

Motivated False Memory

Soo Hong Chew* Wei Huang[†] Xiaojian Zhao[‡]

This Version: September 2015

Abstract

People often forget and sometimes fantasizes. This paper reports an experiment on memory errors and their relation to preferential traits such as present bias and risk and ambiguity attitude. We find systematic incidence of memory biases in forgetting negative events and exhibiting false memory in favor of positive events. Intriguingly, positive delusion and positive confabulation are significantly related to the degree of present bias, but this is not the case for positive amnesia. In an intra-person, multiple-self model, false memory, rather than amnesia, serves to enhance confidence for one's future selves to account for our experimental findings.

Keywords: false memory, amnesia, delusion, confabulation, present bias.

JEL Classification: C91, D03, D83, Z13

*Department of Economics and Department of Finance, National University of Singapore, Singapore. Email: chew.soohong@gmail.com

[†]Economics and Management School, Wuhan University, China. E-mail: whuang.ust@gmail.com

[‡]Department of Economics, Hong Kong University of Science and Technology, Hong Kong. E-mail: xjzhao@ust.hk

1 Introduction

Implicit in much of economic analysis is the assumption of full consciousness – the decision maker has unlimited ability for attention towards stimuli, unbounded capacity for recording and storage of events, and perfect accuracy of recall. This runs counter to the fact that as biological beings, we have limited capacity for attention, recording, and recall. At any moment, we cannot be conscious of all stimuli or sensations registered at the biological level nor can we recall accurately all that are stored at various levels of memory. To varying degrees, people are susceptible to limited attention and imperfect recall including forgetfulness and false memory encompassing memory illusion and delusion (see, e.g., Pashler, 1998).

In a recent paper, Howe (2011) reviews the bright or positive side of illusion, including perceptual illusion, cognitive illusion and memory illusion. McKay and Dennett (2009) report many fitness relevant memory illusions, e.g., positive yet illusional self-appraisals, can lead to a sense of confidence which in turn leads to future success. Howe and Derbish (2010) and Howe et. al. (2011) argue that one adaptive value of false autobiographical recollection is the tendency toward positive biasing of one’s past self, which allows revision of the past that facilitates self-enhancement. When people exhibit delusion, they tend to fabricate non-existent evidence in the positive direction. In an extreme case, the neuroscientist Ramachandran (1996) documents how a patient who could not move her left arm claims that she could engage in activities that require the use of both hands, say clapping. Additional examples include delusion of persecution in attributing one’s failure to conspiracies (Bortolotti and Mameli 2012) and the “reverse Othello syndrome” when being delusory in the fidelity of one’s partner can work as a defense mechanism to maintain one’s self-esteem (see Burler (2000) and McKay, Langdon, and Coltheart (2005)).¹

It is important to investigate the instrumental value of false memory, encompassing delusion and illusion, from an economics perspective. The presence of false memory has a wide range of real-life relevance, such as enhancing one’s self-image to boost

¹Memory may also be influenced by information supplied after an event occurs (Loftus and Palmer, 1974) and false memory may be subject to manipulation by others (Loftus, 1993), including the possibility of suggestive questions distorting a witness’ memory (Frenda et. al. 2011).

labour market value, building the academic dream to motivate research graduate students and junior professors (Cross, 1977), and creating organizational culture to enhance corporate performance (e.g., Benabou, 2013). Limitations in consciousness, including attention and memory, have appeared in the economics literature. Dow (1991) studies optimal search under limited memory and shows how a decision maker may focus scarce cognitive resources on part of the problem. Sims (2003) links the idea of limited inattention to sources of inertia in prices and wages. Carrillo and Mariotti (2000) offer a link between strategic ignorance and present bias inducing a need to sustain personal motivation by ignoring information that may weaken the individual's self confidence.² In Benabou and Tirole (2002), a functional role of amnesia has emerged in terms of one's decision to suppress recall of negative events. In their model of selective memory involving either a bad signal or no signal at an earlier stage, the incidence of present bias induces a demand for the individual to form motivated over confidence³, giving rise to the tendency to forget a negative signal in favor of no signal in equilibrium in an intra-person, multi-self setting. Still, the study of the cause and effect of false memory including delusion and confabulation has been largely missing in the economics literature.

The present paper studies both experimentally and theoretically amnesia and false memory. In the experiment, we study three types of memory errors: amnesia (forgetting a past event), delusion (fabricate an event that did not actually happen); and confabulation (distorting the memory of a past event to another distinct event). In the initial stage of the experiment, subjects take an incentivized Ravens IQ test after completing a number of decision making tasks including one on temporal discounting. In a subsequent stage several months later, subjects are shown 6 questions, each accompanied by the correct answer, comprising four from the original test and two which are new but similar. For each question, subjects are asked to recall whether

²Recently, Brown, Croson and Eckel (2011) find support of Carrillo-Mariotti's model in an experimental test.

³Besides personal over confidence, Svenson (1981) studies over confidence in terms of social rankings. In the test by Burks et al (2013) of three mechanisms that can deliver over confidence, social signaling is supported while Bayesian updating (Benoit and Dubra, 2011) and concern for self image (Koszegi, 2006) are rejected.

they did it correctly, did it incorrectly, did not see it, or do not remember. Subjects receive S\$1 for each correct recall, lose S\$1 for each incorrect recall, and receive nothing for not remembering. We find systematic incidence of memory biases in forgetting negative events and exhibiting false memory encompassing delusion and confabulation in favor of positive events. Intriguingly, positive delusion and positive confabulation are significantly related to the degree of present bias, but this is not the case for positive amnesia.

To account for the key findings of our experiment, we apply the Benabou-Tirole (B-T) (2002) model with positive amnesia countering the effect of present bias leading the individual to undertake less investment activity than otherwise. In principle, we may re-interpret the B-T model as applying to delusion in terms of no-signal versus good signal, or confabulation in terms of bad signal versus good signal, giving rise to a positive relation between present bias and positive delusion and positive confabulation respectively. Thus, the delusion and confabulation interpretations of the B-T model but not the amnesia interpretation are compatible with our experimental results. While the stylized B-T model allows for reinterpretation in terms of delusion as well as in terms of confabulation, it cannot distinguish between the roles of amnesia and false memory. Extending B-T to a model involving a bad signal, no-signal, and a good signal, we can endogenize amnesia and false memory and relate them to the individual's attitude towards temporal discounting and investigate the interaction between these two memory management instruments. This extended model enables the study of the instrumental values of memory biases on the supply side of over confidence. Following the B-T model, the individual arrives at his belief inducing a demand for over confidence to motivate himself to resolve this under-investment problem in terms of ex ante evaluation. Otherwise, having a correct belief about oneself can lead to under-investment due to present bias. On the supply side, besides positive amnesia, positive false memory may offer additional channels to deliver over confidence. In our model, false memory and amnesia can act as potential substitutes as motivational mechanisms for the individual. For sufficiently severe present bias, the possibility of false memory precludes amnesia from delivering over confidence. This enables us to identify the potential role of positive false memory as a key mechanism to

supply one’s motivation and demonstrate how one instrument of memory management can invalidate another.

