GCOE Discussion Paper Series

Global COE Program
Human Behavior and Socioeconomic Dynamics

Discussion Paper No.117

Vision and Flexibility in a Model of Cognitive Dissonance

Junichiro Ishida

February 2010

GCOE Secretariat
Graduate School of Economics
OSAKA UNIVERSITY
1-7 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan

Vision and Flexibility in a Model of Cognitive Dissonance*

Junichiro Ishida

ISER, Osaka University

February 4, 2010

Abstract

This paper explores the consequences of cognitive dissonance, coupled with time-inconsistent preferences, in an intertemporal decision problem with two distinct goals: acting decisively on early information (vision) and adjusting flexibly to late information (flexibility). The decision maker considered here is capable of manipulating information to serve her self-interests, but a tradeoff between distorted beliefs and distorted actions constrains the extent of information manipulation. Building on this tradeoff, the present model provides a unified framework to account for the conformity bias (excessive reliance on precedents) and the confirmatory bias (excessive attachment to initial perceptions).

JEL Classification Codes: D03, D99.
Key Words: Cognitive dissonance, Vision, Flexibility, Conformity bias, Confirmatory bias.

*Correspondence to: Junichiro Ishida, Institute of Social and Economic Research, Osaka University, 6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan. E-mail address: jishida@iser.osaka-u.ac.jp.

"Wise men change their minds, fools never."

– from Zhouyi ("The Book of Changes")

"Not to amend a fault after you commit it, that is a true fault."

– from Analects of Confucius

1 Introduction

Suppose that you (e.g., CEO, constituency) are in search of someone (e.g., manager, politician) who, on your behalf, conducts some long-term project. Ideally, you would like your decision maker to achieve two distinct goals. First, time is precious and it is hence important to exhibit a "vision" by acting decisively on early information before things unfold themselves. Second, as more information becomes available over time, it is also equally important to be open-minded and "flexible" enough to adjust to new information without being prejudiced by any prior beliefs. The problem is that, while both vision and flexibility are clearly indispensable for effective decision making, a body of literature in both social psychology and behavioral economics suggests that they are indeed very rare to be found: on one hand, we often rely excessively on precedents, instead of our own private information, and bias our perceptions and behaviors towards the majority (the conformity bias);[1] on the other hand, we also often let our initial perceptions interfere too much with our decision making by misinterpreting subsequent evidence in favor of the initial perceptions (the confirmatory bias).[2] In either case, we do not process information as objectively as an unbiased Bayesian learner would do, consequently resulting in all sorts of inefficient outcomes.[3]

[1]One of the early experiments to show the conformity bias is provided by Asch (1951) who shows that an individual conform to the judgments of the majority that are obviously incorrect.

[2]A number of experiments, e.g., Ariely et al. (2003), document that our perceptions and valuations are initially malleable but become imprinted and fairly stable after we are called upon to make an initial decision (the anchoring effect). Our notion of the confirmatory bias also implies that first impressions matter. See Rabin and Shrag (1999) for psychological evidence on the confirmatory bias.

[3]For instance, the lack of vision is often touted as a leading cause for business failures which occur ever so frequently. The lack of flexibility is also salient, as exemplified by the two old Chinese proverbs in the opening quotes.

Given that the lack of vision and flexibility in dynamic decision making is so ubiquitous, it is important to understand why our perceptions and behaviors are biased in such a systematic manner. The main purpose of the paper is to provide a unified framework to account for both the conformity bias (the lack of vision) and the confirmatory bias (the lack of flexibility) by building on the theory of cognitive dissonance, coupled with time-inconsistent preferences, as an integral piece of the analysis.[4] To this end, we consider a two-period setting where a decision maker must take some actions sequentially in an uncertain environment. In each period, she can observe a free signal which partially reflects the state of nature: for clarity, we refer to the signal observed in the first (second) period as the first (second) signal. Each observed signal is either informative or noisy depending on some factors beyond her control, and when it is informative, its accuracy depends on her unknown ability type. Based on the available information, the decision maker makes an inference about the state of nature and takes an appropriate action.

A critical aspect of this setting is that the decision maker does not know whether any given signal is informative or noisy beyond its objective probability, which provides some leeway for her interpretation of the available information. While an unbiased Bayesian learner would update her belief based on this objective probability, we deviate from this convention by allowing the decision maker to assign any probability (of the signal being informative) to serve her self-interests. This information manipulation is not costless, however, because any distorted beliefs would persist and later lead to distorted and less efficient choices of action.[5] This tradeoff creates a tension within herself and constrains the extent of information manipulation.

We attempt to capture this entire process by constructing a model of intrapersonal con-

[4]The theory of cognitive dissonance asserts that people experience an uncomfortable feeling when holding contradictory ideas or cognitions. To reduce dissonance, people change their beliefs and behaviors, often by interpreting events or facts in a self-serving manner. For instance, it is well known that people tend to discover flaws in evidence that conflicts with their self-interests but rely excessively on evidence that portrays them in a good light. For instance, several studies show that people tend to be more critical of the validity of tests on which they failed (Wyer and Frey, 1983; Pyszczynski et al, 1985).

[5]A case in point is an instance of workers in a hazardous job, as depicted in Akerlof and Dickens (1982). To escape from the fear of injury, they have an incentive to believe that the job is actually safe, but in this state of self-deception, they also often fail to wear required safety equipments – an apparently inefficient action. This inefficiency comes from the need for dissonance reduction: wearing safety equipments while believing that the job is safe is apparently contradictory and hence puts them in the state of dissonance.

flicts between the objective self and the subjective self. The objective self represents the rational side of the decision maker and is a forward-looking unbiased Bayesian learner, whereas the subjective self represents the more instinctive and myopic side and cares about her self-images (ego preferences). This framework then contains two sources of cognitive bias: on one hand, the subjective self's ego gives rise to an incentive to manipulate information to preserve her self-images; on the other hand, the difference in time horizon between the two selves amounts to time-inconsistent objectives and the need to regulate future selves. With these two forces at work, we obtain the following results.

The conformity bias (the lack of vision): The decision maker relies excessively on precedents by discounting the first signal.

The confirmatory bias (the lack of flexibility): The decision maker exhibits excessive attachment to her initial choice of action, both by exaggerating the second signal if it is consistent with the first signal and discounting it if it is inconsistent.

To see the underlying logic of the current model more clearly, we start with the second period of the model where the decision maker has already accumulated some information. The driving force of the confirmatory bias is the subjective self's preference for better self-images. As stated, the accuracy of each observed signal is dependent on the decision maker's ability. This means that if the second signal is consistent with the first one, chances are that she has made the right observations. In this case, she has a reason to exaggerate the informativeness of the second signal because that will boost her self-confidence. On the other hand, inconsistent signals are a bad news for the decision maker because this may indicate the lack of ability on her part, placing her under the state of dissonance. To reduce dissonance, she then downplays the importance of the second signal by simply disregarding it as uninformative and unreliable. This result thus indicates that "first impressions matter" in that the decision maker is prejudiced by the first signal, one way or the other, to the extent that it is not warranted objectively. In either case, the decision maker lacks flexibility as she fails to respond objectively to new information, consequently ending up with a biased view of the world.

In contrast, the conformity bias concerns what happens in the first period. The underlying logic leading to this bias now stems partially from a self-control problem due to time-inconsistent objectives. To see how this bias arises, notice that while the decision maker's information processing in the second period is biased, its magnitude is heavily dependent on the informativeness of the first signal. The magnitude is larger if she relies strongly on the first signal and takes a clear stance. In contrast, if the decision maker regards the first signal as a complete noise containing no relevant information, the second signal provides no information about her ability type whatsoever, and there arises no need to justify herself. For a decision maker who is hampered by the self-control problem, therefore, discounting the informativeness of the first signal and remaining rather ambiguous are a way to regulate her future selves. In so doing, however, she fails to act decisively on early information and is hence unable to articulate her vision before things unfold themselves, which we take as representing the lack of vision.[6]

The present setup allows us to identify several important factors that affect the way the decision maker processes information over time and hence critical determinants of vision and flexibility. First, the key factor in this entire process is the extent to which the rational and farsighted objective self can regulate the more instinctive and myopic subjective self. The lack of control over the subjective self, or the lack of willpower as we later define, is the source of all behavioral biases in our model. A decision maker with weak willpower must compromise more and consequently ends up with a more biased view of the world. In contrast, all sorts of biases disappear if the objective self has perfect control over the subjective self, which allows the decision maker to rationally and objectively process information.

