# Confidence Cycles

Thomas M. Eisenbach and Martin C. Schmalz*

June 2018

**Abstract**

We provide a model that rationalizes variations in confidence of rational agents, both in the time-series and the cross-section. Combining horizon-dependent risk aversion ("anxiety") and selective memory, we show that over- and under-confidence can arise in the Bayesian equilibrium of an intra-personal game. In the time-series, overconfidence is more prevalent when actual risk levels are high, while underconfidence occurs when risks are low. In the cross-section, more anxious agents are more prone to biased confidence and their beliefs fluctuate more, leading them to buy in booms and sell in crashes. Lastly, fluctuations in confidence can amplify boom-bust cycles.

**Keywords:** Overconfidence, dynamic inconsistency, biases, deception, risk taking
**JEL Classification:** C72, D03, D81, D83, G02

*Eisenbach: Federal Reserve Bank of New York, thomas.eisenbach@ny.frb.org; Schmalz: Stephen M. Ross School of Business, University of Michigan, CEPR, and ECGI, schmalz@umich.edu. A previous version of this paper was circulated under the title "Anxiety, Overconfidence, and Excessive Risk Taking." The views expressed in the paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. For helpful comments and discussion, we would like to thank Nick Barberis, Roland Bénabou, Markus Brunnermeier, Sylvain Chassang, Xavier Gabaix, Daniel Gottlieb, Robin Greenwood, Edoardo Grillo, Andrew Hertzberg (discussant), Daniel Kahneman, Camelia Kuhnen, David Laibson, Stephen Morris, Stefan Nagel, Paulo Natenzon, Wolfgang Pesendorfer, Bryan Routledge, Jacob Sagi, Andrei Shleifer, Paul Tetlock, and Wei Xiong, as well as audiences at Princeton, the University of Michigan, and the Miami Behavioral Finance Conference. Schmalz is grateful for generous financial support through an NTT Fellowship from the Mitsui Life Financial Center. Any errors are our own.

"It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so." — Mark Twain

"[I]t is hard for us with, and without being flippant, to even see a scenario within any kind of realm of reason that would see us losing $1 in any of those [CDO] transactions." — Joe Cassano, CFO of AIG's financial products unit, in August 2007 (AIG, 2007)

# 1 Introduction

Excessive risk taking due to biased beliefs is considered a contributing factor to the recent financial crisis (Barberis, 2013). Interestingly, several outsiders saw the risks building up while they were taken (e.g., Case and Shiller, 2003; Himmelberg et al., 2005; Smith and Smith, 2006; Shiller, 2007). At the same time, the decision makers themselves appeared to underestimate the risks – as evidenced both by public statements as well as personal investment decisions (Cheng et al., 2014; Foote et al., 2012). Why is it that particular people – specifically, those with access to the best information about risk levels and those making the risk-taking decisions – are systematically the most biased, especially at times when risks are high? Reinhart and Rogoff (2009) and Akerlof and Shiller (2010) advocate that fluctuations in confidence are a necessary part of realistic models of market dynamics and the business cycle. We ask: Do such fluctuations imply irrational decision makers, or can a neoclassical model generate such belief dynamics?

We develop a framework that generates overconfidence in equilibrium and that predicts how overconfidence varies in the time-series and in the cross-section. Among the different facets of overconfidence, we focus on overprecision of beliefs, first documented in the 1969 study of Alpert and Raiffa (1982) and popularized by Tversky and Kahneman (1974). A recent example in this tradition is the study of Ben-David et al. (2013) who show that when CFOs are asked to forecast the S&P 500, the realized one-year return falls within a CFO's 80-percent confidence interval only 36 percent of the time.[1]

[1]The data in Gennaioli et al. (2015) indicate that the same CFOs displaying overprecision of beliefs are well-calibrated on average about first moments, i.e. they are not over-optimistic.

Our model predicts this overprecision of beliefs when actual risk levels are high such as at the height of booms – but only among agents who face risk-taking decisions and thus have a material reason for "motivated beliefs." Conversely, our model generates underconfidence among such agents when actual risk levels are low. The model is thus able to explain why, which, and when decision makers in the recent financial crisis appear to have been the most overconfident, and why the confidence level of these agents fluctuates over the market cycle.

How can a rational Bayesian agent have mis-calibrated beliefs? Motivated by empirical evidence linking "anxiety" with belief manipulation, we show that an agent's overconfidence arises endogenously when combining a "demand" for overconfidence due to dynamically inconsistent risk preferences, and a need to make risk-taking decisions with a "supply" of overconfidence via motivated information processing. On the "demand" side, we stay within the standard paradigm of expected utility but assume that the agent exhibits *horizon-dependent risk aversion* as introduced by Eisenbach and Schmalz (2016). Specifically, we assume that risk aversion is higher for imminent than for distant risks, which we refer to as "anxiety." Such preferences have been documented experimentally as early as Jones and Johnson (1973) who have subjects participate in a simulated medical trial. Subjects are told that the probability of experiencing unpleasant side effects increases with the dosage they choose, as does the monetary compensation they receive – a classic risk-return tradeoff. Subjects choose smaller dosages – implying higher risk aversion – when they are to take the drug immediately than when they are to take it the next day.[2]

The key implication of horizon-dependent risk aversion is that agents would like to take more risks in the future but end up backing out as the risks approach. This dynamic inconsistency implies that an agent's earlier self would like to manipulate a later self to take more risk than the later self will be inclined to take on its own. We investigate whether belief manipulation can help align the later self's actions with the earlier self's preferences in situations in which external commitment devices are either unavailable, relatively costly, or less attractive than overconfidence for other

[2]In Section 2, we discuss numerous studies which find the same pattern using standard experimental economics designs with monetary payoffs and rigorous elicitation of risk aversion (Onculer, 2000; Sagristano et al., 2002; Noussair and Wu, 2006; Baucells and Heukamp, 2010; Coble and Lusk, 2010; Abdellaoui et al., 2011).

exogenous reasons.[3]

We find that the internal disagreement about risk taking implies that the earlier self wants to convince the later self that risks are lower than they actually are. If successful, such self-manipulation leads to overly precise beliefs of the later self. As a result, the later self makes decisions that are riskier and more aligned with the earlier self's preferences, compared to the decisions it would have taken under well-calibrated beliefs. The question is then how such a demand for overconfidence can be met and sustained in equilibrium.

The "supply" side of the model addresses the question of how an early self can affect the beliefs of a later self. We take a conservative approach by assuming the belief manipulation happens indirectly through selective memory. The idea that humans display a tendency to selectively forget goes back at least to Freud (1904). Later contributions document that individuals tend to recall their successes more than their failures, and have self-servingly biased recollections of their past performances (Korner, 1950; Crary, 1966; Mischel et al., 1976). Selective memory has since been used in many contributions to the economics literature, as reviewed below. Whereas we allow the earlier self to strategically forget signals it observes, we impose the constraint that the later self is aware of the earlier self's incentive to manipulate and that it processes information in a fully Bayesian way.

The agent's temporal selves therefore interact in a standard sequential game under incomplete information, in which we solve for the perfect Bayesian equilibrium. Analogous to, e.g. Fudenberg and Levine (2006), however, the intra-personal game is a modeling device and not a description of the agent's mental processes. Specifically, we do not mean to imply that the agent is *consciously* aware of the strategic interaction of her temporal selves.

We assume that the different selves each have private, decision-relevant information. The fact that an earlier self's preferred course of action depends on information only available to a later self implies that belief distortions set by the earlier self can lead to *excessive risk taking*, even as judged from the perspective of the earlier, less risk averse, self. We can now combine "demand" and "supply" to solve for the equi-

[3]We show in Eisenbach and Schmalz (2016) that dynamic inconsistency for intra-temporal risk trade-offs is orthogonal to dynamic inconsistency for inter-temporal consumption trade-offs due to non-geometric discounting as in Strotz (1955), Phelps and Pollak (1968), and Laibson (1997). We also discuss the use of external commitment devices to deal with "anxiety."

librium level of overconfidence of the intra-personal game. The equilibrium trades off the benefit of overconfidence which is additional, desirable, risk taking in some states of the world – thus mitigating the effect of "anxiety" – against the costs of overconfidence in terms of excessive risk taking in other states of the world.

Our model generates a rich set of predictions. First, we show that agents may appear overconfident or underconfident to an outside observer, depending on the true state of the world. Specifically, agents displaying overconfidence are observed when actual risks are high and vice versa. We interpret the prediction that confidence positively covaries with current risk levels as consistent with the view of Reinhart and Rogoff (2009) and Akerlof and Shiller (2010) that overconfidence is high at the peak of booms whereas underconfidence prevails in the trough of crises.[4]

Second, the model predicts that in the cross-section, agents in systematically high-risk environments are more overconfident than agents in systematically low-risk environments. This result is consistent with the high levels of overconfidence observed, e.g. for CFOs (Ben-David et al., 2013) and the low levels observed, e.g. for auditors (Tomassini et al., 1982). Third, the agent in our model is only overconfident about immediate uncertainty, but has no reason for overconfidence at longer horizons. This prediction matches the details of the evidence in Ben-David et al. (2013), where CFOs are overconfident only at short horizons and are well calibrated at longer horizons.

Finally, agents with stronger "anxiety" display stronger biases in confidence. This implies that the pro-cyclicality of confidence is strongest for the most anxious agents leading them to overtrading, buying high and selling low. This prediction is consistent with the empirical evidence in Lo et al. (2005), which indicates that more emotional traders generate lower profits.

The model can also generate amplification effects if applied to settings that take overconfidence as an input to generate speculative behavior (Harrison and Kreps, 1978; Scheinkman and Xiong, 2003). Overconfident traders have a greater demand for risk, which leads to more overconfidence and thus generates positive feedback. Conversely, crises trigger an underconfidence spiral. Taken together, the mechanism we propose is a natural amplifier of economic fluctuations.

The model can serve as a reduced-form metaphor for the choice of the agent's in-

[4] Cohn et al. (2015) provide evidence from the lab for the role of "fear" in causing confidence cycles of financial professionals.

formation and communication environment, for example, in an organizational framework. For example, executives may discourage subordinates from warning them about risks, explicitly or implicitly through the choice of reward and punishment systems ("killing the messenger"). This prediction resonates with the marginalization of risk managers in the financial sector before the recent crisis (Flannery et al., 2012). However, the model can also be interpreted more literally as self-manipulation with alcohol or other drugs, which can take place ex ante consciously or subconsciously. For example, Steptoe and Fidler (1987) find that among professional musicians, the use of sedatives to cope with performance anxiety is higher in those with high levels of anxiety than those with low levels.