The paper is organized as follows. Section 2 introduces our experimental design and discusses our experimental findings. Section 3 presents our theory to account for the experimental results. Section 4 offers concluding remarks.

2 Experiment

We conduct an experiment to study three kinds of memory errors – amnesia, delusion and confabulation – in IQ performance and relate them to a range of preferential attitudes including time, risk and ambiguity.

2.1 Design

In the first stage of our experiment, subjects’ discount rates are elicited from comparisons between their tradeoffs in a proximal task (next day versus 30 days later) and a distal task (351 days versus 381 days later). In addition, we elicit subjects’ risk and ambiguity attitudes and have them complete an incentivized Ravens IQ test (see Appendix B). In the second stage after months, we elicit different types of memory errors based on their recall of performance in the IQ test in an incentivized setting. Of 896 subjects from the National University of Singapore to participate the experiment, 701 subjects completed both stages of our study.

In the subsequent stage, subjects are shown 6 test questions one at a time together with the correct answers. Of the 6 questions, 4 appeared in the initial stage and 2 are new. For all subjects, all 6 questions are identical but the order of appearance is randomized. For each of these 6 questions, subjects can choose one of 4 responses:

- a : My response was correct.
- b : My response was incorrect.
- c : I didn’t see this question.

- d : I don't remember.

For the 4 questions which had appeared previously, subjects each receives S\$1 (about US\$0.80) if their choice reflects correctly their performance or if they choose (c) for the 2 questions which had not appeared previously. The subject loses S\$1 when his/her choices reveals a memory error – exhibiting false memory about having seen a question which had not appeared previously or remembering incorrectly. Subjects always receive nothing if they choose (d) – “I don't remember”.

For each of the 6 questions presented in the second stage, the subject either did it right ($s = G$), did it wrong ($s = B$), or did not see it ($s = \emptyset$) at the initial stage. The table below displays subjects' possible responses in relation to s .

	a	b	c	d
$s = G$	a_G : CR	b_G : Negative C	c_G : Negative A	d_G : Weak Negative A
$s = B$	a_B : Positive C	b_B : CR	c_B : Positive A	d_B : Weak Positive A
$s = \emptyset$	a_\emptyset : Positive D	b_\emptyset : Negative D	c_\emptyset : CR	d_\emptyset : Weak CR

There are three types of correct recall (CR), a_G , b_B , and c_\emptyset . Moreover, compared to c_\emptyset (recalling correctly that one has not seen the question previously), d_\emptyset (choosing “I don't remember” when one has not seen the question before) reveals a weak sense of correct recall. There remains 8 types of memory errors: two linked to delusion (D), a_\emptyset and b_\emptyset , which denote the case where the subject remembered a new question incorrectly; two linked to confabulation (C), a_B and b_G , which means that the subject misremembered his/her performance with respect to a question which had appeared previously. In terms of amnesia (A), when the question had appeared before, stating “I don't remember” (option (d)) is weaker than claiming “I didn't see this question” (option (c)). Thus, we denote d_G and d_B as weak amnesia compared with c_G and c_B as amnesia. Given the parallel between responses in (c) and the responses in (d), we have grouped them together in our data analysis.

2.2 Results

We report the findings of an incentivized experiment on different types of memory bias and their relation to temporal discounting and some other personal traits such as IQ, risk preference, gender, etc.

2.2.1 Memory Bias

The overall memory patterns (see Table C1 in Appendix C) reveal considerable incidence of amnesia (33.95%), delusion (64.12%), and confabulation (10.63%) relative to all the possibilities. In each case, we find a consistent tendency for a positive bias which we shall detail below.

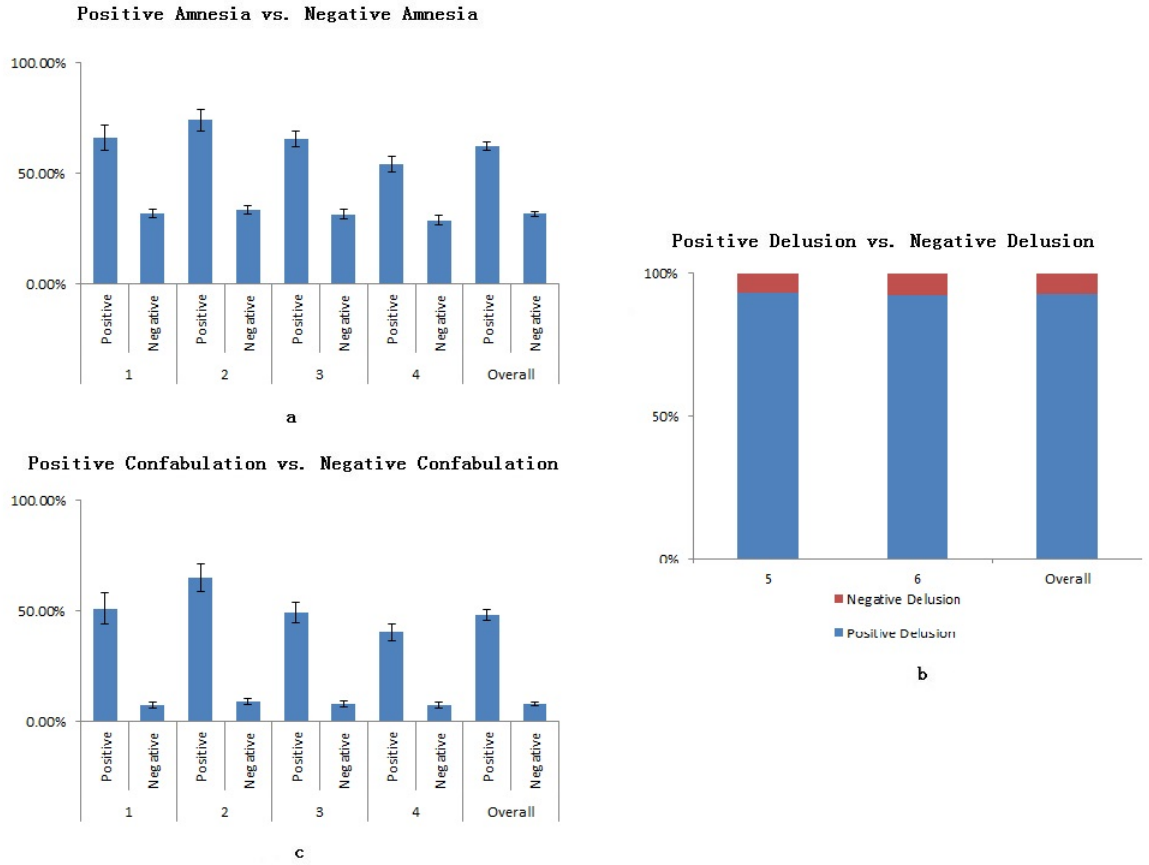


Figure 1: Memory Bias Conditional on Performance

Figure 1a displays the respective rates of positive amnesia $(c_B + d_B)/(b_B + c_B + d_B)$

for Question 1 to Question 4 at 65.22%, 74.07%, 65.52% and 53.67%, which are significantly higher than the corresponding rates of negative amnesia $(c_G + d_G)/(a_G + c_G + d_G)$ at 31.49%, 32.97%, 31.46% and 28.86% (see Table C2). In other words, individuals who did a question incorrectly are significantly more likely to forget than those who did the question correctly.

