economics that motivates our model. Section 3 introduces the baseline model. The costs and benefits of overconfidence are examined in section 4, and a number of extensions of the basic analysis are discussed. Section 5 introduces shame. Section 6 concludes.

1.1 Relationship to the literature

A growing literature is demonstrating the importance of noncognitive skills and traits for a variety of life outcomes[9]. Several theoretical contributions have focused in particular on the role of self-confidence: Bénabou and Tirole (2002) have shown that overconfidence can help to alleviate an under-investment problem arising when preferences are time-inconsistent[10], while Bénabou and Tirole (2011) also consider the psychological benefits of overconfidence in the presence of anticipatory utility, and their implications for identity investments.[11] We share with these papers the assumption that self-confidence can be influenced by biases in information processing and recall. Our focus, however, is on why such biases may be more common in some environments than others, and hence on cultural differences in self-confidence.

In this respect, our work is also related to Alesina and Angeletos (2005), and Bénabou and Tirole (2006). These papers study the interaction between beliefs about the relative importance of effort and luck in determining incomes, and choices of redistributive policies. This leads to the possibility of multiple equilibria, with some societies exhibiting low levels of redistribution and beliefs in the importance of effort, while others exhibit high levels of redistribution and beliefs in the importance of luck. In a similar vein, we show in section 4 how multiple equilibria can arise in our

[9] See Almlund, Duckworth, Heckman and Kautz (2011) for a review and discussion.

[10] For a different approach to the problem of time-inconsistent preferences, see Becker and Mulligan (1997), where individuals devote time and effort to make future pleasures less remote in their mind.

[11] Related papers include Bénabou (2013), which studies denial of bad news in groups when individuals have anticipatory preferences, Compte and Postlewaite (2004), who show that when confidence has a positive effect on performance, biases in information processing can enhance individual welfare, Dessí (2008), where a demand for cultural over-confidence emerges as a solution to the under-investment problem due to the presence of social externalities in cultural investment decisions, and Köszegi (2006), where individuals derive "ego utility" from positive views about their ability. Imperfect self-knowledge is also a key ingredient in the theory of endogenous peer effects developed by Battaglini et al. (2005)

model, with some societies exhibiting greater dynamism and overconfident beliefs, while others exhibit greater stability and no overconfidence.

A simple way to try to explain observed country differences in self-confidence might be to suppose that they are due to country differences in time preference. In our model, overconfidence only emerges in the presence of a bias towards immediate gratification. The observed difference between the U.S. and Japan could then be due to the Japanese being significantly more patient than North Americans. However, we are only aware of one systematic study of country differences in time discounting: Wang, Rieger and Hens (2009) present evidence on β for a sample of 45 countries. Their mean (median) for the U.S. is 0.69 (0.78), *higher* than the corresponding figures for Japan, 0.64 (0.70), implying if anything that the Japanese have a slightly greater present bias. We therefore abstract from differences in time preferences in our model: our results are driven entirely by the trade-off between the costs and benefits of overconfidence, leading to different equilibria in "stable" and "dynamic" societies.

The possibility that observed differences in self-confidence might be due to genetic differences between Japanese and North Americans is sometimes suggested to us, but we are not aware of any evidence supporting this hypothesis. There is, on the other hand, evidence from longitudinal studies of Japanese individuals who moved to Canada and Canadian individuals who moved to Japan, showing a significant tendency for the Canadians' self-confidence to decrease after moving to Japan, while the Japanese' self-confidence increases after moving to Canada (Heine and Lehman (2004), see section 2 below). This suggests that genetic differences, if any, could only be part of the explanation for observed differences in self-confidence.

Another hypothesis that is sometimes put forward concerns the effects of selection in migration patterns, combined with intergenerational transmission of traits. Historically, the argument goes, migrations to North America are likely to have attracted individuals with higher than average self-confidence, who then encouraged and nurtured self-confidence in their children. While the importance of this form of selection would be very difficult to establish empirically[12], the evidence reviewed in section 2.2 does suggest an important role for differences in parenting practices between North America and East Asia. The model we present in section 4, with minor modifications, admits an alternative interpretation in terms of intergenerational transmission, where the "future self" is the child, and the "current self" the parent, who internalizes the child's welfare. Although we do not focus on this interpretation in section 4, to do justice to the evidence on memory and updating biases, we do give it more weight in section 5, where we study the role of shame. Our paper is therefore related to the existing literature on cultural transmission. Bisin and Verdier (2000, 2001) study more generally the intergenerational transmission of cultural traits. We focus instead on specific traits (self-confidence, sensitivity to shame), and examine their role as motivational mechanisms.

The functional role of emotions has attracted economists' attention in recent work, notably in research on envy and regret by Coricelli and Rustichini (2010) and Rustichini (2008). We focus on the emotion of shame and explore the circumstances in

[12] Recent work by Abramitzky, Platt Boustan and Eriksson (2012) has established the importance of selection effects in migrations to America for observable variables, such as occupation and wealth. No corresponding data is available for self-esteem.

which shame emerges as an equilibrium mechanism to induce efficient investment decisions. At the same time, we identify a social cost of reliance on this mechanism, due to the impossibility of tailoring the personal cost of shame to "fit" other, privately known individual characteristics. Thus shame can induce "too much" conformity.

Our approach here builds on the evidence from studies in social anthropology. Reviewing these, Fessler (2007) notes that "shame is prototypically elicited by situations in which i) the actor has failed to live up to some cultural standard for behavior, ii) others are aware of this failure, and iii) the actor is aware of others' knowledge in this regard". It is not clear, in general, to what extent others' disapproval and hostility following the violation of a cultural standard for behavior are a direct reaction to the observed behavior, and to what extent they are derived from preferences over particular individual traits that are inferred from the behavior[13]. Thus our modeling strategy, in which shame attaches to *actions*, seems reasonable in our setting; we view it as complementary to models of conformity where damage to status attaches to inferred predispositions, as in Bernheim (1994).

2 Confidence and shame: evidence for North America and Japan

This section reviews the evidence in psychology, anthropology and economics that motivates our model.

2.1 Overconfidence?

A large literature in psychology has explored people's need for a positive self-view, and, relatedly, the extent to which individuals hold overconfident beliefs about their ability. In this context, overconfidence can be defined in absolute or relative terms: individuals may believe that their ability is greater than it really is, or they may believe that their position in the overall distribution of ability in the relevant population is higher than it really is. We now review the main findings, highlighting the observed differences between North America and Japan.

2.1.1 Self-esteem scores

One very popular approach is to estimate self-esteem scores by asking individuals to report to what extent they agree or disagree[14] with a number of statements intended to capture self-esteem. The ten-item Rosenberg (1965) scale is the most widely used for this purpose, and has been applied in a very large number of studies. Items include "I am able to do things as well as most other people"; "All in all, I am inclined to feel that I am a failure"; and "I take a positive attitude toward myself". The first of

[13] The experimental evidence on how people respond to "unfair" behavior suggests that their reactions are driven by *both*, outcomes *and* inferences about traits/intentions (see, for example, Falk, Fehr and Fischbacher (2003); Fehr and Schmidt (2005) provide an excellent review and discussion).

[14] Possible answers are "strongly disagree", "disagree", "agree" and "strongly agree", with corresponding scores typically from one to four for positive items, and the order reversed for negative items.

these captures specifically beliefs about ability, and clearly does so in relative terms. The other two statements may also capture other influences on self-esteem, and could reflect an absolute comparison (to some standard) or a relative one.

Self-esteem scores appear to differ substantially in North America and Japan, across numerous studies. The distribution of self-esteem scores for *North American* subjects is typically *very skewed towards high self-esteem* (see Baumeister et al. (1989) and Heine et al. (1999) for reviews and discussions); this is *not* the case for *Japanese* subjects (Bond and Cheung (1983), Campbell et al. (1996), Heine et al. (1999), Mahler (1976), Schmitt and Allik (2005)). Moreover, North Americans tend to have significantly higher scores than Japanese for all items but one[15] on the Rosenberg scale, including in particular the item that captures beliefs about (relative) ability (Heine et al. (1999)). Thus while differences in self-esteem may also capture other aspects, they clearly reflect important differences in confidence about ability. Indeed, Schmitt and Allik (2005) decompose global self-esteem scores into subcomponents of self-competence (feeling confident, capable and efficacious) and self-liking: the mean score for *self-competence* is significantly higher for subjects in the United States than in Japan.

An important question then is whether these findings reflect *cultural* differences. Evidence in favor of this interpretation is provided by Heine and Lehman (2004). They obtained self-esteem scores at different points in time for two samples of Japanese students visiting Canada. For one sample they found a significant *increase* in self-esteem with exposure to Canadian culture, while for the other sample the increase was not significant. Heine and Lehman similarly obtained self-esteem scores for a sample of Canadian English teachers who went to live in Japan. They found a significant *decrease* in self-esteem with exposure to Japanese culture.

2.1.2 Other measures of self-confidence and self-enhancement

The findings from studies using self-esteem scores have been confirmed by a large empirical literature in psychology using a variety of related albeit different methods. These include:

(i) studies in which participants evaluate themselves and the average person on the same scale. These studies have found a much greater degree of self-enhancement (the well-known "better-than-average" effect) among North American and Israeli participants than among East Asian (mainly Japanese and Singaporean) participants[16].

(ii) studies in which participants estimate the percentage of people who are more talented than themselves on a variety of dimensions. Here too North American subjects self-enhance much more than Japanese subjects[17], exhibiting the so-called "false uniqueness" effect (a good example of this is given by Svenson (1981): in his US sample, 93% of participants believed themselves to be more skillful than the median in

[15] The exception is the item "I certainly feel useless at times", for which there is no significant difference.

[16] Brown and Kobayashi (2002), Crystal (1999), Endo, Heine and Lehman (2000), Heine and Lehman (1999), Kobayashi and Brown (2003), Kurman (2001, 2003), Kurman and Sriram (2002), Sedikides, Gaertner and Toguchi (2003).

[17] Heine, Kitayama and Lehman (2001), Heine and Lehman (1997), Markus and Kitayama (1991), Norasakkunkit and Kalick (2002).

the group).

(iii) studies in which participants indicate how much their successes and failures are due to their own abilities. American students are much more likely than Chinese or Japanese students to attribute their successes to their ability and their failures to external factors[18].

(iv) studies eliciting participants' memories of their successes and failures. Endo and Meijer (2004) found evidence of self-enhancement among American subjects, but the opposite among Japanese subjects.

All these and other studies have been reviewed in a meta-analysis by Heine and Hamamura (2007): they conclude that North Americans show a clear self-serving bias while East Asians do not.

2.1.3 True or apparent overconfidence?

Benoît and Dubra (2011) have argued that studies where overconfidence is measured by asking individuals to rate themselves relative to the median cannot be used to demonstrate true overconfidence. In particular, the finding that a majority of people rate themselves above the median is consistent with Bayesian updating by individuals with imperfect knowledge of their ability, starting with a common prior. Burks, Carpenter, Goette and Rustichini (2013) have studied the implications of Bayesian updating from a common prior in this context and identified restrictions imposed on the joint distribution of beliefs and true ability. They then tested the restrictions experimentally and rejected them. This, combined with all the other evidence discussed in this section, suggests that overconfidence is an important phenomenon. Yet its importance is significantly greater in North America than in Japan: this observation is the main focus of our paper.

2.1.4 Incentivized beliefs about ability

To economists, beliefs elicited in experiments where subjects are given no monetary incentives to tell the truth may not seem sufficiently reliable. This still would not explain the systematic difference between North American and Japanese responses across a variety of samples. More importantly though, the presence of a substantial bias towards overconfidence among North Americans has been confirmed by Burks et al. (2013), who do address the potential concern over the reliability of answers in the absence of monetary incentives[19]. They administer two tests of cognitive ability to 1016 US subjects, eliciting their beliefs about their ability before and after the test. Each time, subjects are asked to specify which quintile of the group's performance they believe they will be (were) in. Monetary incentives are provided to motivate subjects to give correct answers. The results show that well over 60% of subjects believe they are in the top two quintiles; moreover, overconfident judgements are pervasive wherever possible, i.e. across the first four quintiles of the distribution. Relatedly, Eil and Rao (2011) and Möbius et al. (2013) find, again eliciting incentivized beliefs, that North American subjects revise their beliefs differently in response to good news and bad news (see footnote 2).

[18] Anderson (1999), Endo and Meijer (2004).

[19] See also Hoezl and Rustichini (2005).

2.1.5 Self-esteem maintenance strategies

How are overconfident beliefs sustained? In psychology, a large North American literature has documented the existence of self-serving biases, whereby individuals essentially suppress "bad" signals about their ability and other attributes. This is achieved in a number of ways, including the following:

(a) selective recall of information (e.g. Sanitioso, Kunda and Fong (1990));

(b) subjecting "negative" information to greater scrutiny to find flaws in it or reasons to dismiss its significance (see Baumeister and Newman (1994), Kunda (1990)), and possibly develop alternative explanations that effectively suppress the bad signal (Ditto and Lopez (1992), Ditto et al. (1998));

(c) dismissing the importance of skills one does not have and emphasizing the value of traits one does possess (Dunning and Cohen (1992), Dunning et al. (1989), Tesser and Paulhus (1983));

(d) perceiving own shortcomings as common, own strengths and abilities as uncommon (Muellen and Goethals (1990)).

Yet where attempts have been made to find similar evidence of self-serving biases among Japanese subjects, they have generally failed to do so. For example, as noted earlier, North American subjects tend to attribute their successes to their ability and their failures to external factors such as bad luck (see Zuckerman (1979) for a review). However, studies of Japanese subjects tend to find instead that they attribute failures as much as successes to own (in)abilities (Kitayama et al. (1995), Brown, Gray and Ferrara (2005)).

Relatedly, Baumeister and Jones (1978) found that American participants compensated for negative self-relevant feedback in one domain by inflating their self-assessments in another domain. Heine, Kitayama and Lehman (2001) have investigated whether Canadian and Japanese participants exhibit a similar tendency. All participants were given success or failure feedback following a creativity test; they were then asked to evaluate themselves on dimensions unrelated to creativity. Canadian participants did not show any significant difference in self-evaluations on unrelated dimensions following success or failure feedback on the creativity task. Japanese participants provided *less* favorable self-evaluations on the other dimensions following failure on the creativity test.

Further evidence suggesting that self-esteem maintenance strategies play a more important role for North Americans than for Japanese is provided by studies of self-affirmation and dissonance. In these studies, participants typically choose between two desirable alternatives; they also evaluate the two alternatives before and after making their choice. North American participants usually evaluate their chosen alternative more positively, and the rejected alternative less positively, after making their choice (e.g. Steele, Spencer and Lynch (1993), Heine and Lehman (1997)). This behavior is consistent with a desire to maintain self-esteem by rationalizing one's choices ex post as "the right ones". Japanese participants, in contrast, do not systematically change their evaluations after making their choice (Heine and Lehman (1997)).