The remainder of this section discusses the related literature. In Section 2, we review existing experimental evidence that supports our assumption that temporal distance affects risk-taking behavior. We analyze the model in Section 3. Section 4 suggests interpretations and discusses applications of the model. We conclude in Section 5. All proofs are in the appendix.

**Related Literature.** Our model sheds light on how overconfidence can be formed and sustained in equilibrium, making it "[p]erhaps the most robust finding in the psychology of judgment" (De Bondt and Thaler, 1994). In economics and finance, overconfidence has often been used as an *assumption* to (formally or informally) explain the behavior of speculative investors (e.g. Scheinkman and Xiong, 2003), the ventures of entrepreneurs (e.g. Bernardo and Welch, 2001), corporate decision making (e.g. Malmendier and Tate, 2005; Gervais et al., 2011) or the pricing schedules of firms (e.g. Grubb, 2009).[5]

Research documenting overconfidence (in the strict sense of an overprecision of beliefs) goes back to Alpert and Raiffa (1982) who ask 800 Harvard MBAs to provide percentiles of their subjective uncertainty about 10 quantities, e.g. "The number of 'Physicians and Surgeons' listed in the 1968 Yellow Pages of the phone directory for Boston and vicinity." On average, the true value falls within a student's 50-percent confidence interval only 34 percent of the time and within the 98-percent confidence interval only 66 percent of the time. Subjects are also overconfident when asked

[5]See also, e.g. Harrison and Kreps (1978), Cooper et al. (1988), Daniel et al. (1998), Odean (1998, 1999), Barber and Odean (2001), Shefrin (2001), Koellinger et al. (2007), Hackbarth (2008), and Deaves et al. (2009).

about objectively uncertain random variables, e.g. the Consumer Price Index (Brown, 1973). In the field, overconfidence has been systematically documented among finance practitioners as early as Staël von Holstein (1972) and more recently, e.g. by Deaves et al. (2010), Inoue et al. (2012), Ben-David et al. (2013) or Glaser et al. (2013).

The literature sometimes uses the term "overconfidence" to describe other biases besides "overprecision", including agents overestimating their performance or ability in absolute terms ("overestimation") or relative to others ("overplacement"). Work such as Klayman et al. (1999) shows that overconfidence in the form of overly narrow subjective confidence intervals is the most pronounced phenomenon the literature has labeled "overconfidence." Moore and Healy (2008) systematically disambiguate various alternative interpretations of overconfidence and confirm that overprecision of beliefs is empirically the most robust finding of overconfidence in the literature.[6] Lichtenstein et al. (1982) survey the classic literature; for a recent review of the literature on overconfidence, overoptimism, and related concepts, see Grubb (2015).

Overconfidence as overprecision has been exogenously assumed in existing models of overconfidence, e.g. by Scheinkman and Xiong (2003) where agents treat their signals as more precise than they truly are. In contrast, we microfound why overconfidence arises endogenously. Our model of intra-personal conflict and equilibrium manipulation follows the spirit of Bénabou and Tirole (2002), who show that overestimating personal ability can be a self-motivating strategy for an agent with imperfect willpower, modeled by way of quasi-hyperbolic discounting. Such an agent has dynamically inconsistent preferences with respect to intertemporal trade-offs, i.e. when comparing current costs of effort with future benefits. Carrillo and Mariotti (2000) also study an agent with dynamically inconsistent consumption preferences and show how incomplete learning can arise to effectively commit. In contrast to these approaches, our model generates underestimation of risk, based on dynamically inconsistent preferences with respect to intra-temporal risk trade-offs.[7] Conceptually, our setup falls into the category of "Bayesian persuasion" between a sender

[6] Benoît and Dubra (2011) even show that evidence on better- and worse-than-average effects is *not* inconsistent with rational Bayesian information processing. On the distinction of different typed of overconfidence, see also Moore and Schatz (2017).

[7] Besides intra-personal conflict, overconfidence can also be a useful device to affect tradeoffs in inter-personal conflicts such as in a standard Cournot duopoly (Kyle and Wang, 1997).

and receiver as characterized by Kamenica and Gentzkow (2011); in our case, sender and receiver are two temporal selves of a single agent.

Our assumption of horizon-dependent risk aversion is based on extensive experimental evidence discussed in detail in Section 2 (Jones and Johnson, 1973; Onculer, 2000; Sagristano et al., 2002; Noussair and Wu, 2006; Baucells and Heukamp, 2010; Coble and Lusk, 2010; Abdellaoui et al., 2011). As an assumption about preferences, what we call "anxiety" is orthogonal to the belief-based concept of "cold feet" developed in Epstein and Kopylov (2007). It is more closely related to Epstein (2008), who – in contrast to our work and the experimental evidence we discuss below – assumes risk aversion is higher for distant risks than for imminent risks. Eisenbach and Schmalz (2016) provide a more detailed distinction of horizon-dependent risk aversion from other theories such as time-changing risk aversion, preference for the timing of resolution of uncertainty, and dynamically inconsistent time preferences.

The use of selective memory as a mechanism that generates belief distortions has precedents, for example, in Gennaioli and Shleifer (2010) or Bordalo et al. (2017) (see also Hirshleifer and Welch, 2002). Ericson (2011) also features overconfidence and memory but in the sense that agents overestimate the quality of their memory. Barberis et al. (2015) generate rich implications when beliefs are formed by extrapolation of experiences of the recent past. Greenwood and Nagel (2009), and Malmendier and Nagel (2011, 2016) provide evidence that individual-specific information choice affects expectations and investment behavior. Also related is Compte and Postlewaite (2004), who study biases in information processing motivated by the effect of emotions on performance: to optimize performance, their agents have to manage their emotions, which they achieve through selective information processing. Instead of selective memory, van den Steen (2004, 2011) shows that agents who start with heterogeneous priors but are fully rational can end up with overconfident posteriors. In Gervais and Odean (2001), a bias in learning about ability generates overconfidence over time.

Finally, our model is cast within the standard expected utility framework. This feature contrasts with models that allow the prize space to include mental states (e.g. Caplin and Leahy, 2001; Köszegi, 2006), information entering the utility function directly (e.g. Pagel, 2017), preferences over information due to disappointment effects (e.g. Gul, 1991; Andries and Haddad, 2017), as well as utility from anticipation (e.g. Brunnermeier and Parker, 2006; Brunnermeier et al., 2007, 2017) or memory (e.g. Gilboa et al., 2015). Also, the preferences of our agents don't change over time in anticipated or unanticipated ways (as in Loewenstein et al., 2003); instead, our agents deal with dynamically inconsistent risk preferences at each point of time. See Brunnermeier and Julliard (2008), and Piazzesi and Schneider (2008) on other belief biases that have been linked to the housing frenzy in the run-up to the financial crisis.

## 2 Evidence for Horizon-Dependent Risk Aversion

Temporal distance has been documented to affect risk-taking behavior in field and laboratory experiments. In particular, subjects tend to be more risk averse when a risk is temporally close than when it is distant, both in across-subject and within-subject studies. In this section, we discuss several studies besides the one by Jones and Johnson (1973) mentioned in the introduction. Specifically, we highlight studies with purely monetary payoffs and rigorous elicitation of risk aversion.[8]

Experimental economics widely uses the protocol of Holt and Laury (2002) to elicit risk aversion. Subjects are presented with a list of choices between two binary lotteries. The first lottery always has two intermediate prizes, for example, ($10.00, $8.00), whereas the second lottery always has a high and a low prize, for example, ($19.25, $0.50). The experimenter then changes the respective probabilities of the two prizes, varying from $(0.1, 0.9)$ to $(0.9, 0.1)$. As the probability mass shifts from the second prize to the first prize of both lotteries, the second lottery becomes increasingly attractive compared to the first lottery. Subjects are asked to pick one of two lotteries for each of the probability distributions. The probability distribution at which a subject switches from the "safe" lottery to the "risky" lottery is a proxy for the subject's risk aversion. Noussair and Wu (2006) use this protocol for a within-subjects design with real payoffs, having each subject make choices for resolution and payout that occur immediately and also for risks and payouts that occur three months later. The study finds that more than one third of subjects are more risk averse for the present than for the future. Coble and Lusk (2010) use the protocol for an across-subjects

[8]As another example, Welch (1999) documents preference reversals caused by stage fright. He finds that 67% of subjects who agree to tell a joke in front of a class the following week in exchange for $1 "chicken out" when the moment of truth arrives. By contrast, none of those who decline initially change their mind.

design and find the same pattern, with average risk aversion increasing with the temporal proximity of the risk.

In a different type of experiment, Baucells and Heukamp (2010) let subjects choose between two binary lotteries, a "safer" and a "riskier" one. Different treatments vary the delay until the lotteries are resolved and paid out. The study finds that more subjects choose the riskier lottery as the delay increases. Sagristano, Trope, and Liberman (2002) also have subjects choose between two lotteries, and find the same effect of temporal proximity.

Finally, some studies elicit risk aversion by asking subjects for their certainty equivalents for different lotteries; a lower certainty equivalent corresponds to higher risk aversion. In Onculer (2000), subjects state their certainty equivalent for a lottery to be resolved and paid immediately, as well as for the same lottery to be resolved and paid in the future. The study finds that subjects state significantly lower certainty equivalents for the immediate lottery than for the future lottery. Abdellaoui, Diecidue, and Onculer (2011) conduct a similar study with real payoffs, and find equivalent results.

## 3 Model of Overconfidence

This section lays out the model of endogenous overconfidence. We first describe the "demand" for overconfidence based on horizon-dependent risk aversion. Such preferences imply a dynamic inconsistency: the agent would like to take more risks in the future but prefers to take less risk in the present. We find that the agent's present self would like to make the future self overconfident to induce greater risk-taking. Overconfidence can thus serve as a substitute for external commitment devices.

We then describe the environment our agent faces, including risk trade-offs as well as information available to selves at different times and the resulting possibilities for earlier selves to affect the beliefs of later selves, that is, the "supply" of overconfidence. Finally, we combine "demand" and "supply" in solving for the equilibrium level of overconfidence in the intra-personal game played among the agent's temporal selves.

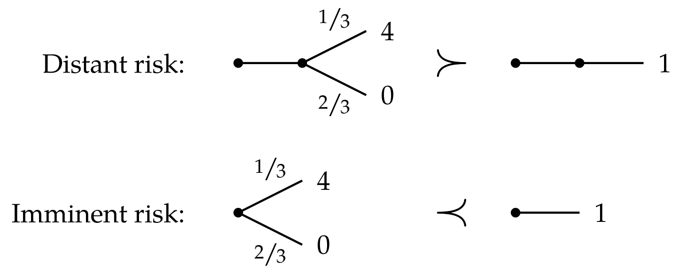


Figure 1: "Anxious" preferences with horizon-dependent risk aversion.