In summary, we have the following result of positive amnesia.

Result 1 *Individuals tend to exhibit positive amnesia rather than negative amnesia.*

Figure 1b displays the rate of positive delusion $a_{\varnothing}/(a_{\varnothing} + b_{\varnothing})$ and negative delusion $b_{\varnothing}/(a_{\varnothing} + b_{\varnothing})$. The rates of positive delusion for Question 5 and Question 6 are respectively 92.92% and 92.42% in comparison with the base rates of correct response for questions 1 to 4: 86.73%, 82.88%, 67.05% and 59.49% (as shown in Table C3). Compared with Question 1, the rates of positive delusion for questions 5 and 6 are significantly higher than the base rate (respectively $p = 0.0006$ and $p = 0.0011$). All the other p values are at the 0.0000 level. Taken together, the pattern of delusion exhibits a significant positive tendency relative to the actual base rates of correct response.

According to Table C1 in Appendix C, 25.55% of those who did it incorrectly in the first stage exhibit positive confabulation. As observed earlier, this is not compatible with our basic model. Figure 1c displays the respective rates of positive confabulation $a_B/(a_B + b_B)$ for Question 1 to Question 4 at 50.00%, 65.00%, 48.72% and 39.52%, which are significantly greater than the corresponding rates of negative confabulation $b_G/(a_G + b_G)$ at 7.04%, 8.73%, 7.58% and 7.26% (see Table C2).

Summarizing, we have the following result where false memory refers to both delusion and confabulation.

Result 2 *Individuals tend to exhibit positive false memory including positive delusion and positive confabulation rather than negative false memory.*

To check the pattern of memory biases such as the relation between positive delusion and positive confabulation, we next observe:

Result 3 *Individuals who do not exhibit positive delusion are less likely to have positive confabulation; individuals with positive confabulation are more likely to have positive delusion.*

This result (see Table C4) describes the relationship between positive delusion and positive confabulation. Among 210 subjects out of 701 who do not exhibit positive delusion, the rates of positive confabulation of 26.67%, 42.86%, 36.67% and 25.71% for these four questions are significantly lower than the corresponding rates of unconditional positive confabulation of 50.00%, 65.00%, 48.72% and 39.52% at p -values of 0.058, 0.065, 0.1202 and 0.0627 respectively. Correspondingly, among 145 subjects who each exhibits some positive confabulation, their rates of positive delusion of 73.19% and 78.36% are significantly higher than the corresponding unconditional rates of positive delusion of 58.72% and 66.02% at p -values of 0.0007 and 0.0026 respectively. In sum, individuals who do not exhibit delusion are less likely to have positive confabulation while individuals with positive confabulation are more likely to have positive delusion.

Together, these findings suggest that positive delusion may be an intermediate step in a process leading to positive confabulation. While Proposition 3 in the two-step model in our general theory below fits naturally into this result, we are not aware of any paper in the existing literature on memory management providing the explanation. This two-step confabulation process appears to be related to the idea of Korsakoff Syndrome (Whitty and Lewin, 1960) in which confabulation can be considered to serve as compensatory pseudo-reminiscence to fill the memory gap. In other words, the brain can produce false memory to make up for memory loss.

2.2.2 Present Bias and Memory Bias

When a subject does not recall whether he has seen a specific question previously, choosing (a), (b), or (c) entails some degree of downside risk or ambiguity. From this perspective, option (d) is free of risk or ambiguity. Before studying the implications of our model relating positive memory biases to the degree of present bias, we first examine whether the frequency of choosing (d) is related to risk attitude and ambiguity attitude measured using two incentivized tasks (see Appendix B). We run an ordered

probit regression on the number of (d) choice, $\#d$ (from 0 to 6), with regressors β , δ , IQ , RA , AA , $Gender$ and $Duration$ where IQ (from 0 to 60) is the score on the Ravens test in the first stage, RA (from 0 to 10) refers to the observed degree of risk aversion in a portfolio choice task, AA (from -10 to 10) refers to the degree of ambiguity aversion, $Gender$ equals 1 if the subject is female and 0 otherwise, and $Duration$ is the number of days between two stages. We do find a significant negative relation between $\#d$ with IQ which corroborates the general finding of a positive relation between IQ and accuracy of recall ($p = 0.000$). The absence of a significant relation between $\#d$ and $Duration$ reveals that time difference does not play an important role in our memory test ($p = 0.919$). There is no significant relation between $\#d$ and subjects' risk attitude, ambiguity attitude, or gender. While the estimated coefficients for RA is negative, it is not significantly different from zero ($p = 0.219$).

In studying the influence of present bias, we first focus on positive memory biases based on subjects' recall of their performance on Question 1 to Question 4. Consider the subject's strategy for positive amnesia which applies when he did a question incorrectly. We define a binary variables h_B to indicate subjects' recall strategy when his/her performance was wrong. If he correctly recalls his performance in the second stage, i.e., choosing (b), the recall strategy is $h_B = 1$. If he cannot recall his initial performance or he does not remember having seen this question, i.e., choosing (c) or (d), we interpret this as positive amnesia with recall strategy $h_B = 0$. We run a probit regression on h_B with regressors β , δ , IQ , RA , AA , $Gender$ and $Duration$.

Result 4 *Positive amnesia is not related to the degree of present bias.*

The estimated coefficients for β in questions 1 to 4 are of different signs and are individually not significant. Pooling them together, the result is still not significant. Thus amnesia does not have a significant probit relation with present bias (see Appendix C.2 on amnesia).