2.2 Shame

North Americans and Japanese appear to differ also in terms of the importance they attach to shame. In an influential early work on this topic, Benedict (1946) characterized Japan as a shame culture. She was subsequently criticized by a number of researchers for defining this in terms of reliance on external sanctions (others' disapproval, losing face etc.) for good behavior - a notion sometimes referred to as "public shame". Some authors have emphasized instead the importance of "private shame", whereby others' critical gaze on the self is internalized. Nevertheless, as Heine et al. (1999) pointed out, "most are in agreement that shame occupies a privileged position for Japanese" - a claim that still applies to date (see for example Creighton (1990); Crystal et al. (2001); Doi (1973); Fessler (2007); Johnson (1993); Kuwayama (1992); Lebra (1983)).

In contrast, research by social anthropologists has found that Californians have a "relatively impoverished cognitive/lexical 'landscape of shame'" (Fessler (2007)), and that for Californians shame as an emotion "is overshadowed by guilt" (which, unlike shame, "is prototypically associated with issues of harm to others").

Recent research has investigated the mechanisms that generate cultural differences in the importance of shame. For example, Miller, Fung and Mintz (1996) studied parental practices in American families in Chicago and Chinese families in Taipei. They found that American parents put considerably more emphasis on protecting their children's *self-esteem* than Chinese parents. In contrast, Chinese parents put more emphasis on inducing *shame* and *self-criticism* following behavioral transgressions. Studies focusing on the comparison of Japanese and American mothers have found that the former are more likely to use moral reasoning, to encourage children to think about how others might perceive their behavior, and to induce empathy, guilt, anxiety and shame in response to discipline problems[20].

To summarize, the evidence reviewed in this section points to a more important role for shame in Japan than in North America, and a more important role for (over)confidence and self-esteem in North America than in Japan. We explore possible reasons for this in the remainder of the paper.

3 Baseline model

Our baseline model modifies the one introduced by Bénabou and Tirole (2002). It has two periods and three dates, $t = 0, 1, 2$. At the beginning of the first period ($t = 0$), each individual starts a project (activity, task, relationship). At this stage, individuals are indistinguishable. For simplicity, there is no cost of starting the project. Once they have started, individuals (privately) receive a signal informative about their ability/skill, θ. They choose their interpretation and recall strategy. At $t = 1$ the individual can continue the same project with probability π. In this case, he can, at a cost, invest in self-improvement, thereby increasing the expected returns from the project. With probability $1 - \pi$, on the other hand, the individual cannot continue

[20] See Hess, Kashiwagi, Azuma, Price and Dickson (1980); Kabayashi-Winata and Power (1989); Lewis (1996); Rothbaum, Pott, Azuma, Miyake and Weisz (2000); Weisz, Rothbaum and Balackburn (1984); Zahn-Waxler, Friedman, Cole, Mizuta and Hiruma (1996).

the existing project. In this case he has to decide whether to undertake a new project. All project outcomes are realized at $t = 2$. The timing is depicted in Figure 1.

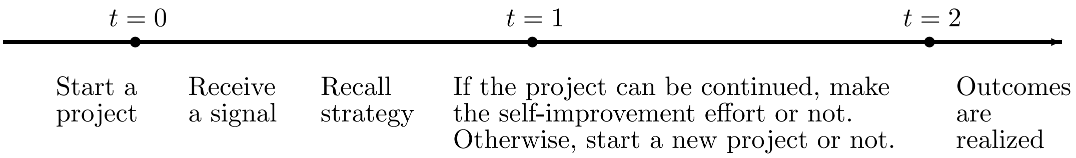
$t = 0$
$t = 1$
$t = 2$
Start a project
Receive a signal
Recall strategy
If the project can be continued, make the self-improvement effort or not. Otherwise, start a new project or not.
Outcomes are realized

Figure 1: Timing

Interpretations of the model

The model, described below, is deliberately stylized, to capture as simply as possible the general tradeoff between the costs and benefits of overconfidence discussed in the Introduction. Several interpretations are possible, each one yielding different insights. According to one interpretation, which will be the main focus of our analysis, individuals receive information about different aspects of their ability/skill from a variety of sources: academic achievements, social interactions, non-academic activities, work, etc. There is plenty of scope for "creative interpretation" of some of the information, and for selective attention to different pieces of information, in ways that generate biased recollections and assessments, as discussed in section 2. In this interpretation, it is today's self (self-0) that influences the information that tomorrow's self (self-1) will recall: the game is intra-personal.

An alternative interpretation, requiring only minor modifications of the model, would be in terms of parental, or more generally inter-generational, transmission of information. Then self-0 would be the older generation (e.g. parents), and self-1 the younger generation: the game is inter-personal. In view of the evidence from the psychology literature discussed in section 2, we choose to focus on the intra-personal game of endogeneous interpretation and recall of information as the main mechanism underlying personal (over)confidence. It should be clear, however, that both mechanisms are at work in determining confidence, and the main insights from our analysis apply to both.

In our model, the early end of the existing project (at $t = 1$) is intended to capture a variety of situations in which individuals cannot continue with the "status quo", and need to decide whether to undertake new activities, initiatives, etc. For example, when a firm is taken over, the change of ownership may bring with it a number of changes in the way the firm is run, so that individual employees have to decide whether to invest in new opportunities within the firm, or possibly search for an alternative employer. Employees who are laid off often have to decide whether to invest in acquiring new skills, or incur the costs of moving. Start-up entrepreneurs whose business fails have to decide whether to seek a "safe" job as employees or invest in trying to start a new business.

These examples mainly concern decisions to do with work in one form or another, but the set of circumstances that may require investment in new activities and initiatives is much broader. A change of government, for instance, may entail significant changes in a variety of policies, making it impossible for many people to hold on to the previous "status quo": each person who is affected by the changes then has to decide how much effort and resources to invest in response to the new circumstances. At a more personal level, changes in family circumstances, such as divorce, also confront individuals with choices about new investments (relationships, home, work, etc.).

In section 4, we shall distinguish between more "stable" societies, in which the probability of being able to hold on to the status quo is higher, and more "dynamic" societies, in which the probability of having to make decisions about investment in new activities is higher. To begin with, these differences will be captured by the exogenous parameter π, reflecting both exogenous factors and institutions. Our approach here will be essentially positive, addressing the following question: for a given set of external factors and society-wide institutions (indexed by π), what patterns of confidence will emerge when individual members of the society attempt to behave (and teach their children to behave) in ways that maximize their expected utility, subject to the constraints implied by those external factors and institutions? In the last part of the section, we will also take a more normative approach and consider welfare consequences. Finally, we will allow for π to be determined endogenously through voting, which enables citizens to choose institutions.

3.1 Projects

The initial project brings a benefit W if it succeeds and zero otherwise. The probability of success depends on the individual's ability; for simplicity, it is equal to θ. We assume that $\theta \in [0, \theta^{\max}]$, where $0 < \theta^{\max} < 1$. Thus even the most talented/skilled individual cannot be sure of success. If the project is continued at $t = 1$, the individual decides whether to exert self-improvement effort: by incurring the cost k, he can increase the probability of success by $\phi(\theta^{\max} - \theta)$, where $1 > \phi > 0$. This assumption captures the idea that by focusing on his failings and weak points, paying attention to criticism and other negative feedback, searching for new information and exploring alternative approaches and ideas, the individual can identify and seek out opportunities for improvement, and thereby achieve a better performance. The scope for such improvement will be greater for individuals with lower initial skill. This specification enables us to model as simply as possible the"complacency" effect of overconfidence discussed in the Introduction[21].

If the existing project cannot be continued at $t = 1$, the individual is faced with a different choice. He can incur a cost c to undertake a new project, which will yield benefit V if successful and zero otherwise. The probability of success in this case is θ. Alternatively, he can undertake another activity whose outcome is less sensitive to ability. For simplicity, we assume that the return from this alternative activity is fixed, and normalize it to zero.

3.2 Preferences

We allow for time-inconsistent preferences by assuming that individuals at $t = 1$ discount expected payoffs at $t = 2$ with a discount factor equal to $\beta\delta$, where δ is

[21] For a colourful account of how overconfidence can inhibit valuable learning and improvement, see also Kroll et al. (2000). Their examples range from strategic decisions at General Motors to Napoleon!

the normal discount rate, while $\beta < 1$ corresponds to hyperbolic discounting. In this case, people give an "excessive" weight to the present.[22]

3.3 Information and beliefs

Self-0 receives a signal s concerning his ability θ. In the baseline model, for simplicity, we focus on the case where s can take just two values: $s = B$ ("bad" signal) and $s = \emptyset$ (no signal). Prior beliefs concerning the signal are described by the probability q; that is, $s = \emptyset$ with probability q and $s = B$ with probability $1 - q$. We can think of q as the proportion of higher-ability individuals in the population. The expected value of θ conditional on each possible realization of the true signal s is given by:

$$\theta_L = E[\theta|s = B] < \theta_H = E[\theta|s = \emptyset].$$

Let $\hat{s}$ be the signal transmitted by self-0 to self-1. We can think of this as (endogeneous) memory. Given our assumptions, if the true signal is $s = \emptyset$, there is no opportunity for signal manipulation; thus $\hat{s} = \emptyset$. On the other hand, if the true signal is $s = B$, self-0 may either communicate the signal truthfully to self-1 ($\hat{s} = B$), or he may decide to suppress the bad signal ($\hat{s} = \emptyset$), as discussed in section 2. At date 1, the state is realized: with probability π the project is continued, otherwise the first project ends and self-1 has to decide whether to undertake a second project. At this date, and before making his investment or effort decision, self-1 privately learns respectively his cost c or k. At date 0, the cost c is known to be uniformly distributed over the interval $[c_L, c_H]$. Similarly the cost k is known to be uniformly distributed over the interval $[k_L, k_H]$.

To make the analysis interesting, we assume that:

$$\delta\phi(\theta^{\max} - \theta_H)W > k_L$$

self-improvement is always efficient if the cost is sufficiently low; and

$$\delta\phi(\theta^{\max} - \theta_L)W < k_H$$

self-improvement is always inefficient if the cost is sufficiently high. Similarly, we assume that:

$$\delta\theta_L V - c_L > 0$$

investment in the new project is always efficient if the cost is sufficiently low, and

$$\delta\theta_H V - c_H < 0$$

investment in the new project is always inefficient if the cost is sufficiently high.

Self-0 has just one decision to make, the recall strategy; that is, the probability that the bad signal will be recalled by self-1:

$$h = \Pr[\hat{s} = B|s = B].$$

We shall denote by h^* the beliefs held by self-1 concerning self-0's strategy.

The recall strategy is depicted in Figure 2.

[22] See Strotz (1955) and Laibson (1997).

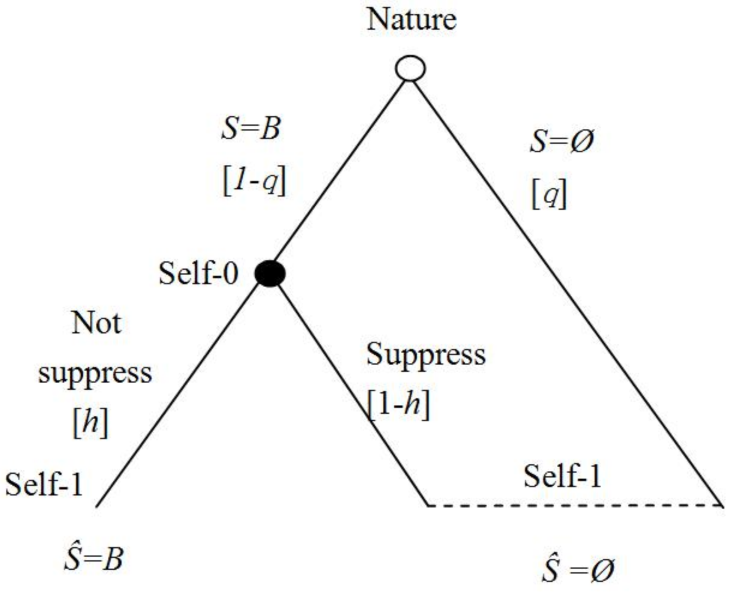

Figure 2: Recall strategy

4 The costs and benefits of overconfidence

4.1 Self-1 belief updating and behavior: sophisticated individuals

Consider self-1's decisions at date 1, in the light of the information available to him. Self-1 has to form expectations over his ability θ. In doing so, he will take into account the possibility that self-0 may have suppressed the true signal s. When $\hat{s} = B$, clearly there has been no suppression; self-1 will therefore have revised beliefs θ_L. When $\hat{s} = \emptyset$, self-1 estimates the following probability that the signal is accurate (the signal's "reliability"):

$$r^* = \Pr[s = \emptyset | \hat{s} = \emptyset; h^*] = \frac{q}{q + (1-q)(1-h^*)}$$

implying that his revised belief is given by:

$$\theta(r^*) = r^*\theta_H + (1 - r^*)\theta_L.$$

Denoting his revised belief by θ^*, clearly self-1 will exert self-improvement effort if, and only if,

$$\beta\delta\phi(\theta^{\max} - \theta^*)W \geqslant k.$$

If the first project has ended, self-1 will undertake the new project if, and only if,

$$\beta\delta\theta^* V - c \geqslant 0.$$

4.2 Self-0 strategy

When $s = B$, self-0 has to choose the recall strategy, h. If he transmits the signal accurately to self-1 ($\hat{s} = B$), his expected utility (ignoring discounting between date 0 and date 1 for simplicity) is given by:

$$\begin{aligned} U_T(\theta_L) = \pi \left[\delta\theta_L W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} \{\delta\phi(\theta^{\max} - \theta_L)W - k\} g dk \right] \\ + (1-\pi) \int_{c_L}^{\beta\delta\theta_L V} \{\delta\theta_L V - c\} f dc \end{aligned}$$

where the subscript T stands for "truth". If on the other hand self-0 suppresses the bad signal ($\hat{s} = \emptyset$), his expected utility depends on self-1's beliefs about the reliability of the signal, r^*, and is given by:

$$\begin{aligned} U_S(\theta_L, \theta(r^*)) = \pi \left[\delta\theta_L W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta^*)W} \{\delta\phi(\theta^{\max} - \theta_L)W - k\} g dk \right] \\ + (1-\pi) \int_{c_L}^{\beta\delta\theta^* V} \{\delta\theta_L V - c\} f dc \end{aligned}$$

where the subscript S stands for "suppression". The net gain from suppressing the bad signal is therefore equal to:

$$\begin{aligned} U_S(\theta_L, \theta(r^*)) - U_T(\theta_L) = -\pi \int_{\beta\delta\phi(\theta^{\max}-\theta^*)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} \{\delta\phi(\theta^{\max} - \theta_L)W - k\} g dk \\ + (1-\pi) \int_{\beta\delta\theta_L V}^{\beta\delta\theta^* V} \{\delta\theta_L V - c\} f dc. \qquad (1) \end{aligned}$$

The first term represents the loss due to overconfidence, which discourages self-improvement effort. The second term represents the impact of overconfidence on the decision to invest in the new project. This yields a gain to the extent that it corrects the under-investment problem due to hyperbolic discounting; if this problem is small, though, there may be excessive confidence and over-investment. For expositional simplicity, we shall focus on the more interesting case where $\beta < \theta_L/\theta_H$, which rules out the possibility of over-investment irrespective of the beliefs held by self-1. We then have a clear tradeoff between the benefits of overconfidence, which alleviates the under-investment problem for new project decisions, and the costs of overconfidence, which exacerbates the problem of under-provision of self-improvement effort.