### 3.1 Preferences

Figure 1 gives a stylized example of the horizon-dependent choice behavior documented experimentally. The agent has to choose between a risky alternative – receiving 4 with probability 1/3 – or a safe alternative – receiving 1 for sure. Because the expected value of the risky alternative is greater than 1, a risk-averse agent may prefer either of the two alternatives, depending on the level of risk aversion. The experimental evidence points to agents who prefer the risky alternative when the risk is temporally distant, but prefer the safe alternative when the risk is temporally close, as indicated in Figure 1. More generally, consider a typical risk-reward trade-off given by two lotteries $\tilde{x}$ and $\tilde{y}$, where $\tilde{x}$ has "higher risk" but also "higher reward" than $\tilde{y}$ if we assume $\tilde{x} = \tilde{y} + \tilde{\varepsilon} + \mu$ with $\tilde{\varepsilon}$ a mean-zero lottery independent of $\tilde{y}$ and $\mu$ a constant. To capture the experimental evidence of agents who prefer the risky lottery $\tilde{x}$ to $\tilde{y}$ if both are delayed, but prefer the safe lottery $\tilde{y}$ to $\tilde{x}$ if both are immediate, we use a two-period setup $t = 0, 1$ and assume a utility specification $U_t$ given by

$$U_0 = \mathbb{E}[v(c_0) + \delta u(c_1)] \quad \text{and} \quad U_1 = \mathbb{E}[v(c_1)],$$

where $\mathbb{E}$ is the expectations operator and $\delta \leq 1$ is a discount factor. More importantly, $v$ and $u$ are von Neumann-Morgenstern utility indexes that depend on whether a risk is imminent or distant. To generate the same choice behavior as documented experimentally, the utility specification has to satisfy the following two conditions for the lotteries $\tilde{x}, \tilde{y}$:

$$\begin{aligned} &\text{For distant lotteries:} \quad \mathbb{E}\left[\delta u(\tilde{x})\right] > \mathbb{E}\left[\delta u(\tilde{y})\right] \\ &\text{For imminent lotteries:} \quad \mathbb{E}\left[v(\tilde{x})\right] < \mathbb{E}\left[v(\tilde{y})\right]. \end{aligned}$$

Given the definitions of $\tilde{x}$ and $\tilde{y}$, these conditions can be satisfied only with $v$ more risk averse than $u$:

$$-\frac{v''(c)}{v'(c)} \geq -\frac{u''(c)}{u'(c)} \quad \text{for all } c.$$

Note that the discount factors $\delta$ play no role in the two conditions above.[9] The above derivation illustrates that intra-temporal risk trade-offs and inter-temporal consumption trade-offs are conceptually very different; the experimental evidence can therefore *not* be addressed by relaxing the standard assumption of geometric discounting.

As an example, let $v(c) = \sqrt{c}$ and $u(c) = c$ and set $\delta = 1$. Then the agent is risk averse with respect to current uncertainty and risk neutral with respect to future uncertainty. Now consider the following lotteries for $\tilde{x}, \tilde{y}$:

$$\tilde{x} = \begin{cases} 4 & \text{with prob. } \alpha \\ 0 & \text{with prob. } 1-\alpha \end{cases} \qquad \text{and} \qquad \tilde{y} = 1.$$

Then $v$ prefers the risky $\tilde{x}$ to the safe $\tilde{y}$ if $\alpha > 1/2$, whereas $u$ prefers $\tilde{x}$ to $\tilde{y}$ if $\alpha > 1/4$ and disagreement exists between $v$ and $u$ for all $\alpha \in (1/4, 1/2)$ as illustrated in Figure 1. In particular, suppose $\alpha = 1/3$ and that the lotteries are resolved and paid out in period 1. Then the agent will choose the safe option $\tilde{y}$ in period 1 but would prefer to commit to the risky option $\tilde{x}$ in the initial period 0.

## 3.2 Environment

Formally, we assume that outright commitment devices are relatively costly or not available at all. Instead, the agent's earlier self may try to distort the later self's beliefs to manipulate the later self's decisions. In particular, the earlier self would like to convince the later self that risks are lower than they actually are. Such a conviction would lead the later self to make riskier decisions that are more in line with the earlier self's preferences. The question is whether such belief manipulation can be achieved in equilibrium.

To make the problem interesting, we add two important elements. First, we as-

[9] The fact that discount factors do not matter in our context is the reason we employ a simple time-separable framework, which would otherwise lead to a conflation of time and risk preferences. Andries et al. (2018) present a utility function that allows for a clean separation in a fully dynamic setting.

sume the agent is rational so the later self is fully aware of the earlier self's incentives to manipulate. The two selves therefore interact in a strategic way, and we have to study the equilibrium of the agent's intra-personal game. Second, we assume the later self has access to additional information that is decision relevant also from the perspective of the earlier self. With this assumption, manipulating the beliefs is costly because doing so may lead to sub-optimal decisions by the later self.

To analyze the intra-personal manipulation game, we use a setting similar to models studying belief manipulation with $\beta$-$\delta$ time inconsistency (Carrillo and Mariotti, 2000; Bénabou and Tirole, 2002). There are two periods $t = 0, 1$. In period 1, the agent faces a risk-reward trade-off, having to choose between a risky and a safe alternative. The risky alternative is given by a lottery with random payoff $x$ characterized by its distribution function $G_\theta$, where $\theta \in \{H, L\}$ denotes a state of the world that determines how risky the lottery is. We assume $G_H$ is a mean-preserving spread of $G_L$, so the risky alternative is unambiguously riskier in state $H$ than in state $L$. The ex-ante probability of the high-risk state $H$ is given by $\pi$. The safe alternative, on the other hand, is given by a constant payoff $a$.

When facing the decision in period 1, the agent evaluates the risk using utility $v$, and therefore wants to take the risky alternative whenever the expected utility is higher than that of the safe alternative:

$$\int_{-\infty}^{\infty} v(x)\, dG_\theta(x) > v(a)$$

Denoting the certainty equivalent of $G_\theta$ given the utility function $v$ by $c_v^\theta$, this condition can be rewritten as

$$c_v^\theta > a.$$

The agent wants to take the risky alternative whenever its certainty equivalent $c_v^\theta$ is greater than the safe alternative $a$.

When thinking about the decision ahead of time (in period 0), the agent evaluates the risk using utility $u$ and therefore wants the future self to take the risky alternative whenever

$$\int_{-\infty}^{\infty} \delta u(x)\, dG_\theta(x) > \delta u(a) \quad \Leftrightarrow \quad c_u^\theta > a.$$

As in the simple numerical example above, we have potential disagreement between

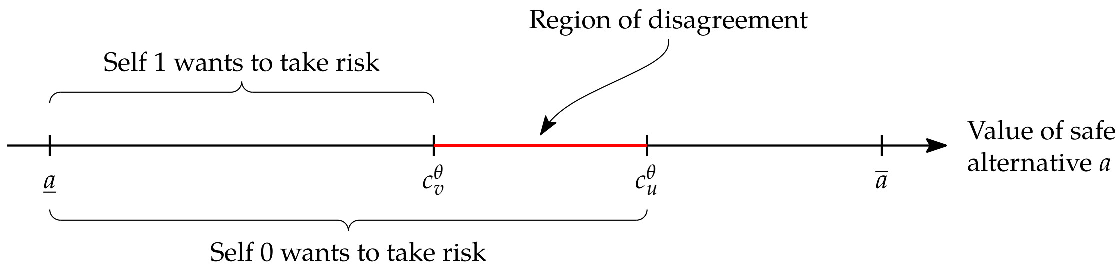


Figure 2: Disagreement between the agent's selves under full information.

self 0 and self 1.

**Lemma 1.** *Because $v$ is more risk averse than $u$, we have $c_u^\theta > c_v^\theta$ for both $\theta \in \{H, L\}$ so the agent in period 0 (self 0) and the agent in period 1 (self 1) will disagree about the right course of action whenever $a \in \left[c_v^\theta, c_u^\theta\right]$.*

Figure 2 illustrates the disagreement between the agent's temporal selves under complete information. If the safe alternative is sufficiently unappealing – the value of $a$ is very low – both selves prefer to take the risky alternative and vice versa if the value of $a$ is very high. For intermediate values of $a$, however, self 0 prefers the risky alternative whereas self 1 prefers the safe alternative.

The problem becomes interesting when information is incomplete. We therefore assume that both of the agent's selves have partial, decision-relevant information. Specifically, the state of the world $\theta$ is revealed to the agent at the beginning of period 0 in the form of a perfectly informative "red flag" warning signal $s$ if the state is high risk:

$$s = \begin{cases} R & \text{if } \theta = H \\ \emptyset & \text{if } \theta = L \end{cases} .$$

The payoff of the safe alternative $a$, however, is not known to the agent until period 1. Self 0 only knows the prior distribution $F$ on $[\underline{a}, \overline{a}]$, but self 1 observes the realized value of $a$.

The only way that we allow self 0 to affect the beliefs of self 1 is through its treatment of the signal $s$. In particular, we assume imperfect memory: if self 0 observes a red flag, $s = R$, then the agent forgets the signal with probability $\varphi \in [0, 1]$, such that self 1's recollection of the signal is $\hat{s} = \emptyset$ with probability $\varphi$ and $\hat{s} = R$ with probability $1-\varphi$. The probability of forgetting the signal $s$ is under the control of self 0 but out of the control of self 1.

Selective memory is a well-established phenomenon in the psychology literature, as reviewed in the introduction. Note that we allow self 0 to affect the beliefs of self 1 only indirectly through forgetting, but don't allow self 0 to directly influence self 1's beliefs. In addition, we allow only the neglect of information that exists and not any fabrication of information that does not exist. Finally and most importantly, we assume self 1 is aware of self 0's incentive to manipulate, and processes information in a fully Bayesian way. In sum, we consider the belief distortion allowed to the agent to be a fairly weak assumption.[10]

These assumptions don't have to be interpreted literally, in the sense of imagining that the individual can directly and consciously suppress memories. Indeed, we find it more difficult to imagine a conscious decision to forget. By contrast, realizing that remembering often takes a conscious effort may be more intuitive. Not making such an effort, which likely results in forgetting, can therefore be viewed as a conscious choice. That said, our model is not necessarily meant to describe the agent's actual mental process. Our model is equally consistent with a Freudian view where some memories get buried in the unconscious, with some probability of reappearance. Whichever the preferred interpretation, as most standard models in economics, the model presented here simply provides a framework with an equilibrium that represents observed behavior and beliefs, and yields insights into the underlying tradeoffs. Section 4 below discusses how our reduced-form structure of belief manipulation can be interpreted in practice as a choice of social environments, information systems, or as self-manipulation with alcohol and drugs.