We first examine the subject's strategy for positive delusion with the two new questions – Question 5 and Question 6. If he indicates that he did it correctly, the delusion strategy is $h_\emptyset = 0$. Otherwise, if he indicates that this question may be new, i.e., answer (c) or (d), his delusion strategy is $h_\emptyset = 1$. We run a probit regression on

h_\emptyset using the regressors β , δ , IQ , RA , AA , $Gender$ and $Duration$. We next consider the subject’s strategy for positive confabulation when he did a question incorrectly. If he indicates that he did it correctly, i.e., choosing (a), the confabulation strategy is $h'_B = 0$ which we interpret as positive confabulation. If he can recall his performance correctly, i.e., choosing (b), the confabulation strategy is $h'_B = 1$. Summarizing:

Result 5 *The likelihood of individual having positive false memory including positive delusion and positive confabulation increases in the degree of present bias (β lower).*

For both Question 5 and 6, the signs of the estimated coefficients for β are consistent with Corollary 1 and are significant ($p = 0.018$ and $p = 0.032$, respectively). Combining Question 5 and Question 6, the result of ordered probit regression is significant at $p = 0.010$ (see Appendix C.2 on delusion). For Question 1 and 2, the coefficients for β are significant with positive signs ($p = 0.028$ and $p = 0.000$). For Question 4, the coefficient is positive, and has a marginal significance ($p = 0.070$). The result for Question 3 is not significant ($p = 0.980$). Pooling four questions together, the result remains significant ($p = 0.015$).

The overall findings in Result 4 and Result 5 support the implication of Corollary 1 of our model without confabulation and Proposition 3 of the model with two-step confabulation, as one can see in the next section. Note that positive amnesia, positive delusion and positive confabulation are not related to the distal discount rate δ . While we have found positive amnesia in subjects’ responses to Question 1 to Question 4, positive delusion in their responses to Question 5 and Question 6, and positive confabulation in their responses to Question 1 to Question 4, we do not find significant relation between 9 of these 10 memory biases and the distal discount rate δ . Only the coefficient of δ in Question 1 for confabulation is significant ($p = 0.041$). This absence of significant relation is compatible with the implications of the B-T model concerning positive amnesia as well as in our model (see below). To further explore the robustness of our findings, we do not find significant relation between any of these three positive memory biases and subjects’ duration between two stages.

For completeness, we next examine the possible influence of IQ on memory patterns along with risk aversion, ambiguity aversion, and gender (see Appendix C.2).

We first focus on unconditional memory patterns. Consistent with what has been reported in the literature, we find IQ to be positively related to accuracy in unconditional recall and actual performance for Question 1 to Question 4. Interestingly, in examining Question 5 and Question 6, besides being positively related to positive delusion, we find that IQ is positively related to unconditional delusion. Then, we examine the relation between IQ memory bias in terms of positive amnesia, positive delusion and positive confabulation. We find that higher IQ is positively related to a higher likelihood of positive false memory but not related to positive amnesia.⁴ The signs of the estimated coefficients of positive delusion and positive confabulation for IQ are significantly negative for Question 1 to Question 6 (all p -values are lower than 0.05).

3 Modeling Motivated False Memory

Our findings of three kinds of positive memory biases under controlled laboratory conditions and their relation to present bias appear to be novel. In this section, we first apply the B-T approach involving two signals to account for our specific findings. In this approach, individuals tend to have the under-investment problem due to their present biased preference. This leads to a demand of over confidence to resolve the insufficient motivation for investments. To explain our findings of motivated false memory more fully, we extend the B-T to a three signal framework both with and without confabulation in order. Our theoretical development may also be recast in an inter-generational setting at the societal level to study collective amnesia and false as will be discuss in Section 4.

⁴One may question if there may be an endogeneity problem since memory and IQ are known to be related. Here, our dependent variables of positive amnesia, positive delusion, and positive confabulation are distinct from unconditional recall for which our subjects' behavior accord with known findings in the literature. It seems implausible that better memory in terms of having a greater capacity for correct unconditional recall would by itself induce the kinds of positive memory biases observed in our study.

3.1 Benabou-Tirole Approach

The original B-T model involves three dates, $t = 0, 1$, and 2 , which may be interpreted as temporary selves in an intra-person, multi-self game. Memory is imperfectly formed during $t = 0$, corresponding to Stage 1 of our experiment, and contributes to the individual's belief at $t = 1$ which corresponds to Stage 2 when memory-based decisions are made. Following the (quasi-)hyperbolic discounting literature (see, e.g., Strotz, 1955; Laibson, 1997), self- t discounts the expected utility of self- $(t + n)$ at $\beta\delta^n$ for $n = 1, 2, 3$, where $\delta = 1$ is the normal discount rate and $\beta < 1$ corresponds to the incidence of present bias.

At the outset, self-0 receives one of two possible private signal concerning his ability θ . Subsequently, self-1 engages in an activity at cost c which if successful will yield benefit V at $t = 2$ and zero otherwise. Self-0 knows the distribution of c which is assumed for simplicity to be uniformly distributed over $[0, \bar{c}]$. The individual's ability is captured by the probability of success θ referred to as personal ability in the costly investment activity, e.g., studying in school, making effort to lose weight, and working in a company, which are associated with motivation in the real world outside the laboratory. Thus his expected benefit at $t = 2$ is θV .

There is a problem of under investments due to individuals' tendency to exhibit present biased preference leading to a demand of over confidence. The B-T model offers a memory management mechanism in terms of positive amnesia – replacing a bad signal by no signal – to supply over confidence and counter this under investment problem. Their main result states that the likelihood of bad signal suppression is positively related to degree of present bias.

More generally, beyond positive amnesia, the individual in the B-T model may mis-remember a lower ability signal as a higher ability one in two additional ways: positive delusion (from no-signal to good signal) and positive confabulation (from bad signal to good signal). Their main proposition then implies that positive false memory – delusion and confabulation – relates positively to the degree of present bias. While these implications of the B-T approach can account for the observed relation between present bias and positive false memory, it cannot concurrently explain the absence of a positive relation between present bias and positive amnesia. Moreover, the B-T

approach is not compatible with Result 3 which suggests that positive delusion is necessary for positive confabulation.

3.2 An Extended Model without Confabulation

To account for the experimental observation differentiating the roles of amnesia and false memory, we extend the B-T model to a 3-signal model. At the outset, self-0 receives a private signal s concerning his ability θ . This may be a bad signal ($s = B$), no signal ($s = \emptyset$), or a good signal ($s = G$). At $t = 1$, he may mis-remember the nature of the signal received at $t = 0$.

We first develop our extended model to account for the case of no confabulation and focus on amnesia (forgetting the receipt of a signal) and delusion (remembering a signal when there was none). Here, $s = \emptyset$ occurs with probability $1 - q$, $s = B$ occurs with probability qp , and $s = G$ occurs with probability $q(1 - p)$. We may interpret q as a measure of the individual's degree of social exposure. Thus, conditional on receiving a signal, the probability of it being bad is given by p and being good is given by $1 - p$. Given feedback from society, p represents the chance that the individual will face a negative event largely due to low ability.