Another important factor is the decision maker's self-confidence in her own ability, which consists of two aspects: the level (the mean) and the fragility (the variance). The self-confidence fragility is about how secure the decision maker feels about herself and measured

[6]There are several papers which explicitly focus on the notion of vision. To name a few, Rotemberg and Saloner (2000) model vision as a bias which makes the manager favor one project over the other. Similarly, Rotemberg and Saloner (1994) argue that a firm could be better off by committing to a specific business strategy. Van den Steen (2005) also formalizes the notion of vision and shows that a leader with strong beliefs would attract employees with similar beliefs. This sorting effect improves coordination within the firm, suggesting a channel through which strong vision can affect the firm's performance. In this paper, we focus on a different aspect of vision, defined as the ability to effectively utilize early information before things unfold themselves.

by the prior variance of the ability type, as perceived by the subjective self. The effect of the self-confidence fragility is complementary to the effect of willpower and generally works in the same direction. If the decision maker is perfectly sure about her own ability, whatever she observes does not change her assessment of herself, and the incentive to preserve self-confidence disappears no matter how severe the self-control problem is. As she becomes less secure about herself, the need to preserve self-confidence intensifies, leading to all sorts of biased behavior.

In contrast, the effect of the self-confidence level, measured by the prior expectation of the ability type, is more complicated compared to the other two factors. In general, a decision maker with more self-confidence exhibits a lager bias in the second period in both directions (both obsession and stubbornness) and hence is less flexible. This is because those with high self-confidence have more trust in the first signal, and the cost of biasing the interpretation of the second signal is relatively low,[7] indicating that high self-confidence *per se* does not necessarily leads to better and objective judgments. In contrast, the effect of the self-confidence level on vision is less certain, as it could go either way. The logic is now partially reversed because the first signal is more reliable for those with high self-confidence and it is relatively more costly for them to bias its interpretation. There is a countervailing effect, however, since the second-period bias is larger for those with high self-confidence and the self-control problem is hence severer, given the same level of willpower and self-confidence fragility. The overall effect is determined by this tradeoff but we argue that the first effect is more likely to dominate the second. In light of this, we claim that high self-confidence makes a decision maker less flexible while it tends to make her more visionary.

The paper proceeds as follows. In the remainder of this section, we briefly discuss related literatures. Section 2 outlines the basic model of subjective information evaluation. Section 3 provides the main results of the paper, with special attention paid to the effects of self-confidence and willpower. Finally, section 4 makes some concluding remarks.

[7]To be more precise, this is because we consider a situation where the prediction ability is tested more in the first period. We explore more on this point later.

Related literature

The paper spans over several distinct areas, and there are accordingly several strands of related literature. Among them, the paper is most closely related to Akerlof and Dickens (1982) in its motivation. In this seminal work, Akerlof and Dickens (1982) posit that: (i) people have preferences not only over states of the world, but also over their beliefs about the state of the world; (ii) they also have some control over the beliefs; (iii) the beliefs once chosen persist over time. The current analysis closely follows these premises in that we consider a decision maker who is capable of deceiving herself and manipulating her belief in any way she can to serve her self-interests. Starting from these premises, however, the current analysis builds on an entirely different analytical framework: we model a detailed process of intrapersonal conflicts and extend it to a dynamic setting to see its dynamic implications.

Besides this, there is now a growing amount of interest on the manipulable nature of our belief system. The importance of the self-serving nature of our belief system has also gained some recognition among economists,[8] and several attempts are made to capture this seemingly robust human nature.[9] Some early attempts to model belief distortions as a result of optimized behaviors are provided by Benabou and Tirole (2002, 2004, 2006), Benabou (2008a, 2008b) and Brunnermeier and Parker (2005).[10] These papers consider various cases where an agent can manipulate her own memory or expectations, thereby explicitly looking into the process of biased belief formation.[11] The current paper intends to add to this line of literature: we view our main contribution in our specific way to apply the notion of cognitive dissonance to the problem of dynamic information processing, which allows us to identify the

[8] A self-serving bias refers to a human tendency to make internal attributions for success and external attribution for failure. The study on this issue has yielded a voluminous literature; see, for instance, Zuckerman (1979). See, for instance, Babcock and Loewenstein (1997).

[9] Formal analyses which incorporate confirmation bias and explore its consequences include Rabin and Schrag (1999) and Compte and Postlewaite (2004). The approach of this paper is different, as we derive it as an optimized behavior of divided selves, instead of assuming it and exploring its consequences.

[10] Also, see Bracha and Brown (2009) and Mayraz (2009).

[11] Several recent works also derive overconfidence, a type of self-serving bias, without explicitly assuming biased information processing. Van den Steen (2004) argues that facing a choice of actions, people are more likely to choose an action of which they overestimate the probability of success, and hence tend to be overoptimistic about the current action that they choose. Santos-Pinto and Sobel (2005) consider a case where different individuals hold different opinions about how skills combine to determine an ability level. In that setting, positive self-image arises because they work more on a skill that is critical to their production function. Also, see Köszegi (2006) for how a rational agent comes to hold an overoptimistic view of herself.

sources of the two frequently cited cognitive biases and make a dynamic link between them. We also provide an analytically tractable framework which allows us to intuitively relate the presence of the biases to deep parameters of our cognitive process.

Whereas the confirmatory bias is deeply rooted in the self-serving nature of our belief system, the conformity bias in our model stems from a type of self-control problem due to time-inconsistent objectives which results from the process of intrapersonal conflicts. In this sense, the paper is related to Carrillo and Mariotti (2000) and Benabou and Tirole (2002, 2004), among many others, who explore various aspects of self-regulation for individuals with time-inconsistent preferences. In the present model, the self-control problem arises due to the difference in time horizon between the two selves, where the myopic subjective self gets in the way to obstruct the objective self to make a fully rational decision. The lack of vision surfaces because taking an ambiguous stance early on is a way to reduce any concern for protecting self-images and hence to regulate the future subjective self.

Besides these so-called "behavioral" aspects, the main thrust of the paper is the dynamic nature of information acquisition. Works along this line include Prendergast and Stole (1996) and Li (2007) who consider dynamic aspects of reputation concerns in signaling models. This paper is especially related to Prendergast and Stole (1996) who consider a situation where an agent signals his ability to acquire information and examine how that signaling incentive changes over the course of his career. They show that the agent exaggerates information when young and tends to become more conservative as he gets older. An important aspect of their model is that changing behavior from previous periods is costly because that signals the lack of ability on his part. Our model shares this aspect, which is the driving force of one side of inflexibility. Aside from this, however, there are several notable differences. First, we consider a totally different setup, which allows us to focus on the flip side of inflexibility, i.e., exaggeration of favorable information.[12] Second, while Prendergast and Stole (1996) only consider static incentives (maximization of the current payoff), we explicitly consider dynamic incentives to regulate the future selves, which results in the discounting of early

[12]In their model, the state space is extended on the real line and hence has no "extreme" points. In this specification, there is no way to exaggerate information in our sense.

information.

2 The model

Our optimization problem rests on the tradeoff between biased beliefs and biased actions, which stems from our inherent desire for coherence and consistency as suggested by the theory of cognitive dissonance.