## 3.3 Intra-personal Game

Given our setup, self 0 and self 1 are playing a sequential intra-personal game with incomplete information. First self 0 chooses the forgetting probability $\varphi$, taking into account self 1's behavior, and then self 1 decides between the risky and the safe al-

[10]Formally, our assumptions about signals and memory are without loss of generality. As shown by Kamenica and Gentzkow (2011), in environments such as ours an analogue of the "revelation principle" applies and we can effectively restrict the number of messages available to self 0 to the number of states of the world $\theta$.

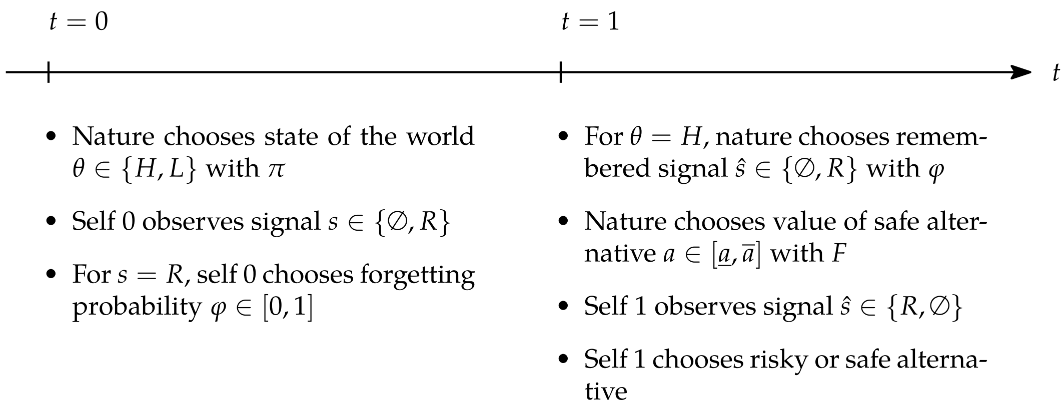


Figure 3: Timeline of intra-personal game.

ternative, taking into account self 0's behavior. Figure 3 illustrates the timeline of the intra-personal game; Figure 8 in the appendix gives an extensive-form representation. We are interested in the perfect Bayesian equilibria of this intra-personal game.

First, we derive self 1's best response in period 1, taking as given an expected forgetting probability $\varphi^e$. If self 1 remembers seeing a red flag, $\hat{s} = R$, she knows the state of the world is high risk and chooses the risky alternative if $a < c_v^H$. If self 1 doesn't remember seeing a red flag, $\hat{s} = \emptyset$, she assigns a posterior probability to the state of the world being high risk given by:

$$\hat{\pi}(\varphi^e) = \frac{\pi \varphi^e}{\pi \varphi^e + 1 - \pi}.$$

Naturally, the posterior probability of being in the high-risk state is increasing in the probability of forgetting a red flag, $\hat{\pi}'(\varphi^e) > 0$.

Self 1's posterior distribution for the risky payoff is then given by:

$$\hat{G}(x|\varphi^e) = \hat{\pi}(\varphi^e)\, G_H(x) + \big(1 - \hat{\pi}(\varphi^e)\big)\, G_L(x).$$

Since $G_H$ is a mean-preserving spread of $\hat{G}$, self 1 will be *overconfident* if the true state of the world is high-risk. Conversely, since $\hat{G}$ is a mean-preserving spread of $G_L$, self 1 will be *underconfident* if the true state of the world is low-risk. Given the Bayesian posterior $\hat{G}(x|\varphi^e)$, self 1 chooses the risky alternative if $a < c_v(\varphi^e)$, where $c_v(\varphi^e)$ is the certainty equivalent of the risky alternative given $\varphi^e$, implicitly defined by

$$\int_{-\infty}^{\infty} v(x)\, d\hat{G}(x|\varphi^e) = v\big(c_v(\varphi^e)\big).$$

Next, we derive self 0's best response in $t = 0$, taking as given self 1's behavior for an expected $\varphi^e$. If self 0 receives a warning signal and chooses a forgetting probability $\varphi$, her expected utility is

$$\begin{aligned}\mathbb{E}U_0(\varphi \mid \varphi^e) = (1-\varphi)\left[F\big(c_v^H\big)\int_{-\infty}^{\infty} \delta u(x)\, dG_H(x) + \int_{c_v^H}^{\overline{a}} \delta u(a)\, dF(a)\right] \\ + \varphi\left[F\big(c_v(\varphi^e)\big)\int_{-\infty}^{\infty} \delta u(x)\, dG_H(x) + \int_{c_v(\varphi^e)}^{\overline{a}} \delta u(a)\, dF(a)\right].\end{aligned}$$

In the first line of $\mathbb{E}U_0$, with probability $1-\varphi$, the agent remembers the warning signal in period 1 and uses the certainty equivalent $c_v^H$ as the threshold; then she chooses the risky alternative for payoffs of the safe alternative below the threshold, $a \in \big[\underline{a}, c_v^H\big)$, and chooses the safe alternative for payoffs above the threshold, $a \in \big[c_v^H, \overline{a}\big]$. In the second line of $\mathbb{E}U_0$, with probability $\varphi$, the agent forgets the warning signal and uses the certainty equivalent $c_v(\varphi^e)$ as the threshold, choosing the risky alternative for $a \in \big[\underline{a}, c_v(\varphi^e)\big)$ and the safe alternative for $a \in \big[c_v(\varphi^e), \overline{a}\big]$.[11]

**Definition.** A perfect Bayesian equilibrium of the agent's intra-personal game is a pair of strategies $(\sigma_0, \sigma_1)$ for the agent's temporal selves 0 and 1, respectively, and a belief $\hat{\omega}$ for self 1 such that:

1. Self 0's strategy $\sigma_0 = \varphi^*$ maximizes $\mathbb{E}U_0(\varphi \mid \varphi^e)$ for $\varphi^e = \varphi^*$.

2. Self 1's strategy $\sigma_1 : \{\emptyset, R\} \times [\underline{a}, \overline{a}] \to \{\text{risky}, \text{safe}\}$ is given by cutoffs $c_v(\varphi^*)$ for $\hat{s} = \emptyset$ and $c_v^H$ for $\hat{s} = R$.

3. Self 1's belief is Bayesian, i.e., $\hat{\omega} = \hat{\pi}(\varphi^*)$ for $\hat{s} = \emptyset$ and $\hat{\omega} = 1$ for $\hat{s} = R$.

[11] Note that it doesn't matter whether self 0 chooses $\varphi$ before or after the risk state $\theta$ is realized as the expected utility before the realization is simply given by

$$\pi\, \mathbb{E}U_0(\varphi \mid \varphi^e) + (1-\pi)\left[F\big(c_v(\varphi^e)\big)\int_{-\infty}^{\infty} \delta u(x)\, dG_L(x) + \int_{c_v(\varphi^e)}^{\overline{a}} \delta u(a)\, dF(a)\right],$$

where the second part doesn't depend on $\varphi$ and is therefore irrelevant for the maximization.

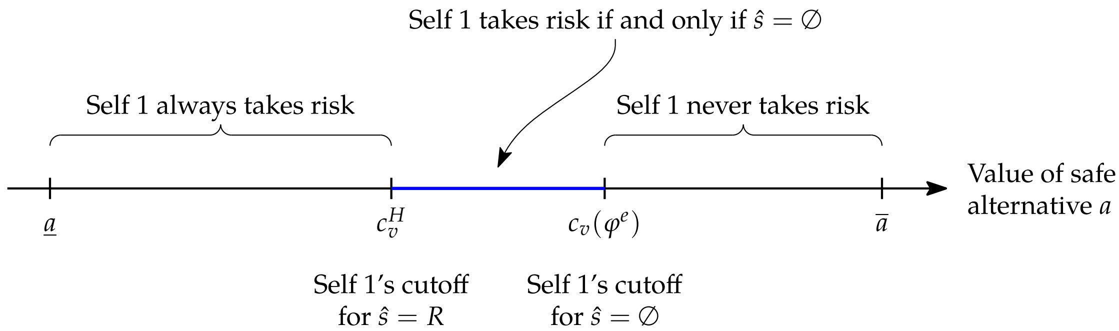


Figure 4: Effect of memory $\hat{s}$ on self 1's decision.

We denote by $D(\varphi^e)$ the derivative of self 0's expected utility with respect to her choice variable $\varphi$ conditional on the value $\varphi^e$ expected by self 1. This marginal benefit of forgetting is given by

$$D(\varphi^e) := \delta \int_{c_v^H}^{c_v(\varphi^e)} \left( \int_{-\infty}^{\infty} u(x)\, dG_H(x) - u(a) \right) dF(a)\,.$$

The expression for $D(\varphi^e)$ has a natural interpretation. As illustrated in Figure 4, the remembered signal $\hat{s}$ affects self 1's decision only for realizations of the safe alternative $a \in \left[c_v^H, c_v(\varphi^e)\right]$. In this interval, self 1 chooses the safe alternative whenever she remembers seeing a red flag, and the risky alternative otherwise. The effect on self 0's expected utility of forgetting the warning signal more often is exactly the difference in utility from the risky action compared to the safe action for the values of $a$ where the decision is affected.

Three types of perfect Bayesian equilibria can exist in this intra-personal game, characterized by the equilibrium forgetting probability $\varphi^*$:

**Honesty:** If $D(0) \leq 0$, an equilibrium with $\varphi^* = 0$ exists. In this equilibrium, the agent never ignores red flags and doesn't influence her future self's beliefs.

**Overconfidence:** If $D(1) \geq 0$, an equilibrium with $\varphi^* = 1$ exists. In this equilibrium, the agent always ignores red flags and makes her future self maximally overconfident.