4.3 Perfect Bayesian equilibria (PBE)

We now characterize the set of Perfect Bayesian equilibria.[23]

[23] While case (i) in Proposition 1 is similar to the case of "defensive pessimism" and case (ii) resembles the leading case in Bénabou and Tirole (2002), even though we vary π rather than β in

Proposition 1 *There exist two threshold values, π_H and π_L (with $\pi_H > \pi_L$), such that: (i) if $\pi > \pi_H$, there is a unique PBE with $h^* = 1$; (ii) if $\pi < \pi_L$, there is a unique PBE with $h^* = 0$; (iii) otherwise, there are three PBEs: the two pure-strategy equilibria with $h^* = 1$ and $h^* = 0$, and a mixed-strategy equilibrium.*

Intuitively, when the probability of continuation of the existing project is sufficiently large, the expected loss from suppressing the bad signal, which discourages self-improvement effort, will be more important than the expected gain, arising from the positive impact of overconfidence on new project investment decisions. Thus the optimal strategy for self-0 will be to transmit the signal truthfully. On the other hand, when the probability of having to choose whether to undertake the new project is high enough, the expected gain from suppression of the bad signal, which alleviates the under-investment problem, will be greater than the expected loss, so that the optimal strategy for self-0 will be to suppress the bad signal. For intermediate values of π, the trade-off is such that there are multiple equilibria: a pure-strategy equilibrium with truthful transmission, a pure-strategy equilibrium with suppression of the bad signal, and a mixed-strategy equilibrium.

The intuition for this result is as follows. Consider some value of π within the intermediate range ($\pi_H > \pi > \pi_L$). When self-1 is very sceptical about the reliability of self-0's signal ($h^* = 0$), θ^* will be relatively low. Note that the benefit of suppressing the bad signal is due to the fact that for some realizations of the cost c which are lower than the expected benefit from investing in the new project, overconfidence will lead to (efficient) investment, whereas in the absence of overconfidence there would be no investment because of hyperbolic discounting. For low θ^* the marginal benefit will be high, since the cost realizations for which this switch to efficient investment will occur will be those for which the net expected benefit from investment is high. As θ^* increases, however, the marginal benefit decreases. The cost of suppressing the bad signal, on the other hand, is due to the fact that for some realizations of the cost k, overconfidence will deter self-1 from exerting self-improvement effort, even though this effort would be efficient. For low θ^*, the marginal cost will be relatively low, since the cost realizations for which the switch away from self-improvement effort will occur will be those for which the net expected benefit from exerting effort is relatively low. As θ^* increases, self-improvement effort is discouraged also for lower cost realizations; i.e. the ones for which the net expected benefit from exerting effort is higher. Thus the marginal cost increases as θ^* increases.

In other words, given π, the gain from suppressing the bad signal increases at a decreasing rate with the level of trust by self-1, while the cost increases at an increasing rate. For π within the intermediate range ($\pi_H > \pi > \pi_L$), therefore, there will be a mixed-strategy equilibrium corresponding to the intermediate level of trust by self-1 which leaves self-0 exactly indifferent between truthful transmission and suppression of the bad signal. In addition, since more trusting beliefs by self-1 will reduce the net gain from suppression, there will be a pure-strategy equilibrium with truthful transmission. Similarly, since less trusting beliefs by self-1 will increase the

this Proposition, it may be worth noting that we cannot simply apply the proof of Proposition 2 in their paper, and thus adopt a different proof method. The details of the difference in the proof are in the Appendix.

net gain from suppression, there will be a pure-strategy equilibrium with suppression of the bad signal.

4.3.1 Overconfidence and underconfidence

The results summarized in Proposition 1 show that different equilibria are possible depending on the value of π, including pure strategy equilibria with accurate recall or complete suppression of the bad signal, as well as mixed strategy equilibria. We now consider the implications for confidence.

In a sufficiently large population, our assumptions mean that a fraction $1 - q$ will observe the bad signal, while the remainder will observe no signal.

In a pure strategy equilibrium with accurate recall, updated beliefs at $t = 1$ will be θ_L for those who observed the bad signal, and θ_H for those who did not: there will be no overconfidence and no underconfidence.

In a pure strategy equilibrium with suppression of the bad signal, updated beliefs at $t = 1$ will be the same for all individuals, equal to $\overline{\theta} \equiv q\theta_H + (1 - q)\theta_L$. Clearly, therefore, there will be both overconfidence and underconfidence in absolute terms. This is because low-ability individuals essentially pool with high-ability individuals: as a consequence, low-ability individuals will have overconfident beliefs, while high-ability individuals will have under-confident beliefs. If we assume that low-ability individuals represent in fact the majority in the population (i.e. $q < 0.5$), the median ability is equal to θ_L, implying that most people will hold overconfident beliefs both in absolute and in relative terms.

4.3.2 Implications and discussion

Our results suggest that overconfidence is more likely to prevail in very "dynamic" societies (low value of π) than in very "stable" societies (high value of π). The US can be thought of as a very dynamic society in the sense of this paper: takeovers play an important role in corporate governance; employee turnover is relatively high; layoffs are common during economic downturns; entrepreneurial activity is high. Politically, two main parties alternate in government. Divorce rates are relatively high.

Japan, during much of the post-war period (the period that shaped the confidence attitudes examined in the psychology studies discussed in section 2), has been a relatively more stable society, with one main party in power during much of the period, an emphasis on lifetime employment with the same firm, a very minor role for takeovers in corporate governance, combined with a tendency to invest for the long term, and to form stable industrial/financial groups.

Our results are therefore consistent with the finding of significantly greater overconfidence in the U.S. than in Japan. They also suggest that confidence attitudes in Japan may change in the future, to the extent that Japan becomes a much more "dynamic" society in the sense of this paper (but see also section 5 on this).

Our main focus is on the US and Japan, two countries for which the distinction between "high-π" and "low-π" societies appears to fit well. They are also the two countries for which the most significant differences in self-esteem scores have been documented in a number of studies. It is nevertheless interesting to look at self-esteem scores for other countries too, presented by Schmitt and Allik (2005). Obviously many countries appear to be "low-π" on some dimensions and "high-π" on others: these countries correspond to those in the intermediate range for π in our model, suggesting that multiple equilibria are possible. We might conjecture, though, that a country like Switzerland will be closer to the "high-π" type, and a country like Israel to the "low-π". Interestingly, Schmitt and Allik (2005) report a relatively low mean self-competence score[24] for Switzerland (14.30) and a high one for Israel (17.50): these can be compared to the reported mean scores for Japan (13.33) and the US (17.21). While our model is too stylized to provide an adequate comprehensive explanation for differences in self-esteem scores across countries, these findings suggest that it may capture part of the explanation, and may usefully inform future empirical work.[25]

4.4 Extensions

The analysis developed in this section can be extended in a number of interesting directions: we review and discuss some of them below.

4.4.1 Richer signal structure

While the main insights of the model emerge clearly in the simplest version with just two signals (bad signal and no signal), it is worth considering what happens if we also allow for a good signal. Formally, the model is modified as follows: s can take one of three values, $s = B$ ("bad" signal) with probability p, $s = \emptyset$ (no signal) with probability q, and $s = G$ ("good" signal) with probability $1 - q - p$. Denote by θ_s the expected value of θ conditional on each possible realization of the true signal s. Naturally, we assume that $\theta_B < \theta_\emptyset < \theta_G$.[26]

If the true signal is $s = \emptyset$, again, there is no opportunity for signal manipulation; thus $\hat{s} = \emptyset$. On the other hand, if the true signal is $s = B$ (or G), self-0 may either communicate the signal truthfully to self-1 ($\hat{s} = B$ (or G)), or he may decide to suppress the signal ($\hat{s} = \emptyset$). Let h_j denote the recall strategy chosen by self-0 when he receives the signal $j \in \{B, G\}$; that is, $h_j = \Pr[\hat{s} = j | s = j]$.

For simplicity, we focus on pure strategy equilibria. The following result rules out the possibility of an equilibrium in which the individual suppresses the bad signal *and* the good signal:

[24] Self-competence scores best capture beliefs about ability, as discussed in section 2. The same rankings emerge if we look instead at global self-esteem scores.

[25] It is tempting to consider possible cross-sectional implications of our analysis: do individuals who expect, for exogenous reasons, to face more frequent new investment decisions tend to hold more confident beliefs, *ceteris paribus*? We are not aware of any empirical study specifically addressing this question. The evidence on CEO overconfidence (Malmendier and Tate (2005, 2008)) is suggestive, but endogeneity is clearly an issue. Casual comparisons between different groups are fraught with difficulties: for example, comparing confidence levels among children of married and divorced parents would need to control for the degree of parental attention, caring and support, which can impact a child's self-esteem directly, generating a potential confound with the effect of divorce on expectations of future stability.

[26] Similarly, we assume that $\delta\phi(\theta^{\max} - \theta_G)W > k_L$, $\delta\phi(\theta^{\max} - \theta_B)W < k_H$, $\delta\theta_B V - c_L > 0$, and $\delta\theta_G V - c_H < 0$. Furthermore, we focus on the more interesting case where $\beta < \theta_B/\theta_G$, which rules out the possibility of over-investment in new projects, and $\beta < (\theta^{\max} - \theta_G) / (\theta^{\max} - \theta_B)$, which rules out the possibility of over-investment in self-improvement, irrespective of the beliefs held by self-1.

Lemma 1 *In any Perfect Bayesian equilibrium, it is impossible to have* $h_B^* < 1$ *and* $h_G^* < 1$.

We therefore have three pure strategy equilibria to consider: one with accurate transmission of both signals, one with suppression of the bad signal, and one with suppression of the good signal. The conditions for each of these three equilibria are given below.

Proposition 2 *There exist four threshold values,* $\pi_L^O < \pi_H^O < \pi_L^U < \pi_H^U$*:*

(i) if $\pi < \pi_L^O$*, there is a unique PBE with* $h_B^* = 0$ *and* $h_G^* = 1$*;*

(ii) if $\pi_L^O < \pi < \pi_H^O$*, there are two PBEs: (a)* $h_B^* = 0$ *and* $h_G^* = 1$*, (b)* $h_B^* = 1$ *and* $h_G^* = 1$*;*

(iii) if $\pi_H^O < \pi < \pi_L^U$*, there is a unique PBE with* $h_B^* = 1$ *and* $h_G^* = 1$*;*

(iv) if $\pi_L^U < \pi < \pi_H^U$*, there are two PBEs: (a)* $h_B^* = 0$ *and* $h_G^* = 1$*, (b)* $h_B^* = 0$ *and* $h_G^* = 0$*;*

(v) otherwise, there is a unique PBE with $h_B^* = 1$ *and* $h_G^* = 0$.

Intuitively, when the expected loss from under-investment in new projects is sufficiently large, it is optimal to suppress the bad signal. On the other hand, when the expected loss from under-provision of self-improvement effort is sufficiently large, it is optimal to suppress the good signal. When the trade-off between these two effects is more balanced, we can have an equilibrium with accurate transmission of both signals.

Thus in very dynamic societies, the bad signal is suppressed in equilibrium, generating a distribution of ex-post beliefs that is skewed towards higher self-confidence (relative to accurate beliefs): individuals who have received the good signal have the highest (and accurate) level of self-confidence, but then those who received the bad signal pool with those who received no signal, achieving a higher level of self-confidence than if they had accurate beliefs. The skewness towards higher self-esteem is consistent with the pattern documented for North Americans, as discussed in section 2.

In relatively more stable societies, we can have the equilibrium with truthful transmission of both signals. Ex post beliefs are then accurate, generating a more symmetric distribution, consistent with the pattern documented for Japan. Finally for societies where the probability of facing new project investment decisions is very low, the equilibrium exhibits suppression of the good signal, generating a distribution of ex-post beliefs skewed towards lower self-confidence. We conjecture that this pattern, which does not resemble those observed for either Japan or the United States, may require a degree of "stability" that is unlikely to be found in the presence of a highly integrated global economy.

4.4.2 Naive agents

Our analysis so far has assumed that individuals are rational and cognitively sophisticated. They are therefore aware of their own incentives to engage in memory-management and self-esteem maintenance strategies, and able to update their beliefs accordingly. In reality, there may also be some naive individuals who suppress bad signals about their ability without being in any way aware that they are doing so, and hence without taking this possibility into account in updating their beliefs ex post. These unaware agents always have *ex-post* beliefs equal to θ_H. Their presence can therefore generate some overconfidence, on average, even in a population where other agents do not engage in self-esteem maintenance strategies.

A different way in which individuals may depart from the assumption of rationality and cognitive sophistication is that they may lack the cognitive skills for full Bayesian updating of beliefs ex post, even though they are aware of the potential scope for memory management ex ante. These agents can behave strategically ex ante, taking into account their cognitive constraints. To see the implications, consider again the baseline version of our model, and suppose that self-1 lacks the cognitive skills for Bayesian updating completely, so that his belief upon observing no signal ($\hat{s} = \emptyset$) is simply $\theta^* = \theta_H$. Knowing this, self-0 expects the net gain from suppressing the bad signal to be equal to:

$$\begin{aligned} U_S(\theta_L, \theta^*) - U_T(\theta_L) = &-\pi \int_{\beta\delta\phi(\theta^{\max}-\theta_H)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} \{\delta\phi(\theta^{\max} - \theta_L)W - k\}gdk \\ &+ (1-\pi)\int_{\beta\delta\theta_L V}^{\beta\delta\theta_H} \{\delta\theta_L V - c\}fdc. \end{aligned}$$

It is straightforward to check that there is then a threshold value π^N such that $0 < \pi^N < 1$ and the net gain from suppressing the bad signal is strictly positive (negative) for $\pi < (>)\pi^N$.

An immediate implication of this (together with Proposition 1) is that for sufficiently low values of π all agents, sophisticated and naive, will suppress the bad signal. Ex post, naive agents will have higher self-confidence (θ_H) than sophisticated agents ($\overline{\theta}$). This is true for naive agents of both types, i.e. those who are unaware and those who are cognitively constrained. In contrast, when π is sufficiently high, only unaware agents will suppress the bad signal. These individuals will always have beliefs equal to θ_H ex post. Cognitively-constrained but aware agents, on the other hand, will have the same ex-post beliefs as sophisticated agents (i.e. accurate beliefs). The average level of self-confidence in the population will therefore be higher, for a given mixture of sophisticated, cognitively-constrained and unaware agents, in very dynamic societies than in more stable ones. This may help to explain the higher average level of self-esteem in the United States, relative to Japan.

4.4.3 Welfare implications

Consider again the baseline version of the model. We have seen that, conditional on observing the "bad" signal concerning their ability, individuals may optimally suppress the signal in some circumstances (depending on the value of π). This leads them to have higher beliefs about their ability than in an equilibrium with accurate recall. On the other hand, in an equilibrium with signal suppression, sophisticated individuals who observe no bad signal will have underconfident beliefs *ex post*, since they rationally take into account the possibility that they may have suppressed a bad signal.