2.1 Setup

Consider a two-period model where a risk-neutral decision maker (DM) engages in some long-term project. In each period, DM observes a signal, evaluates it and takes an action based on the evaluation. The action chosen in each period yields a payoff which is realized and received at the end of period 2. The value of the project depends on the action taken in each period and the (time-invariant) state of nature $\theta \in \{0, 1\}$. The state of nature is not directly observable, and its prior distribution is given by

$$\text{prob}\{\theta = 0\} = \text{prob}\{\theta = 1\} = 0.5.$$

That is, each state occurs equally likely *ex ante*.

In this environment, DM's job is to correctly predict the state and choose an appropriate course of action accordingly. More precisely, let $a_t \in [0, 1]$ denote the action chosen in period $t = 1, 2$. Given some realized state θ, the value of the project (for period t) is given by $v_\theta(a_t)$ where $v_0(a) = -a^2$ and $v_1(a) = -(1 - a)^2$. This specification implies that the actions are *ex ante* symmetric in every respect. Given the action a and some belief $\rho := \text{prob}\{\theta = 1 \mid$ available information$\}$, the expected value of the project is computed as

$$R(a, \rho) = -\rho(1 - a)^2 - (1 - \rho)a^2.$$

The first-order condition implies that the optimal action is $a = \rho$.

2.2 Signals

At the beginning of each period, DM has a chance to observe a signal, which possibly contains some information about the state of nature. The signal is either informative or noisy, and the (objective) probability that any given signal is informative is $\tilde{\gamma} \in (0,1)$. When the signal is noisy, it contains no information about the realized state, so that

$$\text{prob}\{s_t = \theta \mid \text{the signal is noisy}\} = 0.5,\ t = 1, 2.$$

When it is informative, on the other hand, the signal coveys some information about the realized state although its accuracy differs across periods. DM's judgement ability is tested more in period 1, and hence the accuracy depends more on her ability type, denoted by $\eta \in [0,1]$:

$$\text{prob}\{s_1 = \theta \mid \text{the signal is informative}, \eta\} = \frac{1+\eta}{2}.$$

The situation unfolds itself and becomes more predictable as time passes by, so that an informative signal in period 2 perfectly reflects the realized state:

$$\text{prob}\{s_2 = \theta \mid \text{the signal is informative}, \eta\} = 1.$$

It is perhaps worth emphasizing that the second signal contains information not just about the realized state but also about the ability type, because DM with low ability is more likely to observe inconsistent signals. The problem is that the informativeness of any given signal is not known beyond its objective probability (which we refer to as the objective informativeness), and that provides some leeway for the decision maker's interpretation.

2.3 Information processing and belief formation

There are two unknowns in this model, the ability type η and the state of nature θ, and a belief must be formed on each of them. Although DM knows nothing more than the fact that any given signal is informative with probability $\tilde{\gamma}$, we here consider a situation where she can deceive herself and subjectively assigns the informativeness to each observed signal in a self-serving manner. Let $\gamma_t \in [0,1]$ denote the subjective informativeness of the observed signal, i.e., the subjective assessment of the probability that the signal observed in period t is informative.[13] While there are potentially many ways to capture this aspect, we model DM

[13]The important point is that this informativeness is exogenous and, more importantly, independent of DM's ability type, so that it is not ego-threatening. Our stance here is that the accuracy of each signal depends on

as a multi-layered self with divided interests between the objective self (cognition) and the subjective self (affect).[14]

The objective self: The objective self represents a rational side of DM who is forward-looking and processes information in an objective manner as conventionally assumed. The objective self derives utility from the outcome of the project and is hence interested in knowing the true state of nature. Given some informativeness (g_1, g_2), let

$$\rho_1(s_1; g_1) := \text{prob}\{\theta = 1 \mid s_1, g_1\},\ \rho_2(s_1, s_2; g_1, g_2) := \text{prob}\{\theta = 1 \mid s_1, s_2, g_1, g_2\},$$

denote the beliefs about θ where $g_t \in \{\gamma_t, \tilde{\gamma}\}$: we say that a belief is unbiased when $g_t = \tilde{\gamma}$ and compromised when $g_t = \tilde{\gamma}$. We also assume that the objective self knows the true value of η with precision, which is denoted by μ.

When evaluating her expected payoff, the objective self uses the unbiased belief. Assuming that the objective self does not discount the future,[15] the expected (total) payoff for the objective self in period 1 is hence given by

$$\pi_t^{\text{O}}(a_t) = R(a_t, \tilde{\rho}_t) + E[R(a_2, \tilde{\rho}_2) \mid s_1, \gamma_1],$$

where $\tilde{\rho}_1 := \rho_1(s_1; \tilde{\gamma})$ and $\tilde{\rho}_2 := \rho_2(s_1, s_2; \tilde{\gamma}, \tilde{\gamma})$ denote the unbiased beliefs. The expectation here is taken over possible second-period signals. Similarly, the expected payoff in period 2 is

$$\pi_2^{\text{O}}(a_1, a_2) = R(a_2, \tilde{\rho}_2).$$

The subjective self: The subjective self represents a more primitive and instinctive side of DM who is myopic and, with imperfect knowledge about herself, cares about her self-images (ego preferences). To be more precise, the subjective self postulates that her ability type is

many factors, including those that are not ego-threatening: for instance, the accuracy is most likely to depend on environmental factors that are purely external to DM. This informativeness is a formal representation of those external factors that are not ego-threatening.

[14] One of the first attempts to model an individual with divided selves is provided by Thaler and Shefrin (1981). Recent attempts to model divided selves are provided by Lowenstein and O'donoghue (2005) and Fudenberg and Levine (2006). For a different approach, see Brocas and Carrillo (2008) who model the brain as a hierarchical organization.

[15] The assumption is strictly for simplicity, as the role of the discount factor, even if it is introduced, is fairly straightforward.

drawn from some distribution F with mean $\mu := \int \eta dF$ and variance $\sigma^2 := \int \eta^2 dF - \mu^2$.[16] We take that the prior mean reflects DM's initial self-confidence level while the variance reflects its fragility: we say that DM is more secure about her self-images when σ^2 is smaller and close to zero.

When estimating her ability type, the subjective self uses the subjective informativeness. Define

$$\mu_1(s_1;\gamma_1) := E[\eta \mid s_1, \gamma_1],\ \mu_2(s_1, s_2;\gamma_1;\gamma_2) := E[\eta \mid s_1, s_2, g_1, g_2],$$

as the belief about η. Being myopic, the subjective self totally disregards the value of the project which realizes in the future and instead cares only about the immediate gain, i.e., her self-images at each moment.[17] The subjective self's payoff in each period is thus simply given by

$$\pi_t^{\mathrm{S}} = \mu_1(s_1;\gamma_1),\ \pi_t^{\mathrm{S}} = \mu_2(s_1, s_2;\gamma_1;\gamma_2).$$

2.4 The intrapersonal conflict

The subjective informativeness is chosen so as to maximize the overall payoff π_t, which is defined as a weighted average of π_t^{S} and π_t^{O}:

$$\pi_t = \beta\pi_t^{\mathrm{S}} + \pi_t^{\mathrm{O}}.$$

Here, $\beta > 0$ is a parameter which measures the strength of DM's willpower: she is regarded as possessing strong willpower to regulate herself (or the subjective self) when β is small and close to zero. Note that this parameter conflates two fundamentally distinct elements: the weight given to the subjective self (which determines the severity of time inconsistency) and the strength of ego preferences. The parameter β is hence a concept which encompasses these two aspects that happen to play the same role in this particular setup.

[16]This specification implies that the subjective self's prior mean coincides with its true value, although there is no strong reason why this must be so. While we make this assumption mainly for analytical tractability, we argue that this is also realistic because actual ability and self-confidence are generally positively correlated, e.g., Stankov and Crawford (1996). In Appendix B, we relax this assumption to show that our main insights can be obtained even if the prior mean deviates from the true value.