**Mixed:** If $D(\bar{\varphi}) = 0$ for some $\bar{\varphi} \in (0, 1)$, an equilibrium with $\varphi^* = \bar{\varphi}$ exists. In this equilibrium, the agent plays a mixed strategy, ignoring the red flag with

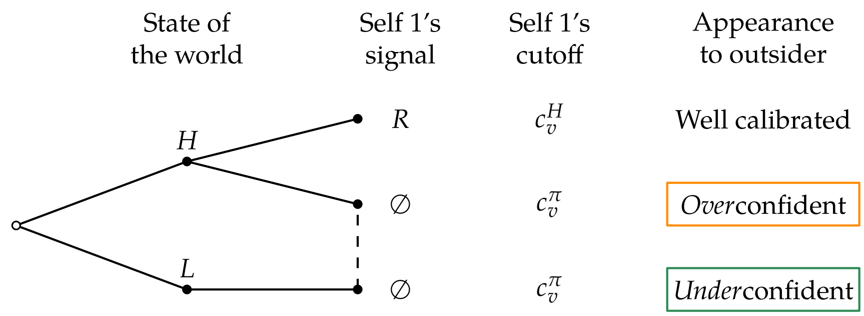


Figure 5: Apparent over- and underconfidence to an outside observer.

probability $\bar{\varphi}$, and makes her future self partially overconfident.

We have the following result on the existence of equilibria.

**Proposition 1.** *One of the extreme equilibria always exists: either the honesty equilibrium or the overconfidence equilibrium or both. If both extreme equilibria exist, a mixed equilibrium also exists.*

Since the mixed equilibrium is not stable (see the proof of Proposition 1), we focus on the extreme equilibria of overconfidence and honesty.

## 3.4 Apparent Over- and Underconfidence

In the overconfidence equilibrium $\varphi^* = 1$, an outside observer will find the agent using one of two cutoffs, as illustrated in Figure 5. If the state of the world is high risk and the red flag signal was remembered, the agent's self 1 is using the cutoff $c_v^H$ when deciding between the risky and the safe alternative. To an outside observer who knows the agent's preferences $v$ and the true state of the world $H$, the agent therefore appears to have well calibrated beliefs.

By contrast, if (i) the state is $H$ but the warning signal was forgotten or (ii) the state is $L$, the agent's self 1 is using the cutoff $c_v(1) = c_v^\pi$, i.e. with the prior probability of high risk $\pi$. While the agent herself cannot distinguish between (i) and (ii), an outside observer who knows the state can and will interpret the agent's behavior accordingly. In case (i), the agent using $c_v^\pi$ appears *over*confident to the outside world, because – based on her preference $v$ – she is expected to use $c_v^H < c_v(\varphi^*)$. In case (ii), the agent using $c_v^\pi$ appears *under*confident, because she is expected to use $c_v^L > c_v^\pi$. We can now state the following corollary.

**Corollary 1.** *In the overconfidence equilibrium, agents can appear to be ex-post over- or underconfident. Moreover, the appearance is linked to the true state of the world:*

- *Agents displaying overconfidence can only be observed if the state is high risk.*
- *Agents displaying underconfidence can only be observed if the state is low risk.*

A benefit of the fully rational framework and the perfect Bayesian equilibrium analysis is that we can interpret the results of Corollary 1 in the time series. As long as the realization of the state of the world $\theta$ is i.i.d. so that learning has no role, we can imagine the simple two-period setup being repeated in sequence.

Empirically, Reinhart and Rogoff (2009) as well as Akerlof and Shiller (2010) argue that at the peaks of booms – when actual risks are high – overconfidence widespread, whereas underconfidence is common in the trough of crises – when actual risk is low. Our model predicts exactly that: agents appear overconfident when risks are high and underconfident when risks are low. In fact, also the CFOs in Ben-David et al. (2013) are more overconfident during times of high volatility than during times of low volatility.

### 3.5 Cross-sectional Differences in Confidence

The existence of each kind of equilibrium depends on all main primitives of the model, which yields the following comparative statics.

**Proposition 2.** *The overconfidence equilibrium is more likely to exist in any of the following situations:*

1. *If the agent is more prone to anxiety – in the sense that $u$ remains unchanged but $v$ is even more risk averse.*
2. *If the high-risk state is more likely ex ante – in the sense that $\pi$ is higher.*
3. *If the high-risk state is more risky – in the sense that $G_L$ remains unchanged but a mean-preserving spread is added to $G_H$.*

Figure 6 illustrates the comparative statics of Proposition 2 by showing how the function $D(\varphi^e)$ is affected by increasing anxiety, and the likelihood or the riskiness

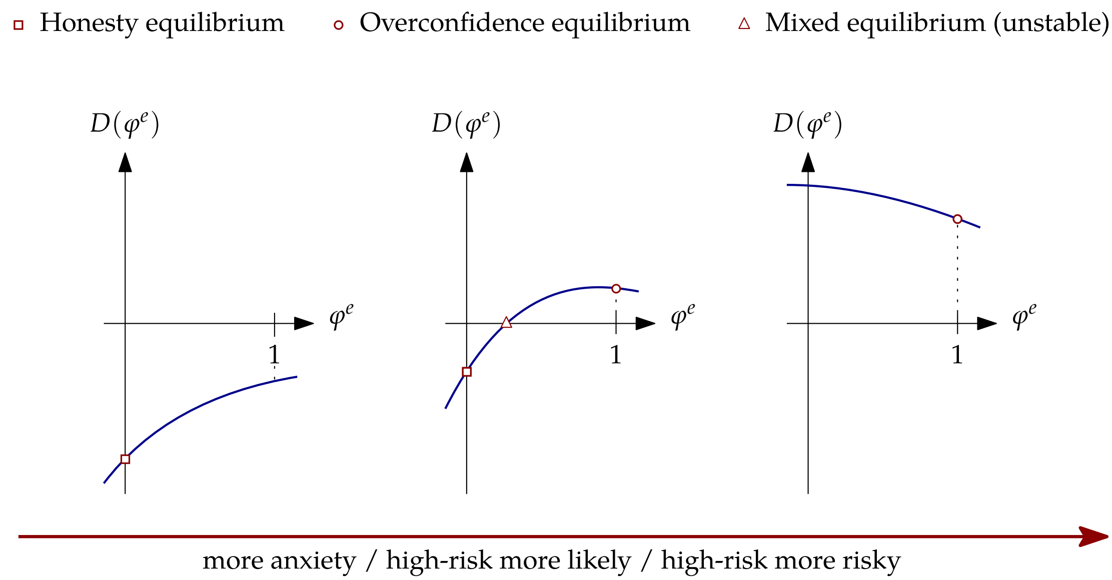


Figure 6: Effect of more anxiety, and higher likelihood or riskiness of the high risk state on $D(\varphi^e)$ and resulting equilibria.

of the high risk state. All three increase $D(\varphi^e)$ which makes $D(1) \geq 0$ and therefore existence of the overconfidence equilibrium more likely (see the Proof of Proposition 2 for more detail). Perhaps counterintuitively, we find that agents who are more prone to anxiety when facing immediate risk are the ones that are more likely to exhibit overconfidence. In terms of the environment agents are in, we find that a riskier environment – both ex ante and ex post – is more conducive to overconfidence.

We can interpreting the results of Proposition 2 in the cross-section of environments faced by different agents. The fact that agents in riskier environments are more likely to exhibit overconfidence is reminiscent of work on cognitive dissonance such as Akerlof and Dickens (1982). Such studies typically assume psychic utility, such as the fear of accidents, as entering the agent's utility directly. By contrast, our framework also applies to environments where the agent's job involves risk taking without risk of bodily harm. For example, according to the model, finance professionals should be particularly likely to display overconfidence, as documented by Ben-David et al. (2013).

Note that, by construction, the agent in our model is only overconfident about immediate uncertainty, that is, self 1 in $t = 1$. However, even if we generalized the model to include selves prior to self 0, no incentive would exist to make any self overconfident other than the one facing the risk and making the decision. This detail of the model predictions squares nicely with the details of the evidence of Ben-David et al. (2013) who find CFOs are overconfident only at short horizons and are unbiased at longer horizons.

Combining the time-series result of Section 3.4 with the cross-sectional results of this section generates additional predictions. Relative to the average market participant, agents more prone to anxiety will exhibit greater swings between over- and underconfidence over time. The greater fluctuation in their "emotion-driven beliefs" leads anxiety-prone agents to trade more, as documented by Odean (1998, 1999). In addition, since their confidence is pro-cyclical, the more "emotional" agents also end up systematically on the wrong side of the market – buying high and selling low – thus loosing money as documented by Lo et al. (2005).

In a more general setting, feedback effects could emerge. Overconfident traders have a greater demand for risk than unbiased traders so that overconfidence sustains and reinforces excessive risk levels. Conversely, in a crisis, an underconfidence feedback could depress price levels below fundamentals.

### 3.6 Excessive Risk Taking

Welfare statements in models with dynamically inconsistent preferences are problematic (Schelling, 1984). Nevertheless, our model allows us to characterize certain risk taking as excessive. Specifically, the future self can end up taking risks that even the less risk-averse current self would have avoided. To an observer who is unaware of the agent's intra-personal conflict and resulting equilibrium level of overconfidence, the agent seems to be taking risks that are greater than can be explained even based on the less risk-averse preference $u$.

**Corollary 2.** *If $c_u^H < c_v^\pi$ and an equilibrium with overconfidence exists, the agent will be observed to take excessive risks, that is, she will appear less risk averse than both $v$ and $u$.*

This seemingly paradoxical situation of an anxious agent taking excessive risks can arise if the true state of the world is high risk, $\theta = H$, but the agent forgets the warning signal, $\hat{s} = \emptyset$. In this case, self 0 would like the cutoff $c_u^H$ to be used, but self 1 actually uses the cutoff $c_v(1) = c_v^\pi$. As illustrated in Figure 7, whenever the payoff

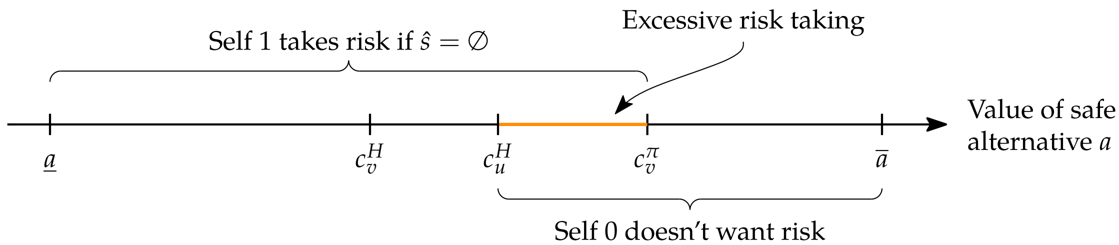


Figure 7: Excessive risk taking for $c_u^H < c_v^\pi$.

of the safe alternative is between the two cutoffs, $a \in \left(c_u^H, c_v^\pi\right)$, the agent takes risks in period 1 that even self 0 considers excessive. Of course, the paradox is due to the fact that self 0 knows the state of the world to be high-risk while self 1 has to rely on her Bayesian posterior.