For $s = B$, \emptyset , or G , we refer to the individual as type- B , type- \emptyset , or type- G , respectively. Let θ_s refer to the expected value of θ conditional on each possible realization of the true signal s , i.e., $\theta_B = E[\theta|s = B]$, $\theta_\emptyset = E[\theta|s = \emptyset]$ and $\theta_G = E[\theta|s = G]$.⁵ Let

$$\theta_\emptyset = p\theta_B + (1 - p)\theta_G.$$

where $\theta_B < \theta_G$. That is, receiving no signal implies that his ability is given by the expected ability in the presence of a signal. Thus, θ_\emptyset lies in between θ_B and θ_G . In order to incorporate self-0's role in memory formation including the possibility of delusion for self-1, we let \hat{s} denote the subjective signal transmitted from self-0 to self-1. Specifically:

- (i) $s = G$: In this case $\hat{s} = G$ in the absence of opportunity for signal manipulation.

⁵Alternatively, we can simply assume that θ_s is the value of θ given the true signal s . Since the value of θ may be stochastic for a given s as discussed in Benabou and Tirole (2002), we take a more general interpretation of θ_s which represents the expected value of θ conditional on s .

(ii) $s = B$: In this case $\hat{s} = B$ or $\hat{s} = \emptyset$. Self-0 may communicate the signal truthfully to self-1 or suppress the bad signal (*amnesia*).

(iii) $s = \emptyset$: In this case $\hat{s} = \emptyset$ or $\hat{s} = G$. Self-0 may leave it as it is or create a fake good signal (*delusion*).

Our modeling setup reflects Result 1 and Result 2 where subjects tend to have positive amnesia and positive delusion.⁶

Let $h_s = \Pr[\hat{s} = s|s]$ denotes the probability that he chooses to transmits the signal s truthfully to self-1 by self-0 of type s . We denote by h_B^* and h_\emptyset^* the respective beliefs held by self-1 concerning self-0 being truthful in the chosen recall and delusion strategies.

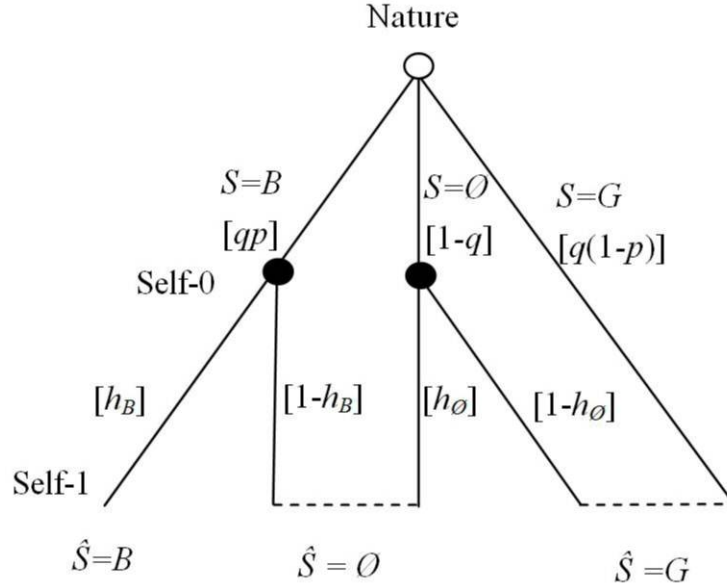


Figure 2: Memory Management without Confabulation

The intra-personal game involving the individual's memory management strategy is depicted in Figure 2. At epoch 1, self-1 forms expectations over his ability θ in light of the recalled \hat{s} , taking into account the possibility that self-0 may have suppressed the true signal or created a fake signal. Let $\theta^*(\hat{s})$ denote self-1's assessment of his ability given \hat{s} and $r^*(\hat{s})$ denote the reliability of the signal \hat{s} , i.e., the probability that the signal \hat{s} is accurate. When $\theta^*(\hat{s}) > \theta_s$, we say the individual exhibits *over*

⁶In terms of our experiments, we classify the recalled signal \hat{s} as follows: $\hat{s} = G$ if we observe a_G or a_\emptyset ; $\hat{s} = B$ if we observe b_B ; and $\hat{s} = \emptyset$ if we observe c_\emptyset , c_B , d_\emptyset , or d_B .

confidence. Clearly, $\theta^*(B) = \theta_B$. For the case of $\hat{s} = \emptyset$, applying Bayes' rule, we have:

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

It follows that

$$\theta^*(\emptyset) = r^*(\emptyset)\theta_\emptyset + (1-r^*(\emptyset))\theta_B \geq \theta_B.$$

Similarly, for the case of $\hat{s} = G$, we have:

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

and

$$\theta^*(G) = r^*(G)\theta_G + (1-r^*(G))\theta_\emptyset > \theta_\emptyset.$$

The last inequality reflects strict over confidence of self-1 when $s = \emptyset$, because his updated belief $\theta^*(G)$ of his ability is always higher than the true ability θ_\emptyset of self-0 of type- \emptyset when self-1 receives a good signal. However, his updated ability given no signal $\theta^*(\emptyset)$ equals θ_B when h_\emptyset^* equals 0 so that $r^*(\emptyset) = 0$. In other words, the incidence of delusion precludes amnesia from delivering over confidence.

Self-1 will incur the cost of investment if and only if

$$\beta\theta^*V - c \geq 0.$$

The divergence in interest between self-0 and self-1 is captured by having $\beta < 1$. Notice that the qualitative features of the analysis above is robust with respect to a form of non-Bayesianism often called partial naivete in which the first term of the denominator in the expression for reliability is modified by a factor λ which may be less than 1.

When $s = B$, self-0 chooses recall strategy h_B . The net gain from suppressing the bad signal ($\hat{s} = \emptyset$) is thus equal to:

$$\int_{\beta\theta_B V}^{\beta\theta^*(\emptyset)V} \{\theta_B V - c\} dF(c) \tag{1}$$

where F refers to the distribution function of c .⁷ When $\beta\theta^*(\emptyset)V$ exceeds $\beta\theta_B V$, amnesia delivers over confidence which in turn gives rise to more investment activities.

Similarly, when $s = \emptyset$, self-0 chooses delusion strategy h_\emptyset . The net gain from creating a fake signal ($\hat{s} = G$) is then given by:

$$\int_{\beta\theta^*(\emptyset)V}^{\beta\theta^*(G)V} \{\theta_\emptyset V - c\} dF(c). \quad (2)$$

Notice that $\beta\theta^*(G)V$ always exceeds $\beta\theta^*(\emptyset)V$, so that delusion delivers over confidence leading further to more investment activities.

We explore the functional role of positive amnesia and positive delusion and study how they depend on present bias β , how motivated delusion interacts with amnesia, and how these motivational mechanisms contribute to the supply of over confidence. A central question of our work is on the relation between the individual's magnitude of present bias and his possible states of motivated memory error in equilibrium.