[17]The current specification means that the subjective self cares only about the mean of the ability estimate but not about the variance. If the subjective self prefers a more (less) accurate estimate for some reasons, that tends to raise (lower) both γ_1 and γ_2.

Formally, we divide each period into two stages and suppose that DM goes through the following two-stage process:

Stage 1 (the information-processing stage): Upon observing s_t, the objective and subjective selves assign γ_t to it so as to maximize the overall payoff π_t. The subjective informativeness chosen in this stage then yields a compromised belief about the state of nature.

Stage 2 (the action stage): The objective self chooses the optimal action based on the compromised belief.

Being the only one to care about the outcome of the project, an action in each period is chosen deliberately by the objective self.[18] This choice is made, however, based on the compromised belief because, once γ_t is imprinted, deviating from that produces cognitive dissonance. This means that, although DM can freely choose any γ_t, that will become a part of her perception and influence the subsequent choice of action: in this particular sense, the beliefs are internally consistent and coherent.

This internal process eventually amounts to a constrained optimization problem in each period. In period 2, taking (s_1, s_2) and γ_1 as given,

$$\max_{\gamma_2} \ \pi_2 = \beta \mu_2(s_1, s_1; \gamma_1, \gamma_2) + R(a_2, \tilde{\rho}_2),$$

subject to

$$a_2 = \rho_2(s_1, s_2; \gamma_1, \gamma_2).$$

Let $\gamma_2^*(s_1, s_2; \gamma_1)$ denote the optimal solution to this problem which possibly depends on (s_1, s_2) and γ_1. Then, in period 1, taking s_1 as given,

$$\max_{\gamma_1} \ \pi_1 = \beta \mu_1(s_1; \gamma_1) + R(a_1, \tilde{\rho}_1) + E[R(a_2, \tilde{\rho}_2) \mid s_1, \gamma_1],$$

subject to

$$a_1 = \rho_1(s_1; \gamma_1), \ a_2 = \rho_2(s_1, s_2; \gamma_1, \gamma_2^*(s_1, s_2; \gamma_1)).$$

[18]One can regard Stage 1 as an automatic process which operates at a subconscious level and is inaccessible to consciousness. In contrast, Stage 2 may be regarded as a controlled process though which deliberate and conscious choices are made. See Camerer et al. (2005) for the distinction between automatic and controlled processes in neural functioning.

In what follows, the entire process is described as if DM singlehandedly solves this constrained problem. Other than the interpretation detailed just above, this setup yields a wide variety of interpretations as to the structure behind it. In any event, though, the bottom line is the tradeoff between distorted beliefs and distorted actions and any structure that can (at least partially) capture this aspect can suffice for our purposes.[19]

3 Optimal information processing

3.1 Preliminary

Before we move on, we first describe in some depth how the beliefs are formed in this model. First, the objective self must form a belief about the state of nature ρ_t. Given some informativeness g_1, the best estimate of θ in period 1 is obtained as

$$\rho_1(1; g_1) = \frac{1 + g_1\mu}{2}. \tag{1}$$

Due to the symmetric nature of the states, $\rho_1(0; g) = 1 - \rho_1(1; g)$ for any g. Since the states are symmetric, we shall hereafter focus on $s_1 = 1$ without loss of generality. The manipulable range of the belief is depicted in figure 1.

[Figure 1]

With an additional signal in period 2, the belief about θ is updated and given by

$$\rho_2(1, 1; g_1, g_2) = \frac{(1 + g_2)(1 + g_1\mu)}{2((1 - g_2)(1 - g_1\mu) + (1 + g_2)(1 + g_1\mu))} = \frac{(1 + g_2)(1 + g_1\mu)}{2(1 + g_1 g_2 \mu)}. \tag{2}$$

$$\rho_2(1, 0; g_1, g_2) = \frac{(1 - g_2)(1 + g_1\mu)}{2((1 + g_2)(1 - g_1\mu) + (1 - g_2)(1 + g_1\mu))} = \frac{(1 - g_2)(1 + g_1\mu)}{2(1 - g_1 g_2 \mu)}. \tag{3}$$

To ease notation, define

$$\rho_C(g_1, g_2) := \rho_2(1, 1; g_1, g_2) \text{ and } \rho_I(g_1, g_2) := \rho_2(1, 0; g_1, g_2).$$

The manipulable range of the belief is depicted in figure 2.

[19] For instance, we could model this process as an intrapersonal game (or bargaining) between the two selves. An advantage of the current specification over potential alternatives is its greater tractability, as it leads to a closed-form solution for the second-period problem. In any case, we would expected to reach similar conclusions as long as a model captures the tradeoffs considered here.

[Figure 2]

Being uncertain about her own ability, on the other hand, the subjective self must form a belief about η. The inference is only relevant in period 2 because one signal conveys no information about η. The beliefs in period 2 are obtained as

$$\mu_2(1,1;\gamma_1,\gamma_2) = \frac{\gamma_1\gamma_2\int_0^1 \eta(1+\eta)dF + (1-\gamma_1\gamma_2)\mu}{\gamma_1\gamma_2\int_0^1 (1+\eta)dF + (1-\gamma_1\gamma_2)} = \frac{\mu+\gamma_1\gamma_2 E[\eta^2]}{1+\gamma_1\gamma_2\mu}, \tag{4}$$

$$\mu_2(1,0;\gamma_1,\gamma_2) = \frac{\gamma_1\gamma_2\int_0^1 \eta(1-\eta)dF + (1-\gamma_1\gamma_2)\mu}{\gamma_1\gamma_2\int_0^1 (1-\eta)dF + (1-\gamma_1\gamma_2)} = \frac{\mu-\gamma_1\gamma_2 E[\eta^2]}{1-\gamma_1\gamma_2\mu}. \tag{5}$$

Similarly as above, define

$$\mu_C(\gamma_1,\gamma_2) := \mu_2(1,1;\gamma_1,\gamma_2) \text{ and } \mu_I(\gamma_1,\gamma_2) := \mu_2(1,0;\gamma_1,\gamma_2).$$

Notice that $\mu_C > \mu > \mu_I$ for any $\gamma_1\gamma_2 > 0$ and $\sigma^2 > 0$, i.e., consistent signals are a good news for DM while inconsistent signals are a bad news.

3.2 Flexibility: the second-period problem

In period 2, DM observes an additional signal, which may nor may not be consistent with the first signal. DM's flexibility, which is defined as the ability to respond rationally and objectively to new information, is tested in this situation. When DM is prejudiced by the first signal and fails to respond objectively to the second signal, the consequent action is distorted and becomes necessarily inefficient. Here, we examine when and to what extent DM exhibits this type of behavioral bias.

In period 2, DM observes s_2 and assigns the subjective informativeness to it, taking (s_1, s_2) and γ_1 as given. Choosing γ_2 then yields a compromised belief $\rho_2(s_1,s_2;\gamma_1,\gamma_2)$, which corresponds to the action chosen by the objective self. The optimization problem is thus defined as

$$\max_{\gamma_2} \quad \beta\mu_2(s_1,s_2;\gamma_1,\gamma_2) + R(\rho_2(s_1,s_2;\gamma_1,\gamma_2), \rho_2(s_1,s_2;\tilde{\gamma},\tilde{\gamma})).$$

The first-order condition is given by

$$\beta\frac{\partial\mu_2}{\partial\gamma_2} - 2(\rho_2(s_1,s_2;\gamma_1,\gamma_2) - \rho_2(s_1,s_2;\tilde{\gamma},\tilde{\gamma}))\frac{\partial\rho_2}{\partial\gamma_2} = 0,$$

assuming that there exists an interior solution. The solution is obtained as a function of γ_1 and denoted as $\rho_2^*(s_1, s_2; \gamma_1)$. Given some optimal belief $\rho_2^*(s_1, s_2; \gamma_1)$, we refer to $b_2(s_1.s_2; \gamma_1) := \rho_2^*(s_1, s_2; \gamma_1) - \tilde{\rho}_2(s_1, s_2; \tilde{\gamma}, \tilde{\gamma})$ as the optimal bias as perceived by the objective self. We take this as a measure of flexibility where DM is less prejudiced and hence more flexible when the bias is smaller.