In assessing the welfare implications of memory-management and creative interpretation strategies, we take an *ex ante* perspective[27]: if an individual could choose whether to engage in such strategies or refrain from doing so *before* learning his true ability (more precisely, before observing the true signal s), what would he do?

It turns out that *ex ante* it would be optimal to commit *not* to engage in memory-management and creative interpretation strategies: the expected cost of such strategies outweighs the expected gain. The intuition for this result is the following. As we have seen, in an equilibrium with suppression of the bad signal, low-ability individuals will have overconfident beliefs and high-ability individuals will have underconfident beliefs (in absolute terms). This means that low-ability individuals will be discouraged from investing in self-improvement, while high-ability individuals will be encouraged to invest in self-improvement. Yet it is the low-ability individuals who would benefit most from investment in self-improvement. Similarly, low-ability individuals will be encouraged to invest in the new project, while high-ability individuals will be discouraged: yet it is the high-ability individuals who will benefit most from investment in the new project.

4.4.4 Endogenizing dynamism and stability

Our analysis so far has taken π as a given characteristic of the economic and social environment, reflecting existing institutions as well as other exogenous factors influencing the degree of stability. We now extend the model to allow individuals in a society to vote over institutions, and thereby choose π. This enables us to examine the interaction between belief formation and institutional choices underlying the degree of dynamism or stability in the society. In practice, the degree of stability in a country at any given time will reflect both, institutional choices and a variety of other exogenous influences (e.g. shocks to technology and the natural environment, wars, relevant changes in other countries, etc.). Thus we see the analysis presented earlier and the one developed below as complementary perspectives.

To keep the model as simple as possible, the extension has four dates, $t = 0$, 1, 2, 3. At $t = 0$, each individual starts a project, and receives a signal informative about his ability θ. He chooses his recall strategy. At $t = 1$ each individual updates his beliefs. He then votes on institutions that determine π. For simplicity, π may be high, π_H, or low, π_L. At $t = 2$, each individual learns whether the current project is continuing or ending. He then chooses his effort on the continuing project, or if the project has ended, he decides whether to invest in a new project. All outcomes are realized at $t = 3$.

The novel part occurs at $t = 1$, when individuals update their beliefs and decide how to vote. We assume they vote sincerely for the policy (value of π) that maximizes their expected payoff at $t = 1$, given their updated beliefs. Note that if individuals choose accurate recall at $t = 0$, a fraction q will have updated beliefs θ_H at $t = 1$, and a fraction $1 - q$ will have updated beliefs θ_L. They may vote differently. On the other hand, if individuals choose to suppress bad signals at $t = 0$, they will all have the same updated beliefs at $t = 1$, and vote in the same way.

[27] In taking the *ex ante* perspective we follow much of the literature on hyperbolic discounting. Note however that there is no universal agreement on how to analyze welfare implications when the different selves have conflicting preferences (see Bernheim and Rangel (2009)).

Multiple equilibria with endogenous π Our main interest here is to investigate the possibility of multiple equilibria with endogenous π. In particular, we explore conditions for two pure strategy equilibria to arise: one in which individuals suppress the bad signal at $t = 0$, and then vote for π_L, and one in which they choose accurate recall at $t = 0$, and then vote (at least, a majority of them) for π_H.

At $t = 1$, each individual observes (recalls) either $\hat{s} = B$ or $\hat{s} = \emptyset$. We know from our earlier analysis that updated beliefs will be θ_L if $\hat{s} = B$, and θ^* if $\hat{s} = \emptyset$, where

$$\theta^* = r^*\theta_H + (1 - r^*)\theta_L.$$

The individual's expected payoff when $\hat{s} = B$ is $W_1(B) \equiv U_T(\theta_L)$, given by

$$\begin{aligned} U_T(\theta_L) = E(\pi) &\left(\delta\theta_L W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} (\delta\phi(\theta^{\max} - \theta_L)W - k)\, gdk \right) \\ &+ (1 - E(\pi)) \int_{c_L}^{\beta\delta\theta_L V} (\delta\theta_L V - c)\, fdc \end{aligned}$$

where $E(\pi)$ denotes the expected value of π.

The individual's expected payoff when $\hat{s} = \emptyset$ is $W_1(\emptyset) \equiv r^*U_S(\theta_H) + (1-r^*)U_S(\theta_L)$, where

$$\begin{aligned} U_S(\theta_i) = E(\pi) &\left(\delta\theta_i W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta^*)W} (\delta\phi(\theta^{\max} - \theta_i)W - k)\, gdk \right) \\ &+ (1 - E(\pi)) \int_{c_L}^{\beta\delta\theta^* V} (\delta\theta_i V - c)\, fdc. \end{aligned}$$

It follows that the individual's expected payoff when $\hat{s} = B$ increases (decreases) with E(π) whenever $A > (<)B$, where:

$$\begin{aligned} A &\equiv \delta\theta_L W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} (\delta\phi(\theta^{\max} - \theta_L)W - k)\, gdk \\ B &\equiv \int_{c_L}^{\beta\delta\theta_L V} (\delta\theta_L V - c)\, fdc \end{aligned}$$

Similarly, the individual's expected payoff when $\hat{s} = \emptyset$ increases (decreases) with E(π) whenever $X > (<)Y$, where:

$$\begin{aligned} X &\equiv \delta\theta^* W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta^*)W} (\delta\phi(\theta^{\max} - \theta^*)W - k)\, gdk \\ Y &\equiv \int_{c_L}^{\beta\delta\theta^* V} (\delta\theta^* V - c)\, fdc \end{aligned}$$

Clearly for W sufficiently large relative to V, everyone will vote for π_H, irrespective of their updated beliefs on θ. Similarly for V sufficiently large relative to W, everyone will vote for π_L. The more interesting case for our purposes is where W and V are such that voting behavior does depend on updated beliefs. In particular, we see[28] that for some parameter values we can have $A > B$, implying that a low-ability individual who accurately recalls the bad signal prefers π_H, while $X < Y$, implying that a low-ability individual who suppresses the bad signal will prefer π_L. Thus if low-ability individuals are the majority, it is possible to have two equilibria, one where individuals suppress bad signals and then vote for π_L (*overconfidence and dynamism*), and one where individuals choose accurate recall and then vote for π_H *(no overconfidence and stability).* In particular, these two pure strategy equilibria can emerge when $A > B$ and $X' < Y'$, where X' and $Y'$are the values of X and Y above evaluated at $\theta^* = \overline{\theta}$; i.e.,

$$X' \equiv \delta\overline{\theta}W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\overline{\theta})W} \big(\delta\phi(\theta^{\max} - \overline{\theta})W - k\big)\, gdk,$$

$$Y' \equiv \int_{c_L}^{\beta\delta\overline{\theta}V} \big(\delta\overline{\theta}V - c\big)\, fdc.$$

The intuition for this result is straightforward. A more confident individual is more likely to invest in a new project if the old one comes to an end, and less likely to under-invest because of hyperbolic discounting. His expected payoff is higher when faced with a new investment decision; *ex ante*, this increases the expected benefit from a more dynamic environment. Thus Y increases with θ^*. On the other hand, a more confident individual is less likely to exert self-improvement effort if the old project is continued, which exacerbates the under-provision of effort in the presence of hyperbolic discounting. This effect tends to reduce X as θ^* increases. At the same time, in the absence of self-improvement effort, a more confident individual will have higher expectations of success if the old project is continued: this effect tends to increase X as θ^* increases. When this last effect is relatively weak compared to the first two, $X - Y$ will decrease with θ^*, yielding the possibility of multiple equilibria just discussed.

5 Shame

Our analysis so far has focused on how cultural differences in (over)confidence may emerge in equilibrium when individuals can engage in "creative" interpretation and selective attention strategies to manage their self-esteem. We now extend the analysis to study the role of "social" emotions, in particular shame, and how this interacts with confidence.

Our approach builds on two observations, motivated by the evidence discussed in section 2: (i) the vast majority of people (in all societies) are endowed with a capacity

[28] To see this, note that $X - Y$ is strictly decreasing in θ^* when the following condition holds:

$$\beta\delta(2-\beta)[\phi^2 gW^2\,(\theta^{\max} - \theta^*) + V^2\theta^* f] > W(1 + \phi g k_L) + V f c_L.$$

to feel the emotion of shame[29]; however, (ii) individuals' sensitivity to shame may be enhanced, or reduced, as a consequence of their upbringing and experience of social interactions.

In the model, we now suppose that society can impose a cost of shame S on individuals who adopt certain behaviors. The magnitude of this cost depends on society for two reasons: first, as just noted, because individuals' sensitivity to feelings of shame can be fostered, or reduced, by the social environment in which they grow up (family, school, neighborhood, media, etc.). Second, because society determines not only what constitutes "shameful" behavior, but also a variety of sanctions correlated with the degree of "shamefulness", ranging from mild disapproval to social stigma, ostracism and different kinds of prohibitions and punishments.

We assume that the cost of shame can only be imposed for publicly observable behaviors. This is obviously the case for "public shame", where the cost is linked to losing face and being the target of others' disapproval. However, the psychology literature on shame suggests that the assumption is also reasonable for "private shame", since this essentially internalizes others' critical gaze on the self[30].

In our model, we assume that self-improvement effort is only privately known, while an individual's investment in the new project is observable by others. For example, it may be fairly easy for others to observe whether their friend or acquaintance has found a new job, moved to a different location, embarked on a new degree or training course, started a new business, learned a new skill, started a new relationship, etc. It may be considerably harder for them to observe how much effort he is exerting to come up with better ways of doing his existing job, or how hard he is trying to make an existing relationship work well. To capture this distinction as simply as possible, we assume that the cost S is incurred by individuals who, when the first project ends, do not invest in the new project[31]. Moreover, S cannot depend on the individual's realization of c, since the personal cost of the investment (material and psychological) is only known to the individual.

The optimal social choice of S in our setting can be studied as a representative individual's *ex-ante* choice, "behind the veil of ignorance" (i.e. before he observes his private signal s). Since the magnitude of S will depend a great deal on upbringing and on childhood social interactions (see the evidence reviewed in section 2), one way to think about this in practice is in terms of inter-generational cultural transmission. Thus the older generation (parents) chooses S for the younger generation (children), before learning the realizations of the individual children's ability signals. This seems a reasonable interpretation in light of the evidence that sensitivity to shame is influenced by parenting practices and socialization at an early age (e.g. Miller, Fung and

[29] Fessler (2007) provides an evolutionary account of the development of this emotion, arguing that it evolved from an ancestral form functioning as a mechanism for appeasement in dominance relationships, to a specifically human form functioning to enhance conformity to cultural standards for behavior.

[30] Thus even when an audience is not present, the self may react to the evaluation of an imagined audience ("what would they think if they could see this?"). This still requires the behavior to be potentially observable by the imagined audience, and is therefore more applicable to observable actions than to internal states of mind like "effort".

[31] We can think of these as individuals who persistently do not produce any visible signs of new investments, such as the ones just discussed.

Mintz (1996).

Before studying the optimal choice of S, we need to characterize equilibrium behavior in the presence of an exogenously given cost of shame. This is done below.

5.1 Self-1 behavior

Turning first to self-1's behavior, it is immediate that, since self-improvement effort is not observable, self-1 will exert self-improvement effort if, and only if,

$$\beta\delta\phi(\theta^{\max} - \theta^*)W \geqslant k$$

However, if the first project has ended, self-1 will undertake the new project if, and only if,

$$\beta\delta\theta^* V - c \geqslant -S.$$

5.2 Self-0 strategy

How is self-0's strategy affected by the existence of a cost of shame S, associated with not undertaking the new project? Will this make it easier to alleviate the project under-investment problem ex post, and truthful transmission of the bad signal more attractive ex ante?

Suppose that self-0 observes $s = B$. If he transmits the signal accurately to self-1 ($\hat{s} = B$), his expected utility is given by:

$$\begin{aligned} U_T^S(\theta_L) = {} & \pi[\delta\theta_L W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} \{\delta\phi(\theta^{\max} - \theta_L)]W - k\}gdk] \\ & + (1-\pi)[\int_{c_L}^{\min\{\beta\delta\theta_L V+S,\ c_H\}} \{\delta\theta_L V - c\}fdc - \int_{\min\{\beta\delta\theta_L V+S,\ c_H\}}^{c_H} Sfdc]. \end{aligned}$$

If on the other hand self-0 suppresses the bad signal ($\hat{s} = \emptyset$), his expected utility depends on self-1's beliefs about the reliability of the signal, r^*, and is given by:

$$\begin{aligned} U_S^S(\theta_L, \theta(r^*)) = {} & \pi[\delta\theta_L W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta^*)W} \{\delta\phi(\theta^{\max} - \theta_L)]W - k\}gdk] \\ & + (1-\pi)[\int_{c_L}^{\min\{\beta\delta\theta^* V+S,\ c_H\}} \{\delta\theta_L V - c\}fdc - \int_{\min\{\beta\delta\theta^* V+S,\ c_H\}}^{c_H} Sfdc]. \end{aligned}$$

The net gain from suppressing the bad signal is therefore equal to:

$$\begin{aligned} & U_S^S(\theta_L, \theta(r^*)) - U_T^S(\theta_L) \\ & = -\pi[\int_{\beta\delta\phi(\theta^{\max}-\theta^*)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} \{\delta\phi(\theta^{\max} - \theta_L)W - k\}gdk] \\ & + (1-\pi)[\int_{\min\{\beta\delta\theta_L V+S,\ c_H\}}^{\min\{\beta\delta\theta^* V+S,\ c_H\}} \{\delta\theta_L V - c\}fdc + fS\int_{\min\{\beta\delta\theta_L V+S,\ c_H\}}^{\min\{\beta\delta\theta^* V+S,\ c_H\}} dc]. \end{aligned}$$

Consider first the case where $S \leqslant c_H - \beta\delta\theta^* V$. In this case it is straightforward to verify that the value of the expression does not, in fact, depend on S, and is equal to the value of the corresponding expression without shame, $U_S(\theta_L, \theta(r^*)) - U_T(\theta_L)$, given by equation (1). We therefore have the following result:

Proposition 3 *Suppose society imposes a fixed cost of shame $S \leqslant c_H - \beta\delta\theta^* V$ on individuals who, when faced with the choice to invest or not invest in the new project, decide not to invest. Then irrespective of the magnitude of S, the set of Perfect Bayesian equilibria of the signaling game between self-0 and self-1 will be the same as in the absence of shame, and is described by Proposition 1.*

Now consider the case where $S > c_H - \beta\delta\theta^* V$. Clearly if $S \geq c_H - \beta\delta\theta_L V$, self-1 will always invest in the new project (and thereby avoid incurring any cost of shame), irrespective of the realization of c and of the signal transmitted by self-0. Thus without loss of generality we can focus attention on $S \leqslant c_H - \beta\delta\theta_L V$. When this condition holds as an equality (implying that self-1, as just noted, will always invest in the new project), the net gain from suppressing the bad signal is always strictly negative. There is therefore a unique equilibrium with truthful transmission. In the range $c_H - \beta\delta\theta^* V < S < c_H - \beta\delta\theta_L V$, on the other hand, the set of Perfect Bayesian equilibria is characterized by the following result.