Here, we apply the solution concept of perfect Bayesian equilibrium in conjunction with the intuitive criterion refinement (Cho and Kreps, 1987) to shed light on pure-strategy equilibrium outcomes of our memory management game.⁸

We have the following existence result:

Proposition 1 *There are three possible perfect Bayesian equilibria in pure strategies.*

- (i) (PBE1: Correct Recall) *There exists perfect Bayesian equilibrium, $h_B^* = 1$, $h_\emptyset^* = 1$ if $\beta > \beta_1$ for some $\beta_1 \in (0, 1)$.*
- (ii) (PBE2: Positive Amnesia) *For p close to 1, there exists perfect Bayesian equilibrium, $h_B^* = 0$, $h_\emptyset^* = 1$ if $\beta \in (\underline{\beta}_2, \bar{\beta}_2)$ for some $\underline{\beta}_2, \bar{\beta}_2 \in (0, 1)$.*
- (iii) (PBE3: Positive Delusion) *There exists perfect Bayesian equilibrium, $h_\emptyset^* = 0$ and $h_B^* = 1$ or 0, if $\beta < \beta_3$ for some $\beta_3 \in (0, 1]$.*

In case (i) with β large enough, i.e., the problem of present bias is not too severe, there is an equilibrium with perfect recall and no delusion. In this case, information is always valuable because of an alignment in interests between self-0 and self-1.

⁷For simplicity, we ignore the present bias of self-0 when calculating his expected utility.

⁸Under a uniform distribution, we can solve the mixed strategy equilibria explicitly. For simplicity, we do not consider mixed strategy equilibria in the balance of the paper.

Here, truthful reporting is an equilibrium regardless of the signal received by self-0. Otherwise, amnesia or delusion would create a bias towards over-investment of self-1 from the perspective of self-0 because of his (almost) perfect dynamic consistency. Since PBE1 does not involve amnesia or delusion, we call this equilibrium “correct recall”.

In case (ii) where p is close to 1, there is an equilibrium (PBE2) with amnesia and no delusion for intermediate values of β . The amnesia condition requires β to be bounded from above while no-delusion requires β to be bounded from below. Here, self-1 of type- B and type- \emptyset are pooled. If the individual receives a bad signal, he would choose to ignore it leading him to become over confident. On the other hand, if self-1 does not receive any signal, he becomes under confident since he does not know whether it is indeed the case. With p close to 1, θ_\emptyset is closer to θ_B than θ_G . In this case, a small degree of over confidence will make amnesia beneficial. At the same time, with the larger gap between θ_\emptyset and θ_G , any incidence of delusion would induce excessive over investment. Thus, PBE2 with amnesia without delusion emerges.

The equilibrium outcome in case (iii) involves two possibilities.

(a) *Delusion without amnesia*: Here, self-0 of type- B transmits the bad signal to self-1 while self-0 of type- \emptyset creates a fake good signal. In this case, self-1 of type- \emptyset and type- G are pooled. Thus self-1 of type- G will have self-doubt because he is uncertain whether what he receives is a true signal or a fake signal created by self-0, i.e., a delusion strategy delivers under confidence for self-1 of type- G . Here, delusion requires a low β while no amnesia requires a high β . It follows that an intermediate level of β is required to maintain delusion without amnesia. Note that this equilibrium is determined by the off-equilibrium belief $r^*(\emptyset)$. When $r^*(\emptyset) = 0$, self-0 of type- B is indifferent between correct recall and amnesia. When $r^*(\emptyset) > 0$, he will choose to recall the bad signal for β sufficiently large. Under the intuitive criterion, $r^*(\emptyset)$ can only be zero (see Appendix A.1 for details). In this case, the decision maker is indifferent between correct recall and amnesia and would not need a large β to remain self truthful. Thus, when β is small enough, this equilibrium prevails.

(b) *Delusion with amnesia*: When present bias is sufficiently severe, i.e., β small enough, self-0 of both type- B and type- \emptyset will cheat. Nonetheless, this equilibrium

outcome is similar to the case above. There is no actual amnesia in equilibrium because when self-1 receives no signal, he knows that he is of type- B . While delusion and amnesia can act as substitutes as motivational mechanisms to supply over confidence, delusion precludes amnesia from delivering over confidence in this equilibrium. Compared to case (a) above, we call the amnesia in case (b) fake amnesia. Thus, the equilibria in case (a) and case (b) are essentially identical.

Corollary 1 *The likelihood of positive delusion increases in the degree of present bias, but this is not the case for positive amnesia.*

Corollary 1 partly accounts for Result 4 and Result 5 in our experimental finding, by providing a simple mechanism on how the functional role of amnesia can be invalidated. A higher chance of delusion results from a greater present bias in time preference. In contrast with the B-T model, the existence of the equilibrium with amnesia does not depend monotonically on the magnitude of present bias due to the possibility of fake amnesia in case (iiib). Because amnesia in PBE3 is fake, the individual will recall the bad signal when β is either large enough or small enough and suppress the bad signal for intermediate values of β . In other words, given the existence of delusion, amnesia no longer has a role in delivering over confidence when present bias is severe.

3.3 An Extended Model with Confabulation

We extend the previous model to account for false memory more fully and include positive confabulation, observed in our experiment. Compared to amnesia or delusion, confabulation seems potentially more involved. While it may result from a direct transformation of one signal into another (*one step*), it may also arise from forgetting a signal and then exhibiting delusion in seeing a different signal (*two steps*). We first elaborate a one-step confabulation extension followed a two-step confabulation extension.

3.3.1 One-Step Confabulation

Incorporating the possibility of mis-recalling the bad signal as a good one, i.e., self-0 of type- B may recall the bad signal correctly, forget the bad signal, or transform the bad signal in to a good one. All the other parts of the setup remain as before. Specifically, when $s = B$, \hat{s} can take three values: $\hat{s} = B$, $\hat{s} = \emptyset$ or $\hat{s} = G$. That is to say, self-0 may communicate the signal truthfully with probability h_{BB} , suppress the bad signal with probability $h_{B\emptyset}$ (amnesia) or recall the signal wrongly with probability $1 - h_{BB} - h_{B\emptyset}$ (confabulation). Figure 3 displays the one-step extended model with confabulation.