When the signals are consistent, this condition becomes

$$\frac{\beta\gamma_1\sigma^2}{(1+\gamma_1\gamma_2\mu)^2} - (\rho_2 - \tilde{\rho}_C)\frac{1-(\gamma_1\mu)^2}{(1+\gamma_1\gamma_2\mu)^2} = 0,$$

where $\tilde{\rho}_C := \rho_C(\tilde{\gamma}, \tilde{\gamma})$. The optimal belief ρ_C^* in this contingency as a function of γ_1 is given by

$$\rho_C^*(\gamma_1) - \tilde{\rho}_C = \frac{\beta\sigma^2\gamma_1}{1-(\gamma_1\mu)^2}.$$

It follows from this that $\rho_C^* > \tilde{\rho}_C$ for any $\gamma_1 > 0$, meaning that DM exaggerates her information even more when the observed signals are consistent.

Similarly, when the signals are inconsistent, this condition becomes

$$-\frac{\beta\gamma_1\sigma^2}{(1-\gamma_1\gamma_2\mu)^2} + (\rho_2 - \tilde{\rho}_I)\frac{1-(\gamma_1\mu)^2}{(1-\gamma_1\gamma_2\mu)^2} = 0,$$

where $\tilde{\rho}_I := \rho_I(\gamma_1, \tilde{\gamma})$. The optimal belief ρ_I^* as a function of γ_1 is then given by

$$\rho_I^*(\gamma_1) - \tilde{\rho}_I = \frac{\beta\sigma^2\gamma_1}{1-(\gamma_1\mu)^2},$$

which again implies that $\rho_I^* > \tilde{\rho}_I$ for any γ_1. This means that DM undervalues the second signal and instead favors the first one in order to justify her prior stance. Note also that the optimal bias is symmetric, i.e., $b_2(1,0;\gamma_1) = b_2(1,1;\gamma_1)$ for any γ_1. Define $b_2^*(\gamma_1) := b_2(1,0;\gamma_1) = b_2(1,1;\gamma_1)$, which can be seen as a measure of DM's flexibility.

Proposition 1 *The optimal second-period bias is always positive and given by*

$$b_2^*(\gamma_1) = \frac{\beta\sigma^2\gamma_1}{1-(\gamma_1\mu)^2}.$$

For any given γ_1, the bias is increasing in β, σ^2 and μ. Moreover, the bias is increasing in γ_1.

PROOF: Most of the proposition is self-evident from the discussion made thus far. We only show that the optimal bias is increasing in γ_1. This is the case if

$$\frac{\beta\sigma^2(1+(\gamma_1\mu)^2)}{(1-(\gamma_1\mu)^2)^2} > 0, \tag{6}$$

which always holds.

Q.E.D.

Several remarks are in order. First, the optimal second-period bias is always positive, meaning that DM biases her decision in the direction of the first signal (the anchoring effect). There are two sides to this biased information processing. When the observed signals are consistent, DM overreacts to the second signal, becoming rather obsessive about her initial stance. As a flip side, when the observed signals are inconsistent, she downplays the importance of the second signal and sticks stubbornly with her initial stance. In either case, with imperfect willpower $\beta > 0$ and imperfect knowledge about herself $\sigma^2 > 0$, DM in general exhibits a degree of inflexibility as she fails to respond rationally and objectively to new information.

Second, taking γ_1 as given, the size of the bias depends on attributes such as the self-confidence level and fragility as well as the strength of willpower. The effects of the self-confidence fragility and the strength of willpower are identical and relatively straightforward. The bias is smaller when DM is more secure about herself and/or she has strong willpower to regulate herself: in fact, when DM is perfectly secure about herself ($\sigma^2 = 0$) or has perfect willpower ($\beta = 0$), the intrapersonal conflict disappears and she exhibits no bias whatsoever.[20] What is more interesting but intuitively less straightforward is perhaps the effect of the self-confidence level. The analysis reveals that DM who has more confidence in herself is less likely to reverse her prior decisions and hence more prone to exhibiting inflexibility in both ways. This suggests that although high ability itself is helpful in making good judgements, high self-confidence is often an obstacle to making objective decisions. This result

[20]As an alternative interpretation, the self-confidence fragility can be taken as reflecting how the predictability of the underlying environment is related to DM's ability. For instance, if the accuracy of the first signal is less ability-intensive and related only weakly to η, not much is revealed about her ability type. The situation corresponds to a case with a smaller σ^2.

comes from the fact that the accuracy of the first signal is more sensitive to ability than that of the second signal. Since DM with high self-confidence trusts her first observation more than that with low self-confidence, the cost of biasing the interpretation of the second signal is lower in a relative sense. This biases her information processing in order to protect her self-images, making inflexibility more of a problem for DM with high self-confidence.

Finally, it is important to note that the optimal bias is also a function of an endogenous choice variable γ_1. The bias is larger when DM puts more stock in the first signal and decreases as γ_1 decreases, meaning that vision and flexibility are substitutes. In particular, when $\gamma_1 = 0$, i.e., DM believes that the first signal is not at all informative, nothing she observes undermines her self-confidence, and there is hence no need to bias her decision making in order to justify herself. This aspect gives rise to critical dynamic interactions, because the objective self in period 1 can reasonably control the subjective self in period 2 by disregarding the first signal and remaining rather ambiguous. This is precisely the problem faced by DM in period 1, which we will next turn to.

3.3 Vision: the first-period problem

The prediction ability is tested more in period 1, as the accuracy of the first signal depends positively on DM's ability. In this setting, vision is defined as the ability to respond rationally and objectively to early information before things unfold themselves. The key factor at this stage is the presence of time-inconsistent objectives, which stems from the difference in time horizon between the two selves. The objective self knows that her second-period prediction will be biased, to the detriment of her own interests, to accommodate the subjective self's ego preferences. The objective self thus has an incentive to minimize this bias, if it is possible at all. There is indeed a way to achieve this because, as we have seen, the optimal bias in period 2 depends positively on the informativeness of the first signal γ_1. Disregarding the first signal frees the future self from any concern about her self-images and allows her to react more objectively to the second signal, thereby minimizing the prediction bias which works against the objective self's interests.[21] There is of course a cost associated with it, because a failure

[21] Alternatively, exhibiting a vision (by placing a larger weight on the first signal) is a commitment to limit the set of available future choices. The current framework thus naturally and endogenously gives rise to the tradeoff

to objectively utilize the first signal generally results in an inefficient action in period 1. The optimal bias in period 1 is determined by this tradeoff.

Since γ_1 and γ_2 are intertemporally linked, DM can influence her future self through the choice of γ_1. The objective self's expected payoff in period 2 as a function of γ_1 is given by

$$\Pi_2^O(\gamma_1) = \frac{1+\tilde{\gamma}^2\mu}{2}\pi_C(\gamma_1) + \frac{1-\tilde{\gamma}^2\mu}{2}\pi_I(\gamma_1),$$

where

$$\pi_C(\gamma_1) := R(\rho_C^*(\gamma_1), \tilde{\rho}_C) = -\tilde{\rho}_C(1-\rho_C^*(\gamma_1)) - (1-\tilde{\rho}_C)\rho_C^*(\gamma_1),$$

$$\pi_I(\gamma_1) := R(\rho_I^*(\gamma_1), \tilde{\rho}_I) = -\tilde{\rho}_I(1-\rho_I^*(\gamma_1)) - (1-\tilde{\rho}_I)\rho_I^*(\gamma_1).$$

Upon observing $s_1 = 1$, DM's problem in period 1 is defined as

$$\max_{\gamma_1} \quad \beta\mu + R(\rho_1(1;\gamma_1), \rho_1(1;\tilde{\gamma})) + \Pi_2^O(\gamma_1).$$

Notice that the myopic subjective self has nothing at stake in period 1, so that the problem is strictly about how the objective self regulates the future subjective self.