Excessive risk taking can arise because the condition for an equilibrium with overconfidence, $D(\varphi^*|v) \geq 0$ does not necessarily imply that $\mathbb{E}_H[u(x)] > u(a)$ for all $a < c_v(\varphi^*)$, i.e., that self 0 wants the risky alternative where self 1 chooses it. To an outside observer who knows the state is $H$, the anxious agent using the cutoff $c_v(\varphi^*) = c_v^\pi$ appears to be *less* risk averse than the non-anxious preference $u$. This impression is not true, however. Rather, the anxious agent using the cutoff $c_v(\varphi^*)$ is systematically overconfident.

Why such excessive risk taking is an equilibrium outcome can be illustrated as follows. From Lemma 1, we know the certainty equivalents always satisfy $c_v^H < c_u^H$. In an equilibrium with excessive risk taking, we also have $c_u^H < c_v^\pi$. Given these two inequalities, we can decompose the marginal effect of forgetting more often on self 0's utility as follows:

$$\begin{aligned} D(\varphi^*) = \delta \int_{c_v^H}^{c_v(\varphi^*)} \left( \int_{-\infty}^{\infty} u(x)\, dG_H(x) - u(a) \right) dF(a) \\ = \delta \int_{c_v^H}^{c_u^H} \left( \int_{-\infty}^{\infty} u(x)\, dG_H(x) - u(a) \right) dF(a) \\ - \delta \int_{c_u^H}^{c_v^\pi} \left( \int_{-\infty}^{\infty} u(x)\, dG_H(x) - u(a) \right) dF(a)\,. \end{aligned}$$

In an equilibrium $\varphi^*$ with excessive risk taking, we have $D(\varphi^*) \geq 0$. Given the decomposition above, it follows that:

$$\overbrace{\int_{c_v^H}^{c_u^H}\left(\int_{-\infty}^{\infty}\delta u(x)\,dG_H(x)-\delta u(a)\right)dF(a)}^{\text{benefit of extra risk taking}}$$
$$\geq \underbrace{\int_{c_u^H}^{c_v^\pi}\left(\int_{-\infty}^{\infty}\delta u(x)\,dG_H(x)-\delta u(a)\right)dF(a)}_{\text{cost of extra risk taking}}.$$

For values of the safe alternative $a > c_v^H$, self 1 only takes risk if manipulated. For $a \in (c_v^H, c_u^H)$, such risk taking is desirable from self 0's point of view as captured by the utility benefit on the left-hand side of the inequality. For $a \in (c_u^H, c_v^\pi)$, such risk taking is excessive even from self 0's perspective as captured by the utility cost on the right-hand side of the inequality. For excessive risk taking to occur in equilibrium, the benefit of more risk taking when desired has to outweigh the cost of too much risk taking when not desired.

This result sheds light on the apparently excessive risk taking in the financial sector before the financial crisis of 2008–2009. While the actors involved, e.g. in the process of securitization, were best placed to receive signals about the true risks, Cheng et al. (2014) show these actors to be unaware of the risks. At the same time, many outside observers did see the risks that were building up, as cited in the introduction. Our model provides an explanation for this apparent paradox.

# 4 Interpretations of the Model

Our model describes overconfidence as resulting from a choice to forget risk signals. We now discuss two alternative interpretations of the belief-manipulation structure to illustrate the generality of our stylized framework, and to offer explanations that do not require a literal interpretation of the model.

## 4.1 Choice of Information Environment

One interpretation for the model's belief-manipulation framework is as a reduced-form metaphor for the choice of the agent's social or informational environment.

Specifically, given a preference for a biased posterior, an anxiety-prone agent will attempt to implement information and communication systems that render her misinformed about risks.

In an organizational context, management scholars and practitioners have remarked about the scarcity of openly expressed critical upward feedback. Indeed, the lack of informal and open upward feedback is the reason for the establishment of formal, anonymous upward-feedback mechanisms investigated by the personnel psychology literature (Atwater et al., 1995; Smither et al., 1995; Walker and Smither, 1999; Atwater et al., 2000). Lack of upward feedback is often said to be implicitly or explicitly mandated by the head of the organization ("killing the messenger"). Such lack of upward feedback – especially to risk managers of financial firms – is considered an important contributing factor to the financial crisis of 2007-2009 (Flannery et al., 2012).

In the context of our model, an anxiety-prone leader will design incentives for subordinates to systematically hide risk signals from her, especially in risky environments. As a result, the more severe the dynamic inconsistency in the leader's preferences and the higher the actual risk level, the less upward feedback subordinates will provide.[12]

Whereas these examples resonate with informal accounts of the informational environments in Wall Street firms before the recent crises, direct evidence on the biased choice of information from financial decision making also exists. Karlsson et al. (2009) find that investors look up their portfolio performance less often after receiving a signal about increased risks – behavior known as the "Ostrich Effect."[13] Bhattacharya et al. (2012) find that retail investors have little demand for unbiased advice – especially those who need it the most.

[12] An alternative interpretation is that the leader is naive about her horizon-dependent risk preferences, but her subordinates are aware and support her long-term self's risk-taking plans by supplying her with biased information.

[13] The original finding is that investors tend to not look up their portfolio's performance after market-wide declines about which they are likely to become informed via generic news reports. Falling prices are a signal for increased risk because either (i) increases in risk levels may cause price drops, or (ii) falling prices increase volatility estimates.

## 4.2 Self-manipulation with Alcohol and Drugs

A second interpretation of how the belief manipulation of our model may be implemented in practice is through the use of alcohol and other drugs. This section gives a brief review of psychological evidence on the effect of alcohol and other drugs on (i) risky behavior, (ii) forgetting and confidence, and (iii) performance changes. In addition, we discuss evidence on anxiety-prone individuals' strategic use of alcohol and other drugs to induce effects (i)–(iii).

The finding that alcohol is associated with more risky behavior is robust across domains. Field studies have shown alcohol consumption leads to risky sexual behavior (Halpern-Felsher et al., 1996; Cooper, 2002), accident-related injuries (Cherpitel et al., 1995), and dangerous driving patterns (Donovan et al., 1983). Pathological gambling is more common among people with alcohol-use disorders, and vice versa (Grant et al., 2002; Petry et al., 2005). In the lab, Lane et al. (2004) establish causality from alcohol consumption to risky behavior.

Low risk aversion or a low perception of risk can drive risky behavior. Cohen et al. (1958) show that the riskier driving behavior caused by alcohol consumption is associated with a decreased perception of risk, that is, a higher degree of overconfidence. Supporting the mechanism our model suggests, alcohol has also been shown to lead to forgetfulness, especially of negative signals (Nelson et al., 1986; Maylor and Rabbitt, 1987).

Evidence also shows that drugs are used strategically to improve performance, particularly by individuals with greater degrees of anxiety. As an example, Rimm (2002) recounts that composer-pianist Sergei Rachmaninoff was anxious about playing a particularly difficult passage in the 24th variation of his "Rhapsody on a Theme of Paganini." Based on a friend's recommendation, Rachmaninoff – otherwise a complete teetotaler – drank a glass of crème de menthe (a mint-flavored alcoholic beverage) before the premiere, which he then executed faultlessly. Rachmaninoff subsequently had the same drink before all performances of the piece and marked the 24th the "Crème de Menthe Variation."

More generally, Steptoe and Fidler (1987) find that 17% of professional musicians with high performance anxiety take sedatives as a method of coping. This number compares to 4% of musicians with medium levels of performance anxiety that take sedatives to cope, and 0% of the respondents with low performance anxiety. Based on our model, the performance of anxiety-prone individuals should improve with moderate levels of drug-induced overconfidence. James et al. (1977) as well as Brantigan et al. (1982) show that the use of beta-blockers improves the performance of musicians who suffer from stage fright. Lastly, anecdotal evidence on the "widespread use of [...] cocaine by professional traders" (Bossaerts, 2009) is consistent both with strategic self-manipulation and with our observations about cross-sectional differences in the level of overconfidence across environments.[14] Note that our interpretation of the use of alcohol and other drugs is in stark contrast to strategic self-handicapping (Bénabou and Tirole, 2002) or detrimental effects on intertemporal trade-offs (Schilbach, 2018).

## 5 Conclusion

Using standard tools in economics, this paper shows that horizon-dependent risk aversion ("anxiety") supplies a rationale for overconfident beliefs, wherein selective information processing is used as a tool to accomplish self-delusion. The model predicts salient features of organizational design, such as a tendency to suppress upward feedback about impending risks especially during times of high risk levels, individuals' choice of information systems and drug use that alter the perception of risks, as well as observed equilibrium levels of overconfidence as measured in surveys. Importantly, the model provides a rationale why individuals with access to the most precise information about risk levels can hold the most inaccurate beliefs about these risks, while outside observers have a more accurate view. Relatedly, we give a precise meaning to the notion of "excessive risk taking," and discuss the potential of endogenously generated confidence levels to amplify economic fluctuations. We leave an application to equilibrium asset-pricing models for future research.

[14] "Did cocaine use by bankers cause the global financial crisis?" The Guardian, Monday, April 15, 2013. http://www.theguardian.com/business/shortcuts/2013/apr/15/cocaine-bankers-global-financial-crisis

# References

Abdellaoui, M., E. Diecidue, and A. Onculer (2011). Risk preferences at different time periods: An experimental investigation. *Management Science 57*(5), 975–987.

AIG (2007). American International Group Q2 2007 earnings call transcript. http://seekingalpha.com/article/44048-american-international-group-q2-2007-earnings-call-transcript.

Akerlof, G. A. and W. T. Dickens (1982). The economic consequences of cognitive dissonance. *American Economic Review 72*(3), 307–319.

Akerlof, G. A. and R. J. Shiller (2010). *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press.

Alpert, M. and H. Raiffa (1982). A progress report on the training of probability advisors. In D. Kahneman and A. Tversky (Eds.), *In Judgement under Uncertainty: Heuristics and Biases*, pp. 294–305. Cambridge University Press.

Andries, M., T. M. Eisenbach, and M. C. Schmalz (2018). Horizon-dependent risk aversion and the timing and pricing of uncertainty. Working Paper.

Andries, M. and V. Haddad (2017). Information aversion. Working Paper.

Atwater, L., P. Roush, and A. Fischthal (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology 48*(1), 35–59.

Atwater, L. E., D. A. Waldman, D. Atwater, and P. Cartier (2000). An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology 53*(2), 275–297.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics 116*(1), 261–292.