Proposition 4 *Suppose society imposes a fixed cost of shame S on individuals who, when faced with the choice to invest or not invest in the new project, decide not to invest, and this cost satisfies the condition $c_H - \beta\delta\theta^* V < S < c_H - \beta\delta\theta_L V$. Then irrespective of the magnitude of S, the set of Perfect Bayesian equilibria of the signaling game between self-0 and self-1 will be as follows. There exist two threshold values, π_H^S and π_L^S (with $\pi_H^S > \pi_L^S$), such that: (i) if $\pi > \pi_H^S$, there is a unique PBE with $h^* = 1$; (ii) if $\pi < \pi_L^S$, there is a unique PBE with $h^* = 0$; (iii) otherwise, there are three PBEs: the two pure-strategy equilibria with $h^* = 1$ and $h^* = 0$, and a mixed-strategy equilibrium.*

5.3 How much shame?

Is it ever desirable to have a strictly positive cost of shame S? How much shame, if any, is socially optimal? Our model can help to shed light on these questions. In this section, we focus on the two most interesting cases, in terms of comparing very stable (in the sense of having a high value of π) and very dynamic (low value of π) societies.

5.3.1 Very stable societies

We first consider very stable societies, where $\pi > \max(\pi_H, \pi_H^S)$. We know from the results so far that in these societies there is a unique Perfect Bayesian equilibrium of the signaling game between self-0 and self-1, whereby self-0 always transmits his observed signal truthfully.

To study the socially optimal choice of S, we consider a representative individual's choice at date 0 before learning the true value of his signal s. As discussed earlier, we can think of this in terms of intergenerational cultural transmission, with the older generation choosing S for the younger generation before learning the individual realizations of each personal signal s. The choice of S then corresponds to a choice of child-rearing, education and socialization practices, as suggested by the evidence reviewed in section 2.

At date 0, the representative individual expects to observe the "bad" signal, $s = B$, with probability $1 - q$, and no signal, $s = \emptyset$, with probability q. His expected utility is therefore equal to $W_T^S \equiv qU_T^S(\theta_H) + (1-q)U_T^S(\theta_L)$, where

$$\begin{aligned} U_T^S(\theta_i) &= \pi[\delta\theta_i W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_i)W} \{\delta\phi(\theta^{\max}-\theta_i)W - k\}gdk] \\ &\quad + (1-\pi)[\int_{c_L}^{\beta\delta\theta_i V+S} \{\delta\theta_i V - c\}fdc - \int_{\beta\delta\theta_i V+S}^{c_H} Sfdc] \end{aligned}$$

for $i = H, L$. Differentiating by S yields

$$\frac{\partial U_T^S(\theta_i)}{\partial S} = (1-\pi)f[\delta\theta_i V - c_H + S].$$

Remembering that $\overline{\theta} = q\theta_H + (1-q)\theta_L$, we have

$$\frac{\partial W_T^S}{\partial S} = (1-\pi)f[\delta\overline{\theta}V - c_H + S); \quad \frac{\partial^2(W_T^S)}{\partial S^2} = (1-\pi)f > 0$$

which implies that there is no interior solution for S. Thus without loss of generality we can focus attention on two possibilities: $S = 0$ and $S = c_H - \beta\delta\theta_L V$. It can be easily verified that W_T^S may be written as the sum of a term which depends on S and a term which does not depend on S, $W_T^S \equiv W_0 + W(S)$, with

$$\begin{aligned} W(S) &\equiv q(1-\pi)f(\delta\theta_H VS - Sc_H + \frac{1}{2}S^2) + (1-q)(1-\pi)f(\delta\theta_L VS - Sc_H + \frac{1}{2}S^2) \\ &= (1-\pi)f(\delta\overline{\theta}VS - Sc_H + \frac{1}{2}S^2). \end{aligned}$$

We therefore need to compare $W(0)$ and $W(c_H - \beta\delta\theta_L V)$. Clearly $W(0) = 0$. This yields the following result.

Proposition 5 *In very stable societies, where $\pi > \max(\pi_H, \pi_H^S)$, it will be socially optimal to impose a strictly positive cost of shame $S = c_H - \beta\delta\theta_L V$ if, and only if, the following condition holds: $\delta V(2\overline{\theta} - \beta\theta_L) > c_H$.*

The result shows that if time-inconsistency is sufficiently important (β is sufficiently small), the cost of investing in the new project is not too high for any individual (c_H is not too high), and the proportion of high-ability individuals in the population (q) is sufficiently high, the cost of shame is optimally chosen so that in equilibrium everyone undertakes the new project when the old one has ended. Moreover, in equilibrium nobody incurs the cost of shame. However, individuals with a high personal cost c of undertaking the new project will bear a cost in excess of the expected benefit: there will be over-investment.

Thus it can be optimal for shame to play an important role in very stable societies, where it can alleviate the problem of under-investment in new projects. However, since the cost of shame cannot be individually tailored to the (privately known) personal cost of investment, an over-investment problem will arise. When this is less costly than the potential under-investment in the absence of shame, there is an efficiency role for shame.

5.3.2 Very dynamic societies

We now turn to very dynamic societies, where $\pi < \min(\pi_L, \pi_L^S)$. We know from the results obtained earlier that in these societies there is a unique Perfect Bayesian equilibrium of the signaling game between self-0 and self-1, whereby self-0 always suppresses the bad signal, *unless* the cost of shame is set so high that everyone invests in the new project when the old project has ended ($S = c_H - \beta\delta\theta_L V$), irrespective of the signal received by self-0 and the realization of the personal investment cost c. In the latter case, the unique PBE entails truthful transmission of the signal by self-0.

Once again, we study a representative individual's choice at date 0 before learning the true value of his signal s. At this stage, the individual expects to observe the "bad" signal, $s = B$, with probability $1 - q$, and no signal, $s = \emptyset$, with probability q. To begin with, consider the case where $S = c_H - \beta\delta\theta_L V$. Expected utility is then equal to W_T^S, evaluated at $S = c_H - \beta\delta\theta_L V$. Denote the value of expected utility in this case by $W_T^S(c_H - \beta\delta\theta_L V)$. Now consider the case where $S < c_H - \beta\delta\theta_L V$. Expected utility is equal to $W_S^S \equiv qU_S^S(\theta_H) + (1-q)U_S^S(\theta_L)$, where

$$
\begin{aligned}
U_S^S(\theta_i) = \pi[\delta\theta_i W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\bar{\theta})W} \{\delta\phi(\theta^{\max} - \theta_i)W - k\} g dk] \\
+ (1-\pi)[\int_{c_L}^{\beta\delta\bar{\theta}V+S} \{\delta\theta_i V - c\} f dc - \int_{\beta\delta\bar{\theta}V+S}^{c_H} S f dc]
\end{aligned}
$$

for $i = H, L$. Differentiating by S yields

$$
\frac{\partial U_S^S(\theta_i)}{\partial S} = (1-\pi) f [\delta\theta_i V - c_H + S]
$$

and hence

$$
\frac{\partial W_S^S}{\partial S} = (1-\pi) f [\delta\bar{\theta} V - c_H + S); \quad \frac{\partial^2 (W_S^S)}{\partial S^2} = (1-\pi) f > 0
$$

just as in the case of very stable societies examined earlier, implying that there is no interior solution for S. Thus here too, without loss of generality, we can focus attention on two possibilities: $S = 0$ and $S = c_H - \beta\delta\bar{\theta}V$. The second of these two corresponds to the case where the desire to avoid incurring the cost of shame is strong enough to motivate everybody to invest in the new project if the old project ends. But we know that setting $S = c_H - \beta\delta\theta_L V$ achieves the same outcome, *and* induces truthful transmission in equilibrium, which is optimal from an ex ante perspective (see welfare analysis in the previous section).

We can therefore obtain the following result.

Proposition 6 *In very dynamic societies, where $\pi < \min(\pi_L, \pi_L^S)$, it will be socially optimal to impose a strictly positive cost of shame $S = c_H - \beta\delta\theta_L V$ if, and only if, the following condition holds: $\pi A + \frac{1}{2}(1-\pi)f(c_H - \beta\delta\overline{\theta}V)[B - C] > 0 \qquad (C1)$, where $A \equiv g\delta\phi W^2(1 - \frac{1}{2}\beta^2\delta\phi)q(1-q)(\theta_H - \theta_L)^2 > 0$, $B \equiv \delta\overline{\theta}V - \beta\delta\overline{\theta}V > 0$, and $C \equiv c_H - \delta\overline{\theta}V > 0$. Thus we have two possible equilibria:*

(i) An equilibrium with overconfidence (suppression of bad signal) and no cost of shame ($S = 0$), when $C1$ does not hold

(ii) An equilibrium without overconfidence (truthful transmission) and a high cost of shame ($S = c_H - \beta\delta\theta_L V$), when $C1$ does hold.

Condition $C1$ has an intuitive interpretation: A represents the expected gain from reliance on shame when the status quo project is continued in the long term. Since the equilibrium with shame entails no overconfidence, while the one without shame entails overconfidence, clearly reliance on shame provides better incentives to exert self-improvement effort: this is captured by A. However, this gain occurs with relatively low probability in very dynamic societies. With relatively high probability, the individual will need instead to decide whether to invest in a new project. The term $B - C$ captures the net gain from reliance on shame in this case. The presence of a high cost of shame makes it possible to"correct" more efficiently the under-investment incentives associated with time-inconsistent preferences than would be possible through memory management: this effect is captured by B. There is a price for this though: when the cost of shame is high, individuals whose investment cost is higher than the expected benefit will nevertheless invest, to avoid shame. This loss is captured by C.

Clearly if $B \geqslant C$, condition $C1$ will be satisfied for all values of π in the relevant range ($\pi < \min(\pi_L, \pi_L^S)$), and irrespective of the magnitude of A. However, it is straightforward to verify that $B \geqslant C$ is a stronger condition than the necessary and sufficient condition for shame to be optimal in very stable societies, given in Proposition 5. If $B < C$, shame may not be efficient in very dynamic societies. In particular, shame is less likely to be efficient as π decreases, and as A decreases relative to $C - B$.

5.3.3 Implications and discussion

Our analysis in this section has shown that reliance on shame as a motivational device can be efficient in stable *and* in dynamic societies, depending on parameter values. In stable societies, two types of equilibria can emerge: neither of the two will entail overconfidence, while one of them will entail an important role for shame. In dynamic societies there are also two types of equilibria: one with overconfidence and no motivational role for shame, and the other with no overconfidence and an important role for shame.

This is consistent with the evidence reviewed in section 2. Moreover, it suggests that even if "stable" societies become more "dynamic", in the sense of this paper, this may not lead to cultural convergence in terms of the relative importance of social emotions like shame, and self-esteem maintenance or self-enhancement. The example of the "Four Asian Tigers" (Hong Kong, Singapore, South Korea and Taiwan) is interesting in this respect: their mean self-competence scores are relatively low

(Schmitt and Allik (2005))[32], and we saw in section 2 that shame plays an important role in Taiwan.

Our discussion has focused on the socially optimal choice of S, the cost of shame. However, as mentioned earlier, our results can be interpreted in terms of intergenerational cultural transmission: parents choosing S for their children before learning the individual realizations of the children's ability signals would make the same choice. This interpretation fits well with the evidence on how parenting practices and socialization at an early age emphasize sensitivity to shame, or alternatively the importance of self-confidence, as discussed in section 2.

6 Conclusion

Comparisons across cultures provide a very valuable opportunity for understanding how economics and psychology interact. In this paper, we have focused on self-esteem and shame, both of which have received considerable attention in the psychology literature but far less attention in the economics literature. The available evidence from numerous studies by psychologists suggests that overconfidence is a much more important phenomenon in North America than in Japan. Relatedly, North Americans appear to view high self-esteem much more positively than Japanese. The pattern is reversed when it comes to shame, which appears to play a much more important role among Japanese than North Americans.

We have developed an economic model that can rationalize these observed differences. The model studies a potential tradeoff between the benefits of encouraging self-improvement and the benefits of promoting initiative and new investments. In this context, self-esteem maintenance (self-enhancement) and sensitivity to shame emerge as (substitute) mechanisms to induce efficient effort and investment decisions. While exploring their instrumental value, we also identify some important costs associated with the use of each mechanism in equilibrium: reliance on self-esteem maintenance strategies means that in equilibrium the incentives to invest in self-improvement will be reduced for the individuals who could benefit most from such investment, and similarly for investment in new projects. On the other hand, reliance on shame as an incentive mechanism means that in equilibrium there will be over-investment.

The analysis presented here suggests a number of promising directions for future research: for example, the model can be readily extended to study the role of management practices such as *kaizen* (continuous improvement), and the potential tradeoff with innovation. Perhaps most importantly to our minds, our work represents a first step towards integrating the role of social emotions, time-inconsistent preferences and self-serving biases into economic models able to shed light on observed economic and psychological differences across cultures.

[32] Schmitt and Allik present self-competence scores for three of the "Four Asian Tigers"; the missing one is Singapore.

References

[1] Abramitzky, R., L. Boustan, and K. Eriksson (2012). "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102(5), 1832-56.

[2] Alesina, A., and G. Angeletos (2005). "Fairness and Redistribution," *American Economic Review,* 95(4), 960-980.

[3] Almlund, M., A. Duckworth, J. Heckman, and T. Kautz (2011). "Personality Psychology and Economics," *Handbook of the Economics of Education*, E. Hanushek, S. Machin, and L. Woessman (eds.), Vol. 4, pp. 1-181.

[4] Anderson, C.A. (1999). "Attributional Style, Depression, and Loneliness: A Cross-Cultural Comparison of American and Chinese Students," *Personality and Social Psychology Bulletin*, 25, 482-499.

[5] Battaglini, M., R. Bénabou and J. Tirole (2005). "Self-Control in Peer Groups," *Journal of Economic Theory*, 123(2), 105-134.

[6] Baumeister, R.F. and E.E. Jones (1978). "When Self-Presentation is Constrained by the Target's Knowledge: Consistency and Compensation," *Journal of Personality and Social Psychology,* 36, 608-618.

[7] Baumeister, R.F., D.M. Tice and D.G. Hutton (1989). "Self-Presentational Motivations and Personality Differences in Self-Esteem," *Journal of Personality,* 57, 547-579.

[8] Baumeister, R.F. and L.S. Newman (1994). "Self-Regulation of Cognitive Inference and Decision Processes," *Personality and Social Psychology Bulletin*, 20, 3-19.

[9] Becker, G. and C. Mulligan (1997). "The Endogenous Determination of Time Preference," *Quarterly Journal of Economics*, 112(3), 729-58.

[10] Bénabou, R. (2013). "Groupthink: Collective Delusions in Organizations and Markets," *Review of Economic Studies*, 80, 429-462.

[11] Bénabou, R. and J. Tirole (2002). "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics,* 117(3), 871-915.

[12] Bénabou, R. and J. Tirole (2006). "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics,* 121(2), 699-746.

[13] Bénabou, R. and J. Tirole (2011). "Identity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 126(2), 805-855.

[14] Benedict, R. (1946). *The Chrysanthemum and the Sword*, Boston: Houghton Mifflin.