Define $\tilde{\rho}_1 := \rho_1(1;\tilde{\gamma})$ and $b_1^*(\gamma_1) := \rho_1(1;\gamma_1) - \tilde{\rho}_1$ where we take b_1^* as a measure of DM's vision: the larger b_1^* is, the more visionary she is.[22] The first-order condition is then given by

$$-2b_1^*(\gamma_1)\frac{\partial\rho_1}{\partial\gamma_1} + \frac{1+\tilde{\gamma}^2\mu}{2}\frac{\partial\pi_C}{\partial\gamma_1} + \frac{1-\tilde{\gamma}^2\mu}{2}\frac{\partial\pi_I}{\partial\gamma_1} = 0. \tag{7}$$

Since

$$\frac{\partial\pi_C}{\partial\gamma_1} = \frac{\partial\pi_I}{\partial\gamma_1} = -2b_2^*(\gamma_1)\frac{d\rho_C^*}{d\gamma_1},$$

(7) can be written as

$$b_1^*(\gamma_1) = R(\gamma_1)b_2^*(\gamma_1),$$

where

$$R(\gamma_1) := -\frac{\frac{d\rho_C^*}{d\gamma_1}}{\frac{\partial\rho_1}{\partial\gamma_1}} = -\frac{2\beta\sigma^2(1+(\gamma_1\mu)^2)}{\mu(1-(\gamma_1\mu)^2)^2},$$

between commitment and flexibility, like the one analyzed in Amador et al. (2006) and some others.

[22]Evidently, there is a right degree of vision, i.e., more vision is not necessarily better. DM indeed takes an excessively strong stance when $b_1^* > 0$ although, as we will see later, this would not occur in equilibrium.

Proposition 2 *The optimal first-period bias is always negative. The bias is increasing in α, β and σ^2 and disappears as they tend to zero, whereas it is decreasing in μ if*

$$1 - 4(\tilde{\gamma}\mu)^2 - 3(\tilde{\gamma}\mu)^4 > 0.$$

Moreover, $\gamma_1 = 0$ (the complete lack of vision) if μ is sufficiently small.

PROOF: See Appendix.

The optimal first-period bias is always negative, in contrast to the second-period bias, where DM biases her prediction towards her prior mean (excessive reliance on precedents) and strategically settles for an ambiguous initial stance in order to allow her future self to act more flexibly. This is taken as a sign of the lack of vision. As in period 2, the bias is increasing in β and σ^2, because the need to regulate the future self is simply larger. This indicates that DM fails to articulate her vision when she is insecure about herself and/or lacks willpower to regulate herself.

It is, on the other hand, more complicated to see how a change in μ affects γ_1. On one hand, an increase in μ enlarges the second-period bias for any given γ_1, which raises the marginal gain of lowering γ_1. On the other hand, there is also a countervailing effect. To see this, recall that an increase in μ tends to make DM more inflexible in period 2 because with the first signal being more accurate, the cost of disregarding the second signal is relatively small. As can be imagined, the situation is totally reversed in period 1: the cost of disregarding the first signal is now larger for DM with high self-confidence. While the optimal bias could go either way due to this tradeoff, it is decreasing in μ when $\tilde{\gamma}$ is sufficiently small,[23] because the second-period bias becomes negligibly small, and the need for self-regulation diminishes in this case. Then, since the cost of disregarding the first signal is larger, the first-period bias becomes smaller for DM with high self-confidence who articulates a clearer vision.

The effect of the self-confidence level can also be seen from an alternative perspective. Notice that, as $\mu \to 0$, the choice of γ_1 becomes totally irrelevant, because the signal is not

[23] The bias is in fact decreasing in μ for all $\mu \in [0, 1]$ if $\tilde{\gamma} < \bar{\gamma} \approx 0.464$. Of course, this is only a sufficient condition, indicating that the bias is decreasing in μ for a wider range of parameter values.

informative anyway, no matter how she interprets it: in other words, there is little Bayesian updating when there is little information. Since the cost of biasing the interpretation is vanishingly small, there is no reason to assign any positive informativeness to the first signal. As the proposition indicates, the optimal response is the complete lack of vision ($\gamma_1 = 0$) for DM with sufficiently low self-confidence. When that happens, DM totally disregards early information and does not make up her mind until things become sufficiently clear.

3.4 The overall impact on flexibility

In section 3.2, we have discussed how the second-period bias b_2^* varies in response to changes in parameters such as β, σ^2 and μ when γ_1 is exogenously given and fixed. Since γ_1 itself is a function of those parameters, however, indirect effects through γ_1 must also be taken into account to fully understand the overall impact on flexibility. As suggested, vision and flexibility are substitutes in that when one goes up, the other must go down. Changes in exogenous parameters also affect the relative value of vision and flexibility, as captured by $R(\gamma_1)$, thereby yielding some indirect effects. Although our primary focus is on the direct effects, we here briefly discuss the overall impact of exogenous parameters on the degree of flexibility.

We first examine the impact of changes in β or σ^2. Since these two parameters yield exactly the same effect, we focus on changes in β. As we have seen, an increase in β raises γ_1 and, taking γ_1 as given, also γ_2. At the same time, though, a change in β also changes the relative value of vision and flexibility. While the bias increases up to some point with an increase in β, the cost of inflexibility rises as β becomes sufficiently large. Once β reaches some point, therefore, the first-period bias becomes less of a concern, and DM attempts to minimize the second-period bias by decreasing γ_1 towards zero. This indicates that the impact of changes in β (or σ^2 as well) on the second-period bias is non-monotonic, where the bias disappears when β is either too small or too large.

Second, the impact of changes in δ is relatively straightforward. Because it yields no direct effect on the second-period bias, we only need to look at the indirect effect though γ_1. As we have seen, an increase in δ decreases γ_1 away from $\tilde{\gamma}$. This in turn monotonically decreases the second-period bias, making DM more flexible.

Finally, the impact of changes in μ is somewhat more complicated, because the direct effect can in principle go either way. As we have seen, though, the second-period bias is decreasing in μ for a wide range of parameter values. In that case, we can obtain an unambiguous prediction because both the direct and indirect effects work in the same direction. For a sufficiently low value of μ, DM is perfectly flexible although it comes at the expense of the complete lack of vision. An increase in μ then unambiguously raises the second-period bias, where DM becomes less and less flexible. We can thus summarize this finding as follows: while high self-confidence *per se* makes DM more visionary, it makes her less flexible in that she sticks excessively to her initial stance.

4 Conclusion

The paper sheds light on two important aspects of dynamic decision making – vision and flexibility – and provides a framework to illustrate the workings behind these seemingly contradictory notions. As the misalignment of incentives is a typical source of inefficient decision making in organizations, the driving force of the model is the misalignment of incentives within oneself. In the present model, two distinct sources of intrapersonal conflicts are merged into a single framework: on one hand, the objective and subjective selves differ in their objectives, which eventually forces DM to bend the truth to protect her ego; on the other hand, the two selves also differ in their time horizon, thus leading to time-inconsistent objectives and, as a direct consequence, a type of of self-control problem. Our model provides a unified framework to account for both the conformity bias and the confirmatory bias in dynamic decision making.

As a final note, since we set up a tightly specified model to obtain clear predictions, there are some obvious limitations of the analysis. One such limitation is the fact that we only consider a two-period model. Although the two-period framework is just sufficient to illuminate the two cognitive biases of our interest, an extension to a multi-period setup would certainly yield a richer set of dynamic implications. We expect, however, that similar properties would eventually emerge in that framework.[24] To see this, consider a three-period counterpart. First, in the last period, DM still has an incentive to exaggerate consistent signals and discount inconsistent ones, although the meanings of consistent/inconcistent are now history-dependent.[25] This provides an incentive to discount the signal in the first period to mitigate this bias. This extension to a multi-period setup, however, adds a new dimension to the problem in the second period where both of these forces are present. It would be highly interesting to see how these opposing forces balance out in this intermediate period.