Barberis, N. (2013). Psychology and the financial crisis of 2007–2008. In M. Haliassos (Ed.), *Financial Innovation: Too Much or Too Little?* MIT Press.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics 115*(1), 1–24.

Baucells, M. and F. Heukamp (2010). Common ratio using delay. *Theory and Decision 68*(1-2), 149–158.

Ben-David, I., J. R. Graham, and C. R. Harvey (2013). Managerial miscalibration. *Quarterly Journal of Economics 128*(4), 1547–1584.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3), 871–915.

Benoît, J.-P. and J. Dubra (2011). Apparent overconfidence. *Econometrica 79*(5), 1591–1625.

Bernardo, A. E. and I. Welch (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics and Management Strategy 10*(3), 301–330.

Bhattacharya, U., A. Hackethal, S. Kaesler, B. Loos, and S. Meyer (2012). Is unbiased financial advice to retail investors sufficient? Answers from a large field study. *Review of Financial Studies 25*(4), 975–1032.

De Bondt, W. F. M. and R. H. Thaler (1994). Financial decision-making in markets and firms: A behavioral perspective. Working Paper 4777, National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, and A. Shleifer (2017). Memory, attention and choice. Working Paper.

Bossaerts, P. (2009). What decision neuroscience teaches us about financial decision making. *Annual Review of Financial Economics 1*(1), 383–404.

Brantigan, C. O., T. A. Brantigan, and N. Joseph (1982). Effect of beta blockade and beta stimulation on stage fright. *American Journal of Medicine 72*(1), 88–94.

Brown, T. A. (1973). An experiment in probabilistic forecasting. Report R-944-ARPA, RAND Corporation.

Brunnermeier, M. K., C. Gollier, and J. A. Parker (2007). Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review Papers and Proceedings 97*(2).

Brunnermeier, M. K. and C. Julliard (2008). Money illusion and housing frenzies. *Review of Financial Studies 21*(1), 135–180.

Brunnermeier, M. K., F. Papakonstantinou, and J. A. Parker (2017). Optimal time-inconsistent beliefs: Misplanning, procrastination, and commitment. *Management Science 63*(5), 1318–1340.

Brunnermeier, M. K. and J. A. Parker (2006). Optimal expectations. *American Economic Review 95*(4), 1092–1118.

Caplin, A. and J. Leahy (2001). Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics 116*(1), 55–79.

Carrillo, J. D. and T. Mariotti (2000). Strategic ignorance as a self-disciplining device. *Review of Economic Studies 67*(3), 529–544.

Case, K. E. and R. J. Shiller (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity 2003*(2), 299–362.

Cheng, I.-H., S. Raina, and W. Xiong (2014). Wall Street and the housing bubble. *American Economic Review 104*(9), 2797–2829.

Cherpitel, C. J., T. Tam, L. Midanik, R. Caetano, and T. Greenfield (1995). Alcohol and non-fatal injury in the U.S. general population: A risk function analysis. *Accident Analysis & Prevention 27*(5), 651–661.

Coble, K. and J. Lusk (2010). At the nexus of risk and time preferences: An experimental investigation. *Journal of Risk and Uncertainty 41*, 67–79.

Cohen, J., E. Dearnaley, and C. Hansel (1958). The risk taken in driving under the influence of alcohol. *British Medical Journal 1*(5085), 1438.

Cohn, A., J. Engelmann, E. Fehr, and M. A. Maréchal (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review 105*(2), 860–885.

Compte, O. and A. Postlewaite (2004). Confidence-enhanced performance. *American Economic Review 94*(5), 1536–1557.

Cooper, A. C., C. Y. Woo, and W. C. Dunkelberg (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing 3*(2), 97–108.

Cooper, M. L. (2002). Alcohol use and risky sexual behavior among college students and youth: Evaluating the evidence. *Journal of Studies on Alcohol and Drugs 14*, 101–117.

Crary, W. G. (1966). Reactions to incongruent self-experiences. *Journal of Consulting Psychology 30*(3), 246.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreactions. *Journal of Finance 53*(6), 1839–1885.

Deaves, R., E. Lüders, and G. Y. Luo (2009). An experimental test of the impact of overconfidence and gender on trading activity. *Review of Finance 13*(3), 555–575.

Deaves, R., E. Lüders, and M. Schröder (2010). The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior & Organization 75*(3), 402–412.

Donovan, D. M., G. A. Marlatt, and P. M. Salzberg (1983). Drinking behavior, personality factors and high-risk driving: A review and theoretical formulation. *Journal of Studies on Alcohol and Drugs 44*(3), 395–428.

Eisenbach, T. M. and M. C. Schmalz (2016). Anxiety in the face of risk. *Journal of Financial Economics 121*(2), 414 – 426.

Epstein, L. (2008). Living with risk. *Review of Economic Studies 75*(4), 1121–1141.

Epstein, L. G. and I. Kopylov (2007). Cold feet. *Theoretical Economics 2*(3), 231–259.

Ericson, K. M. M. (2011). Forgetting we forget: Overconfidence and memory. *Journal of the European Economic Association 9*(1), 43–60.

Flannery, M. J., P. Glasserman, D. K. Mordecai, and C. Rossi (2012). Forging best practices in risk management. Working Paper 0002, U.S. Department of the Treasury Office of Financial Research.

Foote, C. L., K. S. Gerardi, and P. S. Willen (2012). Why did so many people make so many ex post bad decisions? The causes of the foreclosure crisis. Working Paper 18082, National Bureau of Economic Research.

Freud, S. (1904). *Zur Psychopathologie des Alltagslebens*. S. Karger.

Fudenberg, D. and D. K. Levine (2006). A dual-self model of impulse control. *American Economic Review 96*(5), 1449–1476.

Gennaioli, N., Y. Ma, and A. Shleifer (2015). Expectations and investment. *NBER Macro Annual 30*(1).

Gennaioli, N. and A. Shleifer (2010). What comes to mind. *Quarterly Journal of Economics 125*(4), 1399–1433.

Gervais, S., J. Heaton, and T. Odean (2011). Overconfidence, compensation contracts, and capital budgeting. *Journal of Finance 66*(5), 1735–1777.

Gervais, S. and T. Odean (2001). Learning to be overconfident. *Review of Financial Studies 14*(1), 1–27.

Gilboa, I., A. Postlewaite, and L. Samuelson (2015). Memory utility. Working Paper.

Glaser, M., T. Langer, and M. Weber (2013). True overconfidence in interval estimates: Evidence based on a new measure of miscalibration. *Journal of Behavioral Decision Making 26*(5), 405–417.

Grant, J. E., M. G. Kushner, and S. W. Kim (2002). Pathological gambling and alcohol use disorder. *Alcohol Research and Health 26*(2), 143–150.

Greenwood, R. and S. Nagel (2009). Inexperienced investors and bubbles. *Journal of Financial Economics 93*(2), 239–258.

Grubb, M. D. (2009). Selling to overconfident consumers. *American Economic Review 99*(5), 1770–1807.

Grubb, M. D. (2015). Overconfident consumers in the marketplace. *Journal of Economic Perspectives 29*(4), 9–36.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica 59*(3), 667–686.

Hackbarth, D. (2008). Managerial traits and capital structure decisions. *Journal of Financial and Quantitative Analysis 43*, 843–882.

Halpern-Felsher, B. L., S. G. Millstein, and J. M. Ellen (1996). Relationship of alcohol use and risky sexual behavior: A review and analysis of findings. *Journal of Adolescent Health 19*(5), 331–336.

Harrison, J. M. and D. M. Kreps (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *Quarterly Journal of Economics 92*(2), 323–336.

Himmelberg, C., C. Mayer, and T. Sinai (2005). Assessing high house prices: Bubbles, fundamentals, and misperceptions. Working Paper 11643, National Bureau of Economic Research.

Hirshleifer, D. and I. Welch (2002). An economic approach to the psychology of change: Amnesia, inertia, and impulsiveness. *Journal of Economics & Management Strategy 11*(3), 379–421.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Inoue, K., H. K. Kato, and T. Yamasaki (2012). Managerial overconfidence: Evidence from Japanese CEOs. Working Paper.

James, I., D. Griffith, R. Pearson, and P. Newbury (1977). Effect of oxprenolol on stage-fright in musicians. *The Lancet 310*(8045), 952–954.

Jones, E. E. and C. A. Johnson (1973). Delay of consequences and the riskiness of decisions. *Journal of Personality 41*(4), 613–637.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Karlsson, N., G. Loewenstein, and D. Seppi (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty 38*, 95–115.

Klayman, J., J. B. Soll, C. Gonzalez-Vallejo, and S. Barlas (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes 79*(3), 216–247.

Koellinger, P., M. Minniti, and C. Schade (2007). "I think I can, I think I can": Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology 28*(4), 502–527.

Korner, I. N. (1950). *Experimental Investigation of some aspects of the problem of repression: repressive forgetting*. Number 970. Bureau of Publications, Teachers College, Columbia University.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association 4*(4), 673–707.

Kyle, A. S. and A. Wang (1997). Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? *Journal of Finance 52*(5), 2073–2090.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics 112*(2), 443–477.

Lane, S. D., D. R. Cherek, C. J. Pietras, and O. V. Tcheremissine (2004). Alcohol effects on human risk taking. *Psychopharmacology 172*(1), 68–77.

Lichtenstein, S., B. Fischhoff, and L. Phillips (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgement under Uncertainty: Heuristics and Biases*, pp. 306–334. Cambridge University Press.

Lo, A. W., D. V. Repin, and B. N. Steenbarger (2005). Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review 95*(2), 352–359.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003). Projection bias in predicting future utility. *Quarterly Journal of Economics 118*(4), 1209–1248.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk-taking? *Quarterly Journal of Economics 126*(1), 373–416.

Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *Quarterly Journal of Economics 131*(1), 53–87.

Malmendier, U. and G. Tate (2005). CEO overconfidence and corporate investment. *Journal of Finance 60*(6), 2661–2700.

Maylor, E. A. and P. Rabbitt (1987). Effect of alcohol on rate of forgetting. *Psychopharmacology 91*(2), 230–235.

Mischel, W., E. B. Ebbesen, and A. M. Zeiss (1976). Determinants of selective memory about the self. *Journal of Consulting and Clinical Psychology 44*(1), 92.

Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological Review 115*(2), 502–517.

Moore, D. A. and D. Schatz (2017). The three faces of overconfidence. *Social and Personality Psychology Compass 11*(8), e12331.

Nelson, T. O., M. McSpadden, K. Fromme, and G. A. Marlatt (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General 115*(3), 247.