[15] Benoît, J-P. and J. Dubra (2011). "Apparent Overconfidence," *Econometrica*, 79(5), 1591-1625.

[16] Berner, E.S. and M.L. Graber (2008). "Overconfidence as a Cause of Diagnostic Error in Medicine," *American Journal of Medicine*, 121(5), Supplement, S2-S23.

[17] Bernheim, B.D. (1994). "A Theory of Conformity," *Journal of Political Economy*, 102(5), 841-877.

[18] Bernheim, B.D., and A. Rangel (2009). "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics,* 124(1), 51-104.

[19] Bisin, A. and T. Verdier (2000). ""Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, 115(3), 955-88.

[20] Bisin, A. and T. Verdier (2001). "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97(2), 298-319.

[21] Bond, M.H. and T. Cheung (1983). "College Students' Spontaneous Self-Concept," *Journal of Cross-Cultural Psychology,* 14, 153-171.

[22] Brown, J.D. and C. Kobayashi (2002). "Self-enhancement in Japan and America," *Asian Journal of Social Psychology*, 5, 145-168.

[23] Brown, R.A., R.R. Gray and M.S. Ferrara (2005). "Attributions for Personal Achievement Outcomes among Japanese, Chinese, and Turkish University Students," *Information and Communication Studies,* 33, 1-13.

[24] Burks, S.V., J.P. Carpenter, L. Goette, and A. Rustichini (2013). "Overconfidence and Social Signalling," *Review of Economic Studies*, 80(3), 949-983.

[25] Campbell, J.D., P.D. Trapnell, S.J. Heine, I.M. Katz, L.F. Lavallee and D.R. Lehman (1996). "Self-Concept Clarity: Measurement, Personality Correlates, and Cultural Boundaries," *Journal of Personality and Social Psychology,* 70, 141-156.

[26] Chen ,C.C., P.G. Greene, and A. Crick (1998). "Does Entrepreneurial Self-efficacy Distinguish Entrepreneurs from Managers?" *Journal of Business Venturing*, 13, 295-316.

[27] Compte, O. and A. Postlewaite (2004). "Confidence-enhanced Performance," *American Economic Review*, 94(5), 1536-57.

[28] Coricelli, G. and A. Rustichini (2010). "Counterfactual Thinking and Emotions: Regret and Envy Learning," *Philosophical Transactions of the Royal Society B*, 365, 241-247.

[29] Creighton, M.R. (1990). "Revisiting Shame and Guilt Cultures: A Forty-Year Pilgrimage," *Ethos,* 18, 279-307.

[30] Crystal, D.S. (1999). "Attributions of Deviance to Self and Peers by Japanese and U.S. Students," *Journal of Social Psychology*, 139, 596-610.

[31] De Noble, A.F., D. Jung, and S.B. Ehrlich (1999). "Entrpreneurial Self-efficacy: the Development of a Measure and its Relationship to Entrepreneurial Action," *Frontiers of Entrepreneurship Research*, Reynolds P., Bygrave W., Manigart S., Mason C., Meyer G., Sapienza H. & K. Shaver (eds.), Babson College: Babson Park, MA.

[32] Dessí, R. (2008). "Collective Memory, Cultural Transmission and Investments," *American Economic Review,* 98(1), 534-560.

[33] Ditto, P.H. and D.F. Lopez (1992). "Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions," *Journal of Personality and Social Psychology*, 63, 568-584.

[34] Ditto, P.H., J.A. Scepansky, G.D. Munro, A.M. Apanovitch, and L.K. Lockhart (1998). "Motivated Sensitivity to Preference-Inconsistent Information," *Journal of Personality and Social Psychology*, 75, 53-69.

[35] Doi, T. (1973). *The Anatomy of Dependence,* Tokyo: Kodansha.

[36] Dunning, D. and G.L. Cohen (1992). "Egocentric Definitions of Traits and Abilities in Social Judgement," *Journal of Personality and Social Psychology*, 63, 341-355.

[37] Dunning, D., J.A. Meyerowitz, and A.D. Holzberg (1989). "Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability," *Journal of Personality and Social Psychology*, 57, 1082-1090.

[38] Eil, D. and J. Rao (2011). "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 3(2), 114-38.

[39] Endo, Y., Heine, S.J. and D.R. Lehman (2000). "Culture and Positive Illusions in Relationships: How my relationships are better than yours," *Personality and Social Psychology Bulletin*, 26, 1571-1586.

[40] Endo, Y. and Z. Meijer (2004). "Autobiographical Memory of Success and Failure Experiences," *Progress in Asian social psychology*, Kashima, Y., Endo, Y., Kashima, E.S., Leung, C. and J. McClure (eds.), Seoul, Korea: Kyoyook-Kwahak-Sa.

[41] Falk, A., E. Fehr, and U. Fischbacher (2003). "On the Nature of Fair Behavior," *Economic Inquiry*, 41, 20-26.

[42] Fehr, E. and K.M. Schmidt (2005). "The Economics of Fairness, Reciprocity and Altruism - experimental evidence and new theories," chapter 8, *Handbook of the Economics of Giving, Altruism and Reciprocity*, Volume 1.

[43] Fessler, D.M.T. (2007). "From Appeasement to Conformity: Evolutionary and Cultural Perspectives on Shame, Competition, and Cooperation," *The Self-Conscious Emotions: Theory and Research*, Jessica L. Tracy, Richard W. Robins and June Price Tangney (eds.), The Guilford Press, New York.

[44] Fiske, S. and S. Taylor (2008). *Social Cognition: from brains to culture*, McGraw-Hill Press.

[45] Hashimoto, M. and J. Raisian (1985). "Employment Tenure and Earnings Profiles in Japan and the United States," *American Economic Review*, 75(4), 721-735.

[46] Heine, S.J. and T. Hamamura (2007). "In Search of East-Asian Self-enhancement," *Personality and Social Psychology Review*, 11(1), 4-27.

[47] Heine, S.J., S. Kitayama. and D.R. Lehman (2001). "Cultural Differences in Self-Evaluation," *Journal of Cross-Cultural Psychology,* 32(4), 434-443.

[48] Heine, S. J., and D.R. Lehman (1997). "Culture, Dissonance, and Self-Affirmation," *Personality and Social Psychology Bulletin,* 23, 389-400.

[49] Heine, S. J., and D.R. Lehman (1999). "Culture, Self-Discrepancies, and Self-Satisfaction," *Personality and Social Psychology Bulletin,* 25, 915-925.

[50] Heine, S. J., and D. R. Lehman (2004). "Move the Body, Change the Self: Acculturative Effects on the Self-Concept," *Psychological Foundations of Culture*, M. Schaller and C. Crandall (eds.), 305-331.

[51] Heine, S. J., D.R. Lehman, H. R. Markus, and S. Kitayama (1999). "Is There a Universal Need for Positive Self-Regard?" *Psychological Review,* 106, 766-794.

[52] Hess, R., K. Kashiwagi, H. Azuma, G.G. Price, and W.P. Dickson (1980). "Maternal Expectations for Mastery of Developmental Tasks in Japan and the United States," *International Journal of Psychology,* 15, 259-271.

[53] Hoezl, E. and A. Rustichini (2005). "Overconfident: Do You Put Your Money on It?" *Economic Journal*, 115(503), 305-318.

[54] Imai, Y. and M. Kawagoe (2000). "Business Start-ups in Japan: Problems and Policies," *Oxford Review of Economic Policy,* 16(2), 114-123.

[55] Johnson, F.A. (1993). *Dependency and Japanese Socialization,* New York: New York University Press.

[56] Kabayashi-Winata, H., and T.G. Power (1989). "Child Rearing and Compliance: Japanese and American families in Houston," *Journal of Cross-Cultural Psychology,* 20, 333-356.

[57] Kitayama, S., H. Takagi and H. Matsumoto (1995). "Causal Attribution of Success and Failure: Cultural Psychology of the Japanese Self," *Japanese Psychological Review,* 38, 247-280.

[58] Kobayashi, C. and J.D. Brown (2003). "Self-Esteem and Self-Enhancement in Japan and America," *Journal of Cross-Cultural Psychology*, 34, 567-580.

[59] Köszegi, B. (2006). "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, 4(4), 673-707.

[60] Kroll, M.J., L.A. Toombs and P. Wright (2000). “Napoleon’s Tragic March Home from Moscow: lessons in hubris,” *Academy of Management Executive,* 14(1), 117-128.

[61] Kunda, Z. (1990). “The Case for Motivated Reasoning,” *Psychological Bulletin*, 108, 480-498.

[62] Kurman, J. (2001). “Self-Enhancement: is it restricted to individualistic cultures?” *Personality and Social Psychology Bulletin*, 12, 1705-1716.

[63] Kurman, J. (2003). “Why is Self-Enhancement Low in Certain Collectivist Cultures? An investigation of two competing explanations,” *Journal of Cross-Cultural Psychology*, 34, 496-510.

[64] Kurman, J. and N. Sriram (2002). “Interrelationships among Vertical and Horizontal Collectivism, Modesty, and Self-Enhancement,” *Journal of Cross-Cultural Psychology*, 33, 71-86.

[65] Kuwayama, T. (1992). “The Reference Other Orientation,” *Japanese Sense of Self,* N.R. Rosenberger (eds.), (pp. 121-149) Cambridge, England: Cambridge University Press.

[66] Laibson, D. (1997). “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics,* 62, 443-477.

[67] Lebra, T.S. (1983). “Shame and Guilt: a psychological view of the Japanese self,” *Ethos,* 11, 192-209.

[68] Leung, F.K.S. (2002). “Behind the High Achievement of East Asian Students,” *Educational Research and Evaluation*, 8(1), 87-108.

[69] Lewis, C. (1995). *Educating Hearts and Minds*, Cambridge, England: Cambridge University Press.

[70] Lewis, C. (1996). “Fostering Social and Intellectual Development: The roots of Japanese educational success,” *Teaching and Learning in Japan,* T.P. Rohlen and G.K. Le Tendre (Eds.), New York: Cambridge University Press.

[71] Mahler, I. (1976). “What is the Self-Concept in Japan?” *Psychologia,* 19, 127-132.

[72] Malmendier, U. and G. Tate (2005). “CEO Overconfidence and Corporate Investment,” *Journal of Finance,* 60(6), 2661-2700.

[73] Malmendier, U. and G. Tate (2008). “Who Makes Acquisitions? CEO Overconfidence and the Market’s Reaction,” *Journal of Financial Economics,* 89(1), 20-43.

[74] Markman, G.D. , R.A. Baron, and D.B. Balkin (2002). “Inventors and New Venture Formation: the effects of general self-efficacy and regretful thinking,” *Entrepreneurship Theory and Practice*, 27(2), 149-166.

[75] Miller, P.J., H. Fung, and J. Mintz (1996). "Self-Construction through Narrative Practices: A Chinese and American comparison of early socialization," *Ethos,* 24(2), 237-280.

[76] Möbius M., M. Niederle, P. Niehaus, and T. Rosenblat (2013). "Managing Self-Confidence: Theory and Experimental Evidence," mimeo.

[77] Moriguchi, C. and H. Ono (2004). "Japanese Lifetime Employment: a century's perspective", EIJS Working Paper, The European Institute of Japanese Studies.

[78] Mullen, B. and G.R. Goethals (1990). "Social Projection, Actual Consensus and Valence," *British Journal of Social Psychology*, 29, 279-282.

[79] Norasakkunkit, V. and M.S. Kalick (2002). "Culture, Ethnicity, and Emotional Distress Measures: The role of self-construal and self-enhancement," *Journal of Cross-Cultural Psychology*, 33, 56-70.

[80] Ono, H. (2006). "Divorce in Japan: why it happens, why it doesn't," EIJS Working Paper, The European Institute of Japanese Studies.

[81] Ono, H. (2010). "Lifetime Employment in Japan: concepts and measurements", *Journal of the Japanese and International Economies*, 24(1), 1-27.

[82] Rosenberg, M. (1965). *Society and Adolescent Self-Image,* Princeton, NJ: Princeton University Press.

[83] Rothbaum, F., M. Pott, A. Azuma, K. Miyake, and J. Weisz (2000). "The Development of Close Relationships in Japan and the United States: Paths of symbiotic harmony and generative tension," *Child Development,* 71, 1121-1142.

[84] Rustichini, A. (2008). "Dominance and Competition," *Journal of the European Economic Association*, 6, 647-656.

[85] Sanitioso, R., Z. Kunda, and G.T. Fong (1990). "Motivated Recruitment of Autobiographical Memory," *Journal of Personality and Social Psychology*, 59, 229-241.

[86] Schmitt, D. and J. Allik (2005). "Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem," *Journal of Personality and Social Psychology,* 89(4), 623–642.

[87] Sedikides, C., L. Gaertner, and Y. Toguchi (2003). "Pancultural Self-enhancement," *Journal of Personality and Social Psychology*, 84, 60-79.

[88] Steele, C.M., S.J. Spencer and M. Lynch (1993). "Self-Image Resilience and Dissonance: The role of affirmational resources," *Journal of Personality and Social Psychology,* 64, 885–896.

[89] Strotz, R. (1955). "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies,* 23, 165-180.

[90] Svenson, O. (1981). "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" *Acta Psychologica,* 47, 143-148.

[91] Tesser, A. and D. Paulhus (1983). "The Definition of Self: Private and public self-evaluation management strategies," *Journal of Personality and Social Psychology*, 44, 672-682.

[92] Vancouver, J.B. and L.N. Kendall (2006). "When Self-efficacy Negatively Relates to Motivation and Performance in a Learning Context," *Journal of Applied Psychology*, 91, 1146-1153.

[93] Wang, M., M. O. Rieger and T. Hens (2009). "An International Survey on Time Discounting," SSRN working paper.

[94] Weisz, J., F.M. Rothbaum, and T.C. Balackburn (1984). "Standing Out and Standing In: the psychology of control in America and Japan," *American Psychologist,* 39, 995-969.

[95] Whyte, G. and A.M. Saks (2007). "The Effects of Self-efficacy on Behaviour in Escalation Situations," *Human Performance*, 20, 23-42.

[96] Wood, S.L. and J.G. Lynch (2002). "Prior Knowledge and Complacency in New Product Learning," *Journal of Consumer Research*, 29(3), 416-26.

[97] Zahn-Waxler, C., R.J. Friedman, P.M. Cole, I. Mizuta, and N. Hiruma (1996). "Japanese and United States Preschool Children's Responses to Conflict and Distress," *Child Development,* 67, 2462-2477.

[98] Zuckerman, M. (1979). "Attribution of Success and Failure Revisited, or: the motivational bias is alive and well in attribution theory," *Journal of Personality,* 47, 245–287.

7 Appendix

7.1 Proof of Proposition 1

Proof. Define

$$\begin{aligned}&X\left(r^*,\pi\right)\\&\equiv U_S(\theta_L,\theta(r^*))-U_T(\theta_L)\\&=-\pi X_1+(1-\pi)X_2\end{aligned}$$

where $X_1 \equiv \int_{\beta\delta\phi(\theta^{\max}-\theta^*)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W-k\}dG(k)>0$ and $X_2 \equiv \int_{\beta\delta\theta_L V}^{\beta\delta\theta^* V}\{\delta\theta_L V-c\}dF(c)>0$.