Appendix A: Proof of Proposition 2

Since

$$\frac{\partial \rho_1}{\partial \gamma_1} = \frac{\mu}{2},$$

$$\frac{\partial \rho_C^*}{\partial \gamma_1} = \frac{\partial \rho_I^*}{\partial \gamma_1} = \frac{\beta\sigma^2(1+(\gamma_1\mu)^2)}{(1-(\gamma_1\mu)^2)^2},$$

the first-order condition becomes

$$\frac{(\gamma_1-\tilde{\gamma})\mu^2}{4} + \frac{\beta\sigma^2\gamma_1}{1-(\gamma_1\mu)^2}\frac{\beta\sigma^2(1+(\gamma_1\mu)^2)}{(1-(\gamma_1\mu)^2)^2} = 0.$$

Define

$$D(\beta,\sigma^2,\mu,\gamma_1) := \frac{4(\beta\sigma^2)^2(1+(\gamma_1\mu)^2)\gamma_1}{\mu^2(1-(\gamma_1\mu)^2)^3},$$

so that the optimal informativeness is obtained as the following fixed point:

$$\gamma_1 = \tilde{\gamma} - D(\beta,\sigma^2,\mu,\gamma_1).$$

It is immediately clear from this that $\gamma_1 < \tilde{\gamma}$. The properties of γ_1 depend on how D respond to changes in exogenous parameters such as β, σ^2 and μ.

[24] Of course, the equilibrium properties depend crucially on the minute details of the information structure, especially how the accuracy of each signal is related to the ability type. A sensible specification in a T-period setup might have prob$\{s_t = \theta \mid \text{the signal is informative}, \eta\} = \lambda_t + (1-\lambda_t)(1+\eta)/2$ where λ_t is aligned in a non-decreasing order, i.e., $1 \geq \lambda_T \geq \lambda_{T-1} \geq \ldots \geq \lambda_1 \geq 0$. The current two-period setup is a special case of this with $\lambda_1 = 0$ and $\lambda_2 = 1$.

[25] More precisely, they depend both on the history of observed signals and associated choices of informativeness.

We now examine how γ_1 responds to changes in β, σ^2 and μ. It is easy to see that $\partial D/\partial\beta > 0$, implying that γ_1 is decreasing and the size of the bias is hence increasing in β. By the same logic, one can show that the size of the bias is increasing in σ^2. The effect of μ is, on the other hand, more complicated. To see this, note that $\partial D/\partial\mu < 0$ if

$$(\gamma_1\mu)^2(1-(\gamma_1\mu)^2) - (1+(\gamma_1\mu)^2)(1-4(\gamma_1\mu)^2) < 0.$$

which can be written as

$$1 - 4(\gamma_1\mu)^2 - 3(\gamma_1\mu)^4 > 0.$$

Since we already know that $\gamma_1 < \tilde{\gamma}$, a sufficient condition for this is

$$1 - 4(\tilde{\gamma}\mu)^2 - 3(\tilde{\gamma}\mu)^4 > 0.$$

When this condition holds, the size of the bias is decreasing in μ.

Finally, we show that $\gamma_1 \to 0$ as $\mu \to 0$. This is relatively straightforward since

$$\lim_{\mu\to 0} D(\beta,\sigma^2,\mu,\gamma_1) = \infty,$$

for any $\gamma_1 > 0$. An interior solution thus fails to exist and the optimal choice is bounded at $\gamma_1 = 0$.

Q.E.D.

Appendix B

In the main body of the analysis, we have assumed that the subjective self's prior expectation of the ability type coincides with its true value. This assumption not only makes the analysis more tractable but also allows us to obtain more realistic implications since actual ability and self-confidence are generally positively correlated.[26] In principle, though, there is no convincing reason why the prior mean must coincide with the true value. Here, we relax this assumption to show that the main results of the paper basically go through with this alteration to the baseline model.

[26] Our comparative static results regarding changes in μ are also valid to the extent that actual ability and self-confidence are positively correlated.

Since the subjective self is totally disinterested in period 1, the first-order condition in period 1 is virtually identical, so that similar conclusions would follow if the basic properties in period 2 were to hold. Now suppose that the subjective self's prior mean is fixed at μ^{S} which may or may not coincide with μ. With this differing prior, the first-order condition in period 2 is modified as

$$\frac{\beta\sigma^2\gamma_1}{(1+\gamma_1\gamma_2\mu^{\mathrm{S}})^2} - (\rho_2 - \tilde{\rho}_C)\frac{1-(\gamma_1\mu)^2}{(1+\gamma_1\gamma_2\mu)^2} = 0, \tag{8}$$

for consistent signals and

$$-\frac{\beta\sigma^2\gamma_1}{(1-\gamma_1\gamma_2\mu^{\mathrm{S}})^2} + (\rho_2 - \tilde{\rho}_I)\frac{1-(\gamma_1\mu)^2}{(1-\gamma_1\gamma_2\mu)^2} = 0, \tag{9}$$

for inconsistent signals. It is evident from these that the same conclusions hold in this case as well: consistent signals are exaggerated while inconsistent ones are discounted, and the bias disappears as $\beta \to 0$ or $\sigma^2 \to 0$.

What remains to be seen is the impact of γ_1 on the optimal bias. To see this, from (8), we obtain

$$\rho_2 - \tilde{\rho}_C = d_C(\gamma_1,\gamma_2)^2 b_2^*(\gamma_1).$$

where

$$d_C(\gamma_1,\gamma_2) := \frac{1+\gamma_1\gamma_2\mu}{1+\gamma_1\gamma_2\mu^{\mathrm{S}}}.$$

We need to show that the optimal bias is increasing in γ_1. Taking γ_2 as a constant, it suffices to show that

$$2d_C\frac{\partial d_C}{\partial\gamma_1}b_2^* + d_C\frac{db_2^*}{d\gamma_1} = 2d_C\frac{\gamma_2(\mu-\mu^{\mathrm{S}})}{(1+\gamma_1\gamma_2\mu^{\mathrm{S}})^2}b_2^* + d_C\frac{db_2^*}{d\gamma_1} > 0,$$

which can be written as

$$\frac{2\gamma_1\gamma_2(\mu-\mu^{\mathrm{S}})}{(1+\gamma_1\gamma_2\mu^{\mathrm{S}})^2} + \frac{1+(\gamma_1\mu)^2}{1-(\gamma_1\mu)^2} > 0.$$

This condition holds if

$$\frac{2\gamma_1\gamma_2(\mu-\mu^{\mathrm{S}})}{(1+\gamma_1\gamma_2\mu^{\mathrm{S}})^2} > -1 \;\Leftrightarrow\; 2\gamma_1\gamma_2\mu > -1-(\gamma_1\gamma_2\mu^{\mathrm{S}})^2.$$

Similarly, it follows from (9) that

$$\rho_2 - \tilde{\rho}_I = d_I(\gamma_1, \gamma_2)^2 b_2^*(\gamma_1).$$

where

$$d_I(\gamma_1, \gamma_2) := \frac{1 - \gamma_1\gamma_2\mu}{1 - \gamma_1\gamma_2\mu^{\mathrm{S}}}.$$

To show that the optimal bias is increasing in γ_1, it suffices to show that

$$2d_I \frac{\partial d_I}{\partial \gamma_1} b_2^* + d_I \frac{db_2^*}{d\gamma_1} = 2d_I \frac{\gamma_2(\mu^{\mathrm{S}} - \mu)}{(1 + \gamma_1\gamma_2\mu^{\mathrm{S}})^2} b_2^* + d_I \frac{db_2^*}{d\gamma_1} > 0,$$

which can be written as

$$\frac{2\gamma_1\gamma_2(\mu^{\mathrm{S}} - \mu)}{(1 - \gamma_1\gamma_2\mu^{\mathrm{S}})^2} + \frac{1 + (\gamma_1\mu)^2}{1 - (\gamma_1\mu)^2} > 0.$$