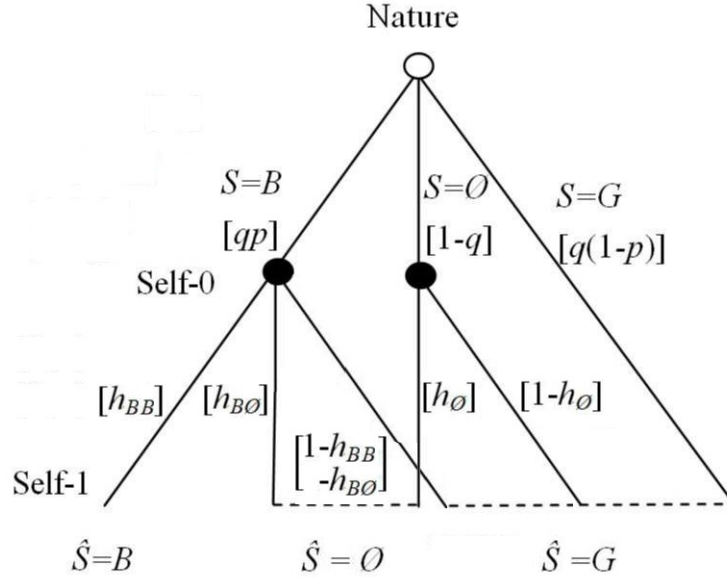


Figure 3: Memory Management with One-Step Confabulation

Following the exposition of our no-confabulation model earlier, it is straightforward to identify 6 possible PBE's as shown in Table 1 four of which are inherited from that model. There are two additional confabulation equilibria consisting of PBE5 and PBE6 where present bias is sufficiently severe. In these cases, an individual of type- B exhibits confabulation to deliver over confidence. Here, the expected assessment of self-1 upon receiving signal G is always $\theta^*(G) = p\theta_B + (1 - p)\theta_G$, no matter what memory strategy self-0 of type- \emptyset chooses, i.e., self-0 of type- \emptyset is always indifferent between creating a fake good signal and not. In other words, the existence of con-

fabulation invalidates the instrumental value of delusion, and only confabulation can deliver over confidence when present bias is severe enough. This key implication is summarized in the following proposition.

Proposition 2 *In the one-step confabulation model only the likelihood of positive confabulation increases in the degree of present bias, but it is not the case when it comes to positive amnesia or positive delusion.*

Equilibria	h_{BB}	$h_{B\emptyset}$	h_{\emptyset}	Present Bias β	Positive Amnesia	Positive Delusion	Positive Confabulation
PBE1	1	0	1	Large enough	N	N	N
PBE2	0	1	1	Intermediate	Y	N	N
PBE3	1	0	0	Intermediate	N	Y	N
PBE4	0	1	0	Intermediate	Y	Y	N
PBE5	0	0	1	Small enough	N	N	Y
PBE6	0	0	0	Small enough	N	Y	Y

Table 1. Equilibrium Memory Bias for One-Step Extended Model

We note that the implication of this proposition is not consistent with two key results of our experiment: positive delusion and positive confabulation are each related to positively present bias (Result 5) and confabulation involving two steps – positive amnesia followed by positive delusion (Result 3). This motivates our two-step confabulation extension of the B-T model to be explicated in the next subsection.

3.3.2 Two-Step Confabulations

We posit a memory strategy involving four epochs: $t = 0, t = 0^+, t = 1$, and $t = 2$. At $t = 0$, the individual chooses his memory strategy after receiving a signal about his ability. At $t = 0^+$, the individual applies his memory strategy again after receiving the reported signal and further chooses how to transmit the reported signal – truthfully or deceptively. At $t = 1$, he decides whether to engage in the activity which delivers a payoff at $t = 2$. In this setting there are multiple ways to exhibit amnesia and delusion, given that motivated memory can happen at both $t = 0$ and $t = 0^+$.

Figure 4 displays the two-step extended model with confabulation. Based on self- 0^+ 's memory, when $\hat{s} = B$, he can choose to recall it with probability \tilde{h}_B or forget it with probability $1 - \tilde{h}_B$. When $\hat{s} = \emptyset$, he can choose to transmit it truthfully to self-2 with probability \tilde{h}_{\emptyset} or exhibit delusion with probability $1 - \tilde{h}_{\emptyset}$. Accordingly,

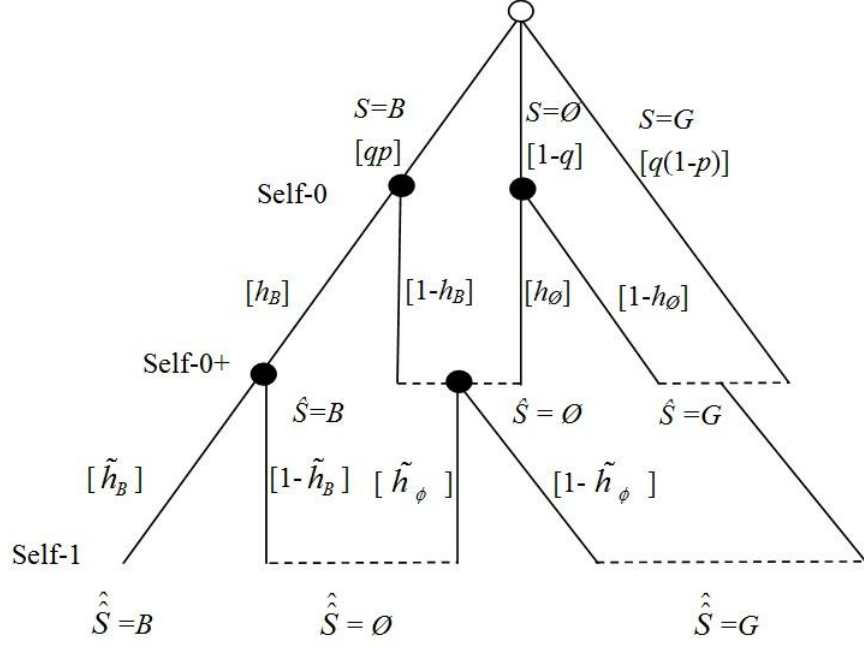


Figure 4: Memory Managment with Two-Step Confabulation

we define the reported signal transmitted from self-0⁺ to self-1 as $\hat{\hat{s}}$. As before, self-1 will decide whether to undertake the task according to his belief about his ability.

Equilibria	h_B	h_ϕ	\tilde{h}_B	\tilde{h}_ϕ	Present Bias β	Positive Amnesia	Positive Delusion	Positive Confabulation
PBE1	1	1	1	1	Large enough	N	N	N
PBE2	0	1	1	1	Intermediate	Y	N	N
PBE3a	1	0	1	1	Intermediate	N	Y	N
PBE3b	0	0	1	1	Intermediate	N(fake)	Y	N
PBE2a	1	1	0	1	Intermediate	Y	N	N
PBE2b	0	1	0	1	Intermediate	Y	N	N
PBE3c	1	0	0	1	Intermediate	N(fake)	Y	N
PBE3d	0	0	0	1	Intermediate	N(fake)	Y	N
PBE3e	1	1	1	0	Intermediate	N	Y	N
PBE3f	1	0	1	0	Intermediate	N	Y	N
PBE3g	1	1	0	0	Intermediate	N(fake)	Y	N
PBE3h	1	0	0	0	Intermediate	N(fake)	Y	N
PBE4a	0	1	1	0	Small enough	N	Y	Y
PBE4b	0	0	1	0	Small enough	N	Y	Y
PBE4c	0	1	0	0	Small enough	N	Y	Y
PBE4d	0	0	0	0	Small enough	N	Y	Y

Table 2. Equilibrium Memory Bias for Two-Step Extended Model

Similarly, we can identify 16 possible PBE's shown in Table 2 after applying the intuitive criterion (see Appendix A.2 for details). We classify the equilibria according to amnesia, delusion and confabulation by comparing the final signal $\hat{\hat{s}}$ with the initial signal s . The first four equilibria are inherited from the previous model without

confabulation. In addition to PBE2, there are two additional amnesia equilibria consisting of PBE2a and PBE2b in the extended model. Similarly, besides PBE3a and PBE3b, there are six additional pure delusion equilibria: PBE3c to PBE3h. The extended model with 2-step confabulation further delivers the desired confabulation equilibria from PBE4a to PBE4d.