It is clear that $X\left(r^*,\pi\right)$ is continuous and decreasing in π for all $r^*\in[q,1]$ as X_1 and X_2 are both positive. Further, we have that $X\left(r^*,1\right)<0$ and $X\left(r^*,0\right)>0$ for all r^*. Thus there is a unique $\pi^*\left(r^*\right)$ such that $X\left(r^*,\pi^*\left(r^*\right)\right)=0$, and $X\left(r^*,\pi\right)>0$ for all $\pi<\pi^*\left(r^*\right)$, and $X\left(r^*,\pi\right)<0$ for all $\pi>\pi^*\left(r^*\right)$ for all r^*.

By the implicit function theorem, we have that

$$\begin{aligned}\frac{d\pi^*}{d\theta^*}&=-\frac{\frac{dX(r^*,\pi)}{d\theta^*}}{\frac{dX(r^*,\pi)}{d\pi}}\\&=\frac{(1-\pi)\beta\delta V^2\left(\delta\theta_L-\beta\delta\theta^*\right)f-\pi\beta\delta\phi W^2[\delta\phi(\theta^{\max}-\theta_L)-\beta\delta\phi(\theta^{\max}-\theta^*)]g}{\int_{\beta\delta\phi(\theta^{\max}-\theta^*)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W-k\}dG(k)+\int_{\beta\delta\theta_L V}^{\beta\delta\theta^* V}\{\delta\theta_L V-c\}dF(c)}\end{aligned}$$

where the denominator is always positive. Notice that the sign of the numerator is ambiguous. Thus, our proof here is not a straightforward extension of the proof of Proposition 2 in Bénabou and Tirole (2002).

There are three cases to consider.

(I) For π sufficiently small, the numerator is positive. Since the numerator is decreasing in θ^*, formally, we have $\frac{d\pi^*}{d\theta^*}>0$ for

$$\pi<\pi_1\equiv\frac{V^2\left(\theta_L-\beta\theta_H\right)f}{V^2\left(\theta_L-\beta\theta_H\right)f+\phi^2W^2[(1-\beta)\theta^{\max}-\left(\theta_L-\beta\theta_H\right)]g}.$$

For values of π satisfying this condition, $\pi^*\left(r^*\right)$ is increasing in r^*, since θ^* is increasing in r^*.

Notice that it is straightforward to verify that $\pi^*\left(q\right)>\pi^*\left(1\right)>\pi_1$.

To show it, since

$$\begin{aligned}&X\left(q,\pi^*\left(q\right)\right)\\&=-\pi^*\left(q\right)\int_{\beta\delta\phi(\theta^{\max}-\overline{\theta})W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W-k\}dG(k)+\left(1-\pi^*\left(q\right)\right)\int_{\beta\delta\theta_L V}^{\beta\delta\overline{\theta}V}\{\delta\theta_L V-c\}dF(c)\\&=0\end{aligned}$$

where $\overline{\theta}\equiv q\theta_H+(1-q)\theta_L$, we have

$$\pi^*\left(q\right)=\frac{\int_{\beta\delta\theta_L V}^{\beta\delta\overline{\theta}V}\{\delta\theta_L V-c\}dF(c)}{\int_{\beta\delta\phi(\theta^{\max}-\overline{\theta})W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W-k\}dG(k)+\int_{\beta\delta\theta_L V}^{\beta\delta\overline{\theta}V}\{\delta\theta_L V-c\}dF(c)}.$$

Similarly, we have

$$\pi^*(1) = \frac{\int_{\beta\delta\theta_L V}^{\beta\delta\theta_H V}\{\delta\theta_L V - c\}dF(c)}{\int_{\beta\delta\phi(\theta^{\max}-\theta_H)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W - k\}dG(k) + \int_{\beta\delta\theta_L V}^{\beta\delta\theta_H V}\{\delta\theta_L V - c\}dF(c)}.$$

The sign of $\pi^*(q) - \pi^*(1)$ equals the sign of

$$\begin{aligned}
&\int_{\beta\delta\phi(\theta^{\max}-\theta_H)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W - k\}dG(k)\int_{\beta\delta\theta_L V}^{\beta\delta\overline{\theta}V}\{\delta\theta_L V - c\}dF(c)\\
&-\int_{\beta\delta\phi(\theta^{\max}-\overline{\theta})W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W - k\}dG(k)\int_{\beta\delta\theta_L V}^{\beta\delta\theta_H V}\{\delta\theta_L V - c\}dF(c)\\
&= \frac{(1-\beta)(\theta_H-\theta_L)\left(\overline{\theta}-\theta_L\right)\left(\theta_H-\overline{\theta}\right)\theta^{\max}gf\beta^3\delta^4\phi^2W^2V^2}{2},
\end{aligned}$$

which is positive.

Further,

$$\begin{aligned}
&\pi^*(1) - \pi_1\\
&= \frac{\int_{\beta\delta\theta_L V}^{\beta\delta\theta_H V}\{\delta\theta_L V - c\}dF(c)}{\int_{\beta\delta\phi(\theta^{\max}-\theta_H)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W - k\}dG(k) + \int_{\beta\delta\theta_L V}^{\beta\delta\theta_H V}\{\delta\theta_L V - c\}dF(c)}\\
&- \frac{V^2(\theta_L-\beta\theta_H)f}{V^2(\theta_L-\beta\theta_H)f + \phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L-\beta\theta_H)]g}\\
= \frac{A}{A+B} &- \frac{V^2(\theta_L-\beta\theta_H)f}{V^2(\theta_L-\beta\theta_H)f + \phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L-\beta\theta_H)]g}
\end{aligned}$$

where

$$A = (\theta_L - \beta\theta_H + (1-\beta)\theta_L)(\theta_H-\theta_L)V^2f$$

and

$$B = ((1-\beta)(2\theta^{\max}-\theta_L) - (\theta_L-\beta\theta_H))(\theta_H-\theta_L)\phi^2W^2g.$$

It further equals

$$\frac{A\phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L-\beta\theta_H)]g - BV^2(\theta_L-\beta\theta_H)f}{(A+B)(V^2(\theta_L-\beta\theta_H)f + \phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L-\beta\theta_H)]g)}$$

The denominator is positive. The numerator equals

$$\phi^2\beta\theta^{\max}fgV^2W^2(\theta_H-\theta_L)^2(1-\beta)$$

which is also positive.

Thus we must have $X(r^*, \pi) > 0$ for all r^*. Therefore, there is a unique PBE with $h^* = 0$.

(II) The numerator is negative for π sufficiently large. Since the numerator is decreasing in θ^*, formally, we have $\frac{d\pi^*}{d\theta^*} < 0$ for

$$\pi > \pi_2 \equiv \frac{V^2\left(\theta_L-\beta\overline{\theta}\right)f}{V^2\left(\theta_L-\beta\overline{\theta}\right)f + \phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L-\beta\overline{\theta})]g}$$

where $\pi_2 > \pi_1$.

For values of π satisfying this condition, $\pi^*(r^*)$ is decreasing in r^*, since θ^* is increasing in r^*. Moreover, it is straightforward to verify that $\pi^*(q) > \pi_2$.

To see it,

$$\begin{aligned}
&\pi^*(q) - \pi_2 \\
&= \frac{\int_{\beta\delta\theta_L V}^{\beta\delta\overline{\theta}V}\{\delta\theta_L V - c\}dF(c)}{\int_{\beta\delta\phi(\theta^{\max}-\overline{\theta})W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W}\{\delta\phi(\theta^{\max}-\theta_L)W - k\}dG(k) + \int_{\beta\delta\theta_L V}^{\beta\delta\overline{\theta}V}\{\delta\theta_L V - c\}dF(c)} \\
&- \frac{V^2\left(\theta_L - \beta\overline{\theta}\right)f}{V^2\left(\theta_L - \beta\overline{\theta}\right)f + \phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L - \beta\overline{\theta})]g} \\
&= \frac{C}{C+D} - \frac{V^2\left(\theta_L - \beta\overline{\theta}\right)f}{V^2\left(\theta_L - \beta\overline{\theta}\right)f + \phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L - \beta\overline{\theta})]g}
\end{aligned}$$

where

$$C = \left(\theta_L - \beta\overline{\theta} + (1-\beta)\theta_L\right)\left(\overline{\theta} - \theta_L\right)V^2 f$$

and

$$D = \left((1-\beta)(2\theta^{\max} - \theta_L) - (\theta_L - \beta\overline{\theta})\right)\left(\overline{\theta} - \theta_L\right)\phi^2W^2g.$$

It further equals

$$\frac{C\phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L - \beta\overline{\theta})]g - DV^2\left(\theta_L - \beta\overline{\theta}\right)f}{(C+D)\left(V^2\left(\theta_L - \beta\overline{\theta}\right)f + \phi^2W^2[(1-\beta)\theta^{\max} - (\theta_L - \beta\overline{\theta})]g\right)}$$

The denominator is positive. The numerator equals

$$\phi^2\beta\theta^{\max}fgV^2W^2\left(\overline{\theta} - \theta_L\right)^2(1-\beta)$$

which is also positive.

We therefore have the following results when $\pi > \pi_2$.

(i) If $\pi > \pi^*(q)$, $X(r^*, \pi) < 0$ for all r^*. Therefore, there is a unique PBE with $h^* = 1$.

(ii) If $\pi < \pi^*(1)$, we have that $X(r^*, \pi) > 0$ for all r^*. Therefore, there is a unique PBE with $h^* = 0$.

(iii) If $\pi^*(1) < \pi < \pi^*(q)$, since $\pi^*(r^*)$ is a decreasing function, the inverse function $r^*(\pi)$ is also decreasing. Thus $X(r^*, \pi)$ has the same sign of $r^*(\pi) - r^*$, implying that there are three PBEs: (a) $r^* = 1$ ($h^* = 1$) with $r^* > r^*(\pi)$, (b) $r^* = q$ ($h^* = 0$) with $r^* < r^*(\pi)$, and (c) a mixed one with h^* such that $X(r^*(\pi), \pi) = 0$.

(III) For intermediate values of $\pi \in [\pi_1, \pi_2]$, there is a threshold value $\theta(\pi)$ such that when $\theta^* < \theta(\pi)$, $\frac{d\pi^*}{d\theta^*} > 0$,and when $\theta^* > \theta(\pi)$, $\frac{d\pi^*}{d\theta^*} < 0$. Thus $\pi^*(r^*)$ increases in r^* as long as r^* is smaller than some cutoff value $\overline{r}$ and decreases thereafter.

We therefore have the following results when $\pi \in [\pi_1, \pi_2]$.

(i) If $\pi < \pi^*(1)$, we have that $X(r^*, \pi) > 0$ for all r^*. Therefore, there is a unique PBE with $h^* = 0$.

(ii) If $\pi^*(1) < \pi < \pi^*(q)$, there are three PBEs: (a) $r^* = 1$ ($h^* = 1$), (b) $r^* = q$ ($h^* = 0$), and (c) a mixed one with h^* such that $X(r^*, \pi) = 0$.

To complete the proof, let $\pi_H \equiv \pi^*(q)$, and $\pi_L \equiv \pi^*(1)$. ■

7.2 Proof of Lemma 1

Proof. When $\hat{s} = B$, or G, clearly there has been no suppression so that the revised belief is $\theta^*(B) = \theta_B$ and $\theta^*(G) = \theta_G$. When $\hat{s} = \emptyset$, self-1's estimate of its reliability is given by:

$$r^*(\emptyset) = \Pr[s = \emptyset | \hat{s} = \emptyset; h_B^*; h_G^*] = \frac{q}{p(1-h_B^*) + q + (1-q-p)(1-h_G^*)},$$

and self-1's belief that this is actually a bad signal is given by:

$$b^*(\emptyset) = \Pr[s = B | \hat{s} = \emptyset; h_B^*; h_G^*] = \frac{p(1-h_B^*)}{p(1-h_B^*) + q + (1-q-p)(1-h_G^*)}.$$

It implies that his revised belief of his ability conditional on no signal $\emptyset$ is given by:

$$\theta^*(\emptyset) = r^*(\emptyset)\theta_\emptyset + b^*(\emptyset)\theta_B + (1 - r^*(\emptyset) - b^*(\emptyset))\,\theta_G$$

which is strictly greater than θ_B and strictly less than θ_G.

When $s = B$, self-0 has to choose the recall strategy, h_B. If he transmits the signal accurately to self-1 ($\hat{s} = B$), his expected utility is given by:

$$U_T(\theta_B) = \pi \left[\delta\theta_B W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_B)W} \{\delta\phi(\theta^{\max} - \theta_B)W - k\} g dk \right]$$
$$+ (1-\pi) \int_{c_L}^{\beta\delta\theta_B V} \{\delta\theta_B V - c\} f dc.$$

If on the other hand self-0 suppresses the bad signal ($\hat{s} = \emptyset$), his expected utility is given by:

$$U_S(\theta_B, \theta^*) = \pi \left[\delta\theta_B W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta^*)W} \{\delta\phi(\theta^{\max} - \theta_B)W - k\} g dk \right]$$
$$+ (1-\pi) \int_{c_L}^{\beta\delta\theta^* V} \{\delta\theta_B V - c\} f dc.$$

The net gain from suppressing the bad signal is therefore equal to:

$$\begin{aligned} U_S(\theta_B, \theta^*) - U_T(\theta_B) &= (1-\pi) \int_{\beta\delta\theta_B V}^{\beta\delta\theta^* V} \{\delta\theta_B V - c\} f dc \\ &\quad - \pi \int_{\beta\delta\phi(\theta^{\max}-\theta^*)W}^{\beta\delta\phi(\theta^{\max}-\theta_B)W} \{\delta\phi(\theta^{\max} - \theta_B)W - k\} g dk \\ &= \frac{\beta\delta^2(\theta^* - \theta_B)}{2} \left[(1-\pi) f V^2 X_2 - \pi\phi^2 g W^2 X_1 \right] \end{aligned}$$

where

$$X_2 = 2\theta_B - \beta\theta_B - \beta\theta^*$$

and

$$X_1 = 2\theta^{\max} - 2\beta\theta^{\max} + \beta\theta^* + \beta\theta_B - 2\theta_B.$$

Similarly, when $s = G$, self-0 has to choose the recall strategy, h_G. If he transmits the signal accurately to self-1 ($\hat{s} = G$), his expected utility is given by:

$$\begin{aligned} U_T(\theta_G) = \pi & \left[\delta\theta_G W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_G)W} \{\delta\phi(\theta^{\max} - \theta_G)W - k\} g dk \right] \\ & + (1-\pi) \int_{c_L}^{\beta\delta\theta_G V} \{\delta\theta_G V - c\} f dc. \end{aligned}$$

If self-0 suppresses the bad signal ($\hat{s} = \emptyset$), his expected utility is given by:

$$\begin{aligned} U_S(\theta_G, \theta^*) = \pi & \left[\delta\theta_G W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta^*)W} \{\delta\phi(\theta^{\max} - \theta_G)W - k\} g dk \right] \\ & + (1-\pi) \int_{c_L}^{\beta\delta\theta^* V} \{\delta\theta_G V - c\} f dc. \end{aligned}$$

The net gain from suppressing the bad signal is therefore equal to:

$$\begin{aligned} U_S(\theta_G, \theta^*) - U_T(\theta_G) &= \pi \int_{\beta\delta\phi(\theta^{\max}-\theta_G)W}^{\beta\delta\phi(\theta^{\max}-\theta^*)W} \{\delta\phi(\theta^{\max} - \theta_G)W - k\} g dk \\ & \quad - (1-\pi) \int_{\beta\delta\theta^* V}^{\beta\delta\theta_G V} \{\delta\theta_G V - c\} f dc \\ &= \frac{\beta\delta^2 (\theta_G - \theta^*)}{2} \left[\pi\phi^2 g W^2 Y_1 - (1-\pi) f V^2 Y_2 \right] \end{aligned}$$

where

$$Y_1 = 2\theta^{\max} - 2\beta\theta^{\max} + \beta\theta^* + \beta\theta_G - 2\theta_G$$

and

$$Y_2 = 2\theta_G - \beta\theta_G - \beta\theta^*.$$

Here, we can show that $X_1 > Y_1$, and $X_2 < Y_2$ as $2 > \beta$ and $0 < \theta_B < \theta_G$.