With some algebra we obtain

$$(2\gamma_1\gamma_2(\mu^{\mathrm{S}} - \mu))(1 - (\gamma_1\mu)^2) + (1 - \gamma_1\gamma_2\mu^{\mathrm{S}})^2(1 + (\gamma_1\mu)^2) > 0,$$

which is further simplified to

$$2\gamma_1\gamma_2(\gamma_1\mu)^2(2\mu^{\mathrm{S}} + \mu) + (1 + (\gamma_1\gamma_2\mu^{\mathrm{S}})^2)(1 + (\gamma_1\mu)^2) - 2\gamma_1\gamma_2\mu > 0.$$

This condition holds since

$$1 + (\gamma_1\mu)^2 > 2\gamma_1\mu \geq 2\gamma_1\gamma_2\mu.$$

This shows that the optimal bias in period 2 is increasing in γ_1 even when the subjective self's prior mean diverts away from the true value. This property gives rise to the same self-control problem as discussed in the main body, leading to qualitatively similar conclusions.

References

Akerlof, Geroge A. and William T. Dickens, 1982, The Economic Consequences of Cognitive Dissonance, *American Economic Review*, 72, 307-19.

Amador, Manuel; Ivan Werning and George-Marios Angeletos, 2006, Commitment vs. Flexibility, *Econometrica*, 74, 365-396.

Ariely, Dan; George Loewenstein and Drazen Prelec, 2003, "Coherent Arbitrariness": Stable Demand Curves without Stable Preferences, *Quarterly Journal of Economics*, 118, 73-105.

Asch, Solomon E., 1951, Effects of Group Pressure upon the Modification and Distortion of Judgements, in Guetzhow, H. ed., *Groups, Leadership and Men*, Pittsburgh, PA: Carnegie Press.

Babcock, Linda and George Loewenstein, 1997, Explaining Bargaining Impasse: the Role of Self-Serving Biases, *Journal of Economic Perspectives*, 11, 109-26.

Benabou, Roland, 2008a, Ideology, *Journal of the European Economic Association*, 6, 321-52.

Benabou, Roland, 2008b, Groupthink: Collective Delusions in Organizations and Markets, mimeo, Princeton University.

Benabou, Roland and Jean Tirole, 2002, Self-Cinfidence and Personal Motivation, *Quarterly Journal of Economics*, 117, 871-915.

Benabou, Roland and Jean Tirole, 2004, Willpower and Personal Rules, *Journal of Political Economy*, 112, 848-86.

Benabou, Roland and Jean Tirole, 2006, Belief in a Just World and Redistributive Politics, *Quarterly Journal of Economic*, 121, 699-746.

Bracha, Anat and Donald J. Brown, 2009, Affective Decision Making: A Behavioral Theory of Choice, mimeo.

Brocas, Isabelle and Juan D. Carrillo, 2008, The Brain as a Hierarchical Organization, *American Economic Review*, 98, 1312-46.

Camerer, Colin; George Loewenstein and Drazen Prelec, 2005, Neuroeconomics: How Neuroscience Can Inform Economics, *Journal of Economic Literature*, 43, 9-64.

Carrillo, Juan D. and Thomas Mariotti, 2000, Strategic Ignorance as a Self-Disciplining Device, *Review of Economic Studies*, 67, 529-544.

Compte, Olivier and Andrew Postlewaite, 2004, Confidence-Enhanced Performance, *American Economic Review*, 94, 1536-57.

Crary, William G., 1966, Reactions to Incongruent Self-Experiences, *Journal of Consulting Psychology*, 30, 246-52.

Falk, Armin; Ernst Fehr and Christian Zehnder, 2006, Fairness Perceptions and Reservation Wages – The Behavioral Effects of Minimum Wage Laws, *Quarterly Journal of Economics*, 121, pp. 1347-81.

Köszegi, Botond, 2006, Ego Utility, Overconfidence, and Task Choice, *Journal of the European Economic Association*, 4, 673-707.

Li, Wei, 2007, Changing One's Mind when the Facts Change: Incentives of Experts and the Design of Reporting Protocols, *Review of Economic Studies*, 74, 1175-94.

Mayraz, Guy, 2009, Priors and Desires: A Model of Payoff-Dependent Beliefs, mimeo.

Prendergast, Canice and Lars Stole, 1996, Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning, *Journal of Political Economy*, 104, 1105-34.

Pyszczynski, Tom; Jeff Greenberg and Kathleen Holt, 1985, Maintaining Consistency between Self-Serving Beliefs and Available Date: a Bias in Information Processing, *Personality and Social Psychology Bulletin*, 11, 179-90.

Rabin, Matthew and Joel L. Schrag, 1999, First Impressions Matter: A Model of Confirmatory Bias, *Quarterly Journal of Economics*, 114, 37-82.

Rotemberg, Julio J. and Garth Saloner, 2000, Visionaries, Managers, and Strategic Direction, *RAND Journal of Economics*, 31, 693-716.

Santos-Pinto, Luis and Joel Sobel, 2005, A Model of Positive Self-Image in Subjective Assessment, *American Economic Review*, 95, 1386-1402.

Stankov, Lazar and John D. Crawford, 1996, Confidence Judgements in Studies of Individual Differences, *Personality and Individual Differences*, 21, 971-86.

Thaler, Richard H. and H.M. Shefrin, 1981, A Theory of Self-Control, *Journal of Political Economy*, 89, 392-406.

Van den Steen, Eric, 2004, Rational Overoptimism (and Other Biases), *American Economic Review*, 94, 1141-51.

Van den Steen, Eric, 2005, Organizaitonal Beliefs and Managerial Vision, *Journal of Law, Economics, and Organization*, 21, 256-83.

Wyer, Robert S. and Frey, Dieter, 1983, The Effects of Feedback about Self and Others on the Recall and Judgements of Feedback-Relevant Information, *Journal of Experimental Social Psychology*, 19, 540-59.

Zakarikis, Andrew L.; G. Dale Meyer and Julio DeCastro, 1999, Differing Perceptions of New Venture Failure: A Matched Exploratory Study of Venture Capitalists and Entrepreneurs, *Journal of Small Business Management*, 37, 1-14.

Zuckerman, Miron, 1979, Attribution of Success and Failure Revisited, or: the Motivational Bias is Alive and Well in Attribution Theory, *Journal of Personality*, 47, 245-87.

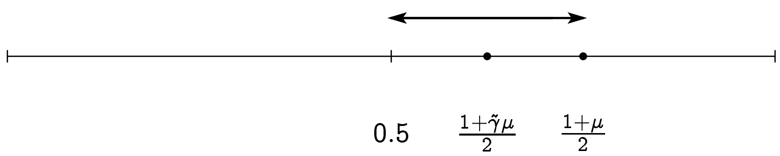

Figure 1: The manipulable range of the belief when $s_1 = 1$ ($\mu = 0.5$, $\tilde{\gamma} = 0.5$).

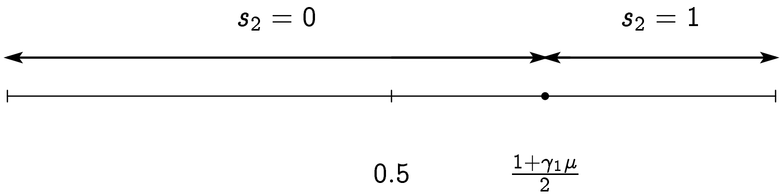

Figure 2: the manipulable range of the belief when $s_1 = 1$ ($\mu = 0.5$, $\tilde{\gamma} = 0.5$, $\gamma_1 = 0.8$).