Noussair, C. and P. Wu (2006). Risk tolerance in the present and the future: An experimental study. *Managerial and Decision Economics 27*(6), 401–412.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance 53*(6), 1887–1934.

Odean, T. (1999). Do investors trade too much? *American Economic Review 89*(5), 1279–1298.

Onculer, A. (2000, April). Intertemporal choice under uncertainty: A behavioral perspective. Working Paper 2000/37/TM, INSEAD.

Pagel, M. (2017). A news-utility theory for inattention and delegation in portfolio choice. *Econometrica* (forthcoming).

Petry, N. M., F. S. Stinson, and B. F. Grant (2005). Comorbidity of DSM-IV pathological gambling and other psychiatric disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry 66*(5), 564–574.

Phelps, E. S. and R. A. Pollak (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies 35*(2), 185–199.

Piazzesi, M. and M. Schneider (2008). Inflation illusion, credit, and asset prices. In J. Y. Campbell (Ed.), *Asset Prices and Monetary Policy*, pp. 147–189. University of Chicago Press.

Reinhart, C. M. and K. S. Rogoff (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press.

Rimm, R. (2002). *The Composer-Pianists: Hamelin and The Eight*. Amadeus Press.

Sagristano, M. D., Y. Trope, and N. Liberman (2002). Time-dependent gambling: Odds now, money later. *Journal of Experimental Psychology: General 131*(3), 364–376.

Scheinkman, J. A. and W. Xiong (2003). Overconfidence and speculative bubbles. *Journal of Political Economy 111*(6), 1183–1219.

Schelling, T. C. (1984). Self-command in practice, in policy, and in a theory of rational choice. *American Economic Review 74*(2), 1–11.

Schilbach, F. (2018). Alcohol and self-control: A field experiment in India. Working Paper.

Shefrin, H. (2001). Behavioral corporate finance. *Journal of Applied Corporate Finance 14*, 113–124.

Shiller, R. J. (2007). Understanding recent trends in house prices and home ownership. Working Paper 13553, National Bureau of Economic Research.

Smith, M. H. and G. Smith (2006). Bubble, bubble, where's the housing bubble? *Brookings Papers on Economic Activity 2006*(1), 1–67.

Smither, J. W., M. London, N. L. Vasilopoulos, R. R. Reilly, R. E. Millsap, and N. Salvemini (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology 48*(1), 1–34.

Staël von Holstein, C.-A. S. (1972). Probabilistic forecasting: An experiment related to the stock market. *Organizational Behavior and Human Performance 8*, 139–158.

van den Steen, E. (2004). Rational overoptimism (and other biases). *American Economic Review 94*(4), 1141–1151.

van den Steen, E. (2011). Overconfidence by Bayesian-rational agents. *Management Science 57*(5), 884–896.

Steptoe, A. and H. Fidler (1987). Stage fright in orchestral musicians: A study of cognitive and behavioural strategies in performance anxiety. *British Journal of Psychology 78*(2), 241–249.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies 23*(3), 165–180.

Tomassini, L. A., I. Solomon, M. B. Romney, and J. L. Krogstad (1982). Calibration of auditors' probabilistic judgments: Some empirical evidence. *Organizational Behavior and Human Performance 30*(3), 391–406.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131.

Walker, A. G. and J. W. Smither (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology 52*(2), 393–423.

Welch, E. (1999). The heat of the moment. Working Paper.

# Appendix

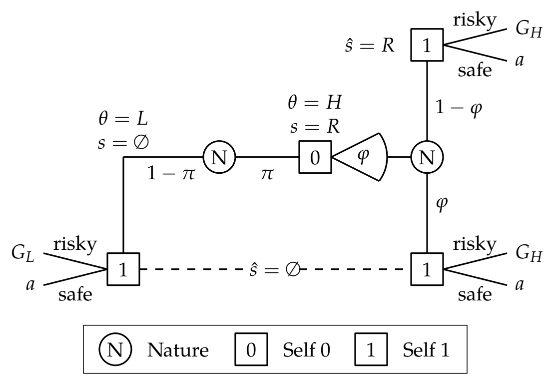


Figure 8: Extensive-form representation of intra-personal game.

**Proof of Proposition 1.** The belief $\hat{\pi}(\varphi^e)$ is continuous and increasing in $\varphi^e$. Therefore, the certainty equivalent $c_v(\varphi^e)$, which forms the upper bound of the integral in $D(\varphi^e)$, is continuous and decreasing in $\varphi^e$, varying between $c_v(0) = c_v^L$ and $c_v(1) = c_v^\pi$, implying $D(\varphi^e)$ is continuous in $\varphi^e$. The integrand in $D(\varphi^e)$ is positive if and only if $a < c_u^H$. The integral starts at $c_v^H < c_u^H$ where the integrand is positive. Over the interval of integration, the integrand decreases, turning negative if $a$ increases beyond $c_u^H$. Whether that happens depends on the upper bound $c_v(\varphi^e) \in \left[c_v^\pi, c_v^L\right]$.

If $c_u^H < c_v^\pi$ then the integrand is negative at the upper bound for all $\varphi$. Reducing the upper bound increases the integral so $D$ is uniformly increasing. For $c_u^H \searrow c_v^H$, the interval $\left[c_v^H, c_u^H\right]$ where the integrand is positive shrinks so $D$ is negative for $c_u^H$ sufficiently close to $c_v^H$. This is low disagreement, self 0 never wants to make self 1 more confident, the only equilibrium is honesty, $D(0) < 0$. As $c_u^H$ increases, $D$ remains increasing but can become positive. Then the mixed and the overconfidence equilibria appear.

If $c_u^H \in \left(c_v^\pi, c_v^L\right)$, the integrand is negative at the upper bound for low $\varphi$ ($D$ increasing) and positive at the upper bound for high $\varphi$ ($D$ decreasing) so $D$ is inverse U-shaped. At $\varphi = 1$, the upper bound is $c_v^\pi < c_u^H$ so the integrand is positive over the entire integration. Therefore $D(1) > 0$, i.e. the honesty equilibrium is guaranteed to exist.

If $c_u^H > c_v^L$, the integrand is always positive so the integral is always positive and reducing the upper bound decreases the integral; so $D$ is positive and decreasing everywhere. This is high disagreement, self 0 always wants to make self 1 more confident, and the only possible equilibrium is overconfidence.

Note that $D$ is increasing at the mixed equilibrium so the equilibrium is not "stable" in the following sense: for a small upward perturbation in self 1's expected $\varphi^e$, self 0 finds it optimal to increase $\varphi$ further, until we reach the overconfidence equilibrium $\varphi = \varphi^e = 1$ and vice versa for a small downward perturbation. $\square$

**Lemma 2.** *Consider two von Neumann-Morgenstern utility functions $v_1$ and $v_2$. If $v_2$ is more risk averse than $v_1$, then $D_{v_2}(\varphi^e) > D_{v_1}(\varphi^e)$ for all $\varphi^e$.*

**Proof of Lemma 2.** If $v_2$ is more risk averse than $v_1$, then $c^H_{v_2} < c^H_{v_1}$ and $c_{v_2}(\varphi^e) < c_{v_1}(\varphi^e)$ for all $\varphi^e$. This implies that for all $\varphi^e$,

$$\begin{aligned} D_{v_2}(\varphi^e) &= \int_{c^H_{v_2}}^{c_{v_2}(\varphi^e)} \Big(\mathbb{E}_H[\delta u(x)] - \delta u(a)\Big)\, dF(a) \\ &> \int_{c^H_{v_1}}^{c_{v_1}(\varphi^e)} \Big(\mathbb{E}_H[\delta u(x)] - \delta u(a)\Big)\, dF(a) \\ &= D_{v_1}(\varphi^e)\,, \end{aligned}$$

as desired. $\square$

**Lemma 3.** *Consider two von Neumann-Morgenstern utility functions $v_1$ and $v_2$. If $v_2$ is more risk averse than $v_1$ and if $\bar{\varphi}_1$ and $\bar{\varphi}_2$ exist such that $D_{v_1}(\bar{\varphi}_1) = 0$ and $D_{v_2}(\bar{\varphi}_2) = 0$, then $\bar{\varphi}_1 > \bar{\varphi}_2$.*

**Proof of Lemma 3.** If $v_2$ is more risk averse than $v_1$, then $c^H_{v_2} < c^H_{v_1}$ by Lemma 1 so the integral in $D_{v_2}(\bar{\varphi}_2)$ has a smaller lower bound. Because $\mathbb{E}_H[\delta u(x)] - \delta u(a)$ is a strictly decreasing function of $a$, for $D_{v_1}(\bar{\varphi}_1) = D_{v_2}(\bar{\varphi}_2) = 0$, it is necessary that $c_{v_2}(\bar{\varphi}_2) > c_{v_1}(\bar{\varphi}_1)$, that is, that the integral in $D_{v_2}(\bar{\varphi}_2)$ must have a greater upper bound. Because $c_{v_2}(\varphi) < c_{v_1}(\varphi)$ for a given $\varphi$, and $c_v(\varphi)$ is decreasing in $\varphi$ for $v_1$ and $v_2$, we have $\bar{\varphi}_2 < \bar{\varphi}_1$. $\square$

**Proof of Proposition 2.** For part 1, from Lemma 2 we know $D_{v_2}(1) > D_{v_1}(1)$ for $v_2$ more risk averse than $v_1$. Therefore, an overconfidence equilibrium exists for $v_2$

if it exists for $v_1$. Again using Lemma 2, we know $D_{v_2}(0) > D_{v_1}(0)$ for $v_2$ more risk averse than $v_1$. Therefore, an honesty equilibrium exists for $v_1$ if it exists for $v_2$. Finally, if a mixed equilibrium exists for $v_1$ and $v_2$, characterized by $\bar{\varphi}_1$ and $\bar{\varphi}_2$ respectively, then by Lemma 3, we have $\bar{\varphi}_1 > \bar{\varphi}_2$.

For part 2, note that $\hat{\pi}(\varphi^e)$ is increasing in $\pi$ similarly as in $\varphi^e$, so analogously to the proof of Proposition 1, we know $D(\varphi^e)$ is increasing in $\pi$. Therefore, for higher $\pi$, the condition $D(0) \leq 0$ for an honesty equilibrium is harder to satisfy, the condition $D(1) \geq 0$ for an overconfidence equilibrium is easier to satisfy, and any solution to $D(\bar{\varphi}) = 0$ will be for a higher $\bar{\varphi}$.

For part 3, note that adding a mean-preserving spread to the distribution $G_H$ has the same effect on certainty equivalents as more disagreement between $v$ and $u$, so the arguments for part 1 apply analogously. $\square$