Suppose $U_S(\theta_B, \theta^*) - U_T(\theta_B) \geq 0$. Then $(1-\pi) f V^2 X_2 \geq \pi\phi^2 g W^2 X_1$ because $\theta^* > \theta_B$.

Given that $X_2 < Y_2$ and $X_1 > Y_1$, we have that $(1-\pi) f V^2 Y_2 > \pi\phi^2 g W^2 Y_1$.

Then $U_S(\theta_G, \theta^*) - U_T(\theta_G) < 0$ because $\theta_G > \theta^*$. ■

7.3 Proof of Proposition 2

Proof. First, we check the existence condition for the PBE with $h_B^* = 0$. By Lemma 1, we know that in this PBE we must have $h_G^* = 1$.

Thus this PBE exists if $U_S(\theta_B, \theta^*) - U_T(\theta_B) \geq 0$ where $\theta^* = (p\theta_B + q\theta_\emptyset) / (p+q)$, that is,

$$(1-\pi) f V^2 X_2 (\theta_B, \theta^*) - \pi\phi^2 g W^2 X_1 (\theta_B, \theta^*) \geq 0$$

where

$$\begin{aligned} & X_2(\theta_B, \theta^*) \\ & = X_2 (\theta_B, (p\theta_B + q\theta_\emptyset) / (p+q)) \\ & = 2\theta_B - \beta\theta_B - \beta (p\theta_B + q\theta_\emptyset) / (p+q) \end{aligned}$$

and

$$\begin{aligned}&X_1\left(\theta_B,\theta^*\right)\\&=X_1\left(\theta_B,\left(p\theta_B+q\theta_\emptyset\right)/\left(p+q\right)\right)\\&=2\theta^{\max}-2\beta\theta^{\max}+\beta\left(p\theta_B+q\theta_\emptyset\right)/\left(p+q\right)+\beta\theta_B-2\theta_B.\end{aligned}$$

It is equivalent to

$$\begin{aligned}&\pi\leq\pi_H^O\\&=\frac{fV^2X_2\left(\theta_B,\left(p\theta_B+q\theta_\emptyset\right)/\left(p+q\right)\right)}{fV^2X_2\left(\theta_B,\left(p\theta_B+q\theta_\emptyset\right)/\left(p+q\right)\right)+\phi^2gW^2X_1\left(\theta_B,\left(p\theta_B+q\theta_\emptyset\right)/\left(p+q\right)\right)}.\end{aligned}$$

Second, we check the existence condition for the PBE with $h_G^*=0$. By Lemma 1, we know that in this PBE we must have $h_B^*=1$.

Thus this PBE exists if $U_S(\theta_G,\theta^*)-U_T(\theta_G)\geq 0$ where $\theta^*=\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right),$ that is,

$$\pi\phi^2gW^2X_1\left(\theta_G,\theta^*\right)-(1-\pi)fV^2X_2\left(\theta_G,\theta^*\right)\geq 0$$

where

$$\begin{aligned}&X_1\left(\theta_G,\theta^*\right)\\&=X_1\left(\theta_G,\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right)\right)\\&=2\theta^{\max}-2\beta\theta^{\max}+\beta\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right)+\beta\theta_G-2\theta_G.\end{aligned}$$

and

$$\begin{aligned}&X_2\left(\theta_G,\theta^*\right)\\&=X_2\left(\theta_G,\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right)\right)\\&=2\theta_G-\beta\theta_G-\beta\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right).\end{aligned}$$

It is equivalent to

$$\begin{aligned}&\pi\geq\pi_L^U\\&=\frac{fV^2X_2\left(\theta_G,\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right)\right)}{fV^2X_2\left(\theta_G,\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right)\right)+\phi^2gW^2X_1\left(\theta_G,\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right)\right)}.\end{aligned}$$

Third, we check the existence condition for the PBE with $h_G^*=1$ and $h_B^*=1$.

This PBE exists if $U_S(\theta_B,\theta^*)-U_T(\theta_B)\leq 0$ and $U_S(\theta_G,\theta^*)-U_T(\theta_G)\leq 0$ where $\theta^*=\theta_\emptyset$.

$$U_S(\theta_B,\theta^*)-U_T(\theta_B)\leq 0$$

is equivalent to

$$(1-\pi)fV^2X_2\left(\theta_B,\theta^*\right)-\pi\phi^2gW^2X_1\left(\theta_B,\theta^*\right)\leq 0$$

where

$$\begin{aligned}&X_2(\theta_B,\theta^*)\\&=X_2\left(\theta_B,\theta_\emptyset\right)\\&=2\theta_B-\beta\theta_B-\beta\theta_\emptyset\end{aligned}$$

and

$$\begin{aligned}&X_1\left(\theta_B,\theta^*\right)\\&=X_1\left(\theta_B,\theta_\emptyset\right)\\&=2\theta^{\max}-2\beta\theta^{\max}+\beta\theta_\emptyset+\beta\theta_B-2\theta_B.\end{aligned}$$

It is equivalent to

$$\begin{aligned}&\pi\geq\pi_L^O\\&=\frac{fV^2X_2\left(\theta_B,\theta_\emptyset\right)}{fV^2X_2\left(\theta_B,\theta_\emptyset\right)+\phi^2gW^2X_1\left(\theta_B,\theta_\emptyset\right)}.\end{aligned}$$

Since $\theta_\emptyset>\left(p\theta_B+q\theta_\emptyset\right)/\left(p+q\right)$, $X_2\left(\theta_B,\theta^*\right)$ is decreasing in θ^*, and $X_1\left(\theta_B,\theta^*\right)$ is increasing in θ^*, it is clear that $\pi_L^O<\pi_H^O$.

Furthermore,

$$U_S(\theta_G,\theta^*)-U_T(\theta_G)\leq 0$$

is equivalent to

$$\pi\phi^2gW^2X_1\left(\theta_G,\theta^*\right)-(1-\pi)fV^2X_2\left(\theta_G,\theta^*\right)\leq 0$$

where

$$\begin{aligned}&X_1\left(\theta_G,\theta^*\right)\\&=X_1\left(\theta_G,\theta_\emptyset\right)\\&=2\theta^{\max}-2\beta\theta^{\max}+\beta\theta_\emptyset+\beta\theta_G-2\theta_G.\end{aligned}$$

and

$$\begin{aligned}&X_2\left(\theta_G,\theta^*\right)\\&=X_2\left(\theta_G,\theta_\emptyset\right)\\&=2\theta_G-\beta\theta_G-\beta\theta_\emptyset.\end{aligned}$$

It is equivalent to

$$\begin{aligned}&\pi\leq\pi_H^U\\&=\frac{fV^2X_2\left(\theta_G,\theta_\emptyset\right)}{fV^2X_2\left(\theta_G,\theta_\emptyset\right)+\phi^2gW^2X_1\left(\theta_G,\theta_\emptyset\right)}.\end{aligned}$$

Since $\theta_\emptyset<\left(\left(1-p-q\right)\theta_G+q\theta_\emptyset\right)/\left(1-p\right)$, $X_2\left(\theta_G,\theta^*\right)$ is decreasing in θ^*, and $X_1\left(\theta_G,\theta^*\right)$ is increasing in θ^*, it is clear that $\pi_H^U>\pi_L^U$.

Since $\pi_L^O<\pi_H^O$ and $\pi_H^U>\pi_L^U$, by Lemma 1, we have only two cases to consider: (1) $\pi_H^O<\pi_L^U$; (2) $\pi_H^U<\pi_L^O$.

Notably, since $X_1\left(\theta,\theta^*\right)$ is decreasing in θ, and $X_2\left(\theta,\theta^*\right)$ is increasing in θ, we have that $\pi_H^U>\pi_L^O$. Thus we rule out case (2).

Therefore, we have $\pi_L^O<\pi_H^O<\pi_L^U<\pi_H^U$, which proves the proposition. ■

7.4 Proof of Proposition 3

Proof. In the case where $S \leqslant c_H - \beta\delta\theta^* V$, we have that

$$\begin{aligned} U_S^S(\theta_L, \theta(r^*)) - U_T^S(\theta_L) &= -\pi[\int_{\beta\delta\phi(\theta^{\max}-\theta^*)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} \{\delta\phi(\theta^{\max} - \theta_L)W - k\}gdk] \\ &+ (1-\pi)[\int_{\beta\delta\theta_L V+S}^{\beta\delta\theta^* V+S} \{\delta\theta_L V - c\}fdc + fS\int_{\beta\delta\theta_L V+S}^{\beta\delta\theta^* V+S} dc] \end{aligned}$$

where the first term is independent of S, and the second term equals

$$\frac{1}{2}\left(1-\pi\right)\delta^2\beta f V^2\left(\theta_L - \theta^*\right)\left(\beta\theta^* - 2\theta_L + \beta\theta_L\right)$$

which is also independent of S. ■

7.5 Proof of Proposition 4

Proof. In the range $c_H - \beta\delta\theta^* V < S < c_H - \beta\delta\theta_L V$, define

$$\begin{aligned} & Y\left(r^*, \pi, S\right) \\ & \equiv U_S^S(\theta_L, \theta(r^*)) - U_T^S(\theta_L) \\ & = -\pi X_1 + (1-\pi)Y_2 \end{aligned}$$

where $X_1 \equiv \int_{\beta\delta\phi(\theta^{\max}-\theta^*)W}^{\beta\delta\phi(\theta^{\max}-\theta_L)W} \{\delta\phi(\theta^{\max} - \theta_L)W - k\}gdk > 0$ and $Y_2 \equiv \int_{\beta\delta\theta_L V+S}^{c_H}\{\delta\theta_L V - c\}fdc + fS\int_{\beta\delta\theta_L V+S}^{c_H} dc > 0$. We can therefore apply similar methods to those used in the proof of Proposition 1. Once again, it is clear that $Y\left(r^*, \pi, S\right)$ is continuous and decreasing in π for all $r^* \in [q, 1]$ as X_1 and Y_2 are both positive. Further, we have that $Y\left(r^*, 1, S\right) < 0$ and $Y\left(r^*, 0, S\right) > 0$ for all r^*. Thus there is a unique $\pi^{**}\left(r^*\right)$ such that $Y\left(r^*, \pi^{**}\left(r^*\right), S\right) = 0$, and $Y\left(r^*, \pi, S\right) > 0$ for all $\pi < \pi^{**}\left(r^*\right)$, and $Y\left(r^*, \pi, S\right) < 0$ for all $\pi > \pi^{**}\left(r^*\right)$ for all r^*.

Note also that Y_2 does not depend on θ^*. Thus we have, by the implicit function theorem,

$$\begin{aligned} \frac{d\pi^{**}}{d\theta^*} &= -\frac{\frac{dY(r^*,\pi,S)}{d\theta^*}}{\frac{dY(r^*,\pi,S)}{d\pi}} \\ &= \frac{-\pi\beta\delta\phi W^2[\delta\phi(\theta^{\max} - \theta_L) - \beta\delta\phi(\theta^{\max} - \theta^*)]g}{X_1 + Y_2} < 0 \end{aligned}$$

implying that $\pi^{**}\left(r^*\right)$ is decreasing in r^*, since θ^* is increasing in r^*.

We therefore have the following results.

(i) If $\pi > \pi^{**}\left(q\right)$, $Y\left(r^*, \pi, S\right) < 0$ for all r^*. Therefore, there is a unique PBE with $h^* = 1$.

(ii) If $\pi < \pi^{**}\left(1\right)$, we have that $Y\left(r^*, \pi, S\right) > 0$ for all r^*. Therefore, there is a unique PBE with $h^* = 0$.

(iii) If $\pi^{**}\left(1\right) < \pi < \pi^{**}\left(q\right)$, since $\pi^{**}\left(r^*\right)$ is a decreasing function, the inverse function $r^*\left(\pi\right)$ is also decreasing. Thus $Y\left(r^*, \pi, S\right)$ has the same sign of $r^*\left(\pi\right) - r^*$, implying that there are three PBEs: (a) $r^* = 1$ ($h^* = 1$) with $r^* > r^*\left(\pi\right)$, (b) $r^* = q$ ($h^* = 0$) with $r^* < r^*\left(\pi\right)$, and (c) a mixed one with h^* such that $Y\left(r^*\left(\pi\right), \pi, S\right) = 0$.

To complete the proof, let $\pi_H^S \equiv \pi^{**}\left(q\right)$, and $\pi_L^S \equiv \pi^{**}\left(1\right)$. ■

7.6 Proof of Proposition 6

Proof. Note that $W_T^S(c_H - \beta\delta\theta_L V) = qU_T^S(\theta_H) + (1-q)U_T^S(\theta_L)$ where

$$U_T^S(\theta_i) = \pi[\delta\theta_i W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\theta_i)W} \{\delta\phi(\theta^{\max}-\theta_i)W - k\}dG(k)] \\ + (1-\pi)\int_{c_L}^{c_H}\{\delta\theta_i V - c\}dF(c).$$

$W_S^S(0) = qU_S^S(\theta_H) + (1-q)U_S^S(\theta_L)$ where

$$U_S^S(\theta_i) = \pi\left(\delta\theta_i W + \int_{k_L}^{\beta\delta\phi(\theta^{\max}-\overline{\theta})W}\left(\delta\phi\left(\theta^{\max}-\theta_i\right)W - k\right)gdk\right) \\ + (1-\pi)\left(\int_{c_L}^{\beta\delta\overline{\theta}V}\left(\delta\theta_i V - c\right)fdc\right).$$

Thus, it is straightforward to get

$$W_T^S(c_H - \beta\delta\theta_L V) - W_S^S(0) \\ = \pi g\delta\phi W^2\left(1 - \frac{1}{2}\beta^2\delta\phi\right)q(1-q)(\theta_H-\theta_L)^2 \\ + (1-\pi)f\left(c_H - \beta\delta\overline{\theta}V\right)\left(\delta\overline{\theta}V - \frac{1}{2}\left(c_H + \beta\delta\overline{\theta}V\right)\right).$$

■