The proposition below summarizes how positive delusion and positive confabulation relate to the degree of present bias in this model.

Proposition 3 *In the two-step confabulation model the likelihood of both positive delusion and positive confabulation increase in the degree of present bias, but it is not the case for positive amnesia.*

The implication of this proposition can account for the key findings of our experimental studies. Particularly, it accounts for Result 3, 4, and 5, which to our knowledge are together not compatible with alternative models in the literature. Note that when self-0 receives a bad signal, self-1 will face three possibilities eventually: a bad signal, no signal and a good signal corresponding to perfect recall, amnesia and confabulation, respectively. When present bias is severe, individuals will exhibit positive memory bias regardless of whether he receives a bad signal or no signal. This induces a motive for transforming a bad signal via confabulation into a good one. In comparison with our model without confabulation, our analysis here offers an alternative mechanism to deliver over confidence without its reliance on the functional role of amnesia.

4 Concluding Remarks

People often forget and sometimes fantasize. Casual introspection suggests that human being cannot consciously choose to forget or be delusional. Yet, while memory bias including false memory is inherently not a conscious process, they tend to possess directionality in terms of a tendency to forget bad signals as well as to fantasize positively. Since the seminal work of Hebb (1949), ample evidence in neuroscience shows that environmental stimuli give rise to dramatic changes in brain functions

including altered learning and memory process (see, e.g., Williams et al., 2001; Meshi et al., 2006). At the same time, we do observe people engaging consciously in specific acts to facilitate their forgetting certain bad signals, e.g., leaving a place to avoid bad memories or burning photos of ex-spouses, and also to induce fake but good signals, such as addiction to soap operas or obsession with video games, possibly to enhance one’s self image. Notably, in a review paper, Howe’s (2011) suggestion of adaptive functionality in delusional disposition corroborates our overall finding of positive amnesia and positive false memory. As such, we posit that memory choice including amnesia and false memory is generally made nonconsciously. Our findings both theoretical and experimental allow for forgetting bad signals and romanticizing fake signals reflects the empirical observations as reported in McKay and Dennett (2009), Howe and Derbish (2010) and Howe et. al. (2011).

Our 3-signal extension of the B-T model admits a natural reinterpretation in a multi-person setting requiring no substantive change in the game form. Benabou (2013) investigates collective memory bias and its function in forming social cognition, and shows that collective delusion in terms of denial of information may be contagious in society with preference over beliefs, generating multiple social cognitions. Such a reinterpretation can provide a rationale for the motivational value of myth making, e.g., telling fairy tales to induce children to form a more rosy view of the world corresponding to a belief in an enhanced chance of future happiness. Our theory suggests that telling tales may conceivably be more functional than omitting information such as a lack of academic achievements. In this case, self-0 represents the older generation while self-1 refers to the younger generation. Besides present bias at the individual level, the β parameter may also capture a degree of altruism of the current generation towards the future generation. Here, forgetfulness has the natural interpretation as the collective amnesia of the younger generation. For example, older generations in Japan could revise historical texts by down playing the event of ‘Nanking massacre’ for the young.

Closest to our collectivist reinterpretation is the work by Dessi (2008) who studies collective memory and cultural transmission, and explains how information suppression at the societal level alleviates the free riding problem. Moreover, our model with

confabulation can offer an account for collective confabulation in transforming past disastrous events into myths, legends, and Utopian tales to be transmitted across generations, e.g., the official account of China’s great leap forward in which the younger generation is taught about its many virtues. Local leaders falsely reported high grain production figures as they witnessed mass starvation and famine (see, e.g., Yang et. al., 2012). Our approach can give rise to a fresh take on how institutional fabrication including collective amnesia, collective delusion, and collective confabulation can enhance confidence at the societal level, thereby motivating people to invest in the collective good from the perspective of the older generations.

References

- [1] Benabou, R. and J. Tirole (2002). “Self-Confidence and Personal Motivation,” *Quarterly Journal of Economics*, 117(3), 871-915.
- [2] Benabou, R. (2013). “Groupthink: Collective Delusions in Organizations and Markets,” *Review of Economic Studies*, 80, 429-462.
- [3] Benoit, J. P. and J. Dubra (2011). “Apparent Overconfidence,” *Econometrica*, 79(5), 1591-1625.
- [4] Bortolotti, M. and M. Mameli (2012). “Self-Deception, Delusion and the Boundaries of Folk Psychology,” *Humanamente*, 20, 203–221.
- [5] Brown, T. L., R. T. A. Croson, and C. Eckel (2011). “Intra- and Inter-personal Strategic Ignorance: A Test of Carrillo and Mariotti,” working paper.
- [6] Burks, S. V., J. P. Carpenter, L. Goette and A. Rustichini (2013). “Overconfidence and Social-Signaling,” *Review of Economic Studies*, 80(3), 949-983.
- [7] Camerer, C., G. Loewenstein, and M. Rabin (2003). *Advances in Behavioral Economics*, Princeton University Press.
- [8] Carrillo, J. and T. Mariotti (2000). “Strategic Ignorance as a Self-Disciplining Device,” *Review of Economic Studies*, 67(3), 529-44.

- [9] Cho, I. K. and D. M. Kreps, (1987). "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179-221.
- [10] Cross, K. P. (1977). "Not Can, but Will College Teaching be Improved?" *New Directions for Higher Education*, 1977 (17), 1-15.
- [11] Dessi, R. (2008). "Collective Memory, Cultural Transmission and Investments," *American Economic Review*, 98(1), 534-560.
- [12] Dow, J. (1991). "Search Decisions with Limited Memory," *Review of Economic Studies*, 58(1), 1-14.
- [13] Frenda, S.J, Nichols, R.M., and E.F.Loftus (2011). "Current Issues and Advances in Misinformation Research," *Current Directions in Psychological Science*, 20, 20-23.
- [14] Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley, 1949.
- [15] Howe, M. L. (2011). "The Adaptive Nature of Memory and Its Illusions," *Psychological Science*, 20(5), 312-315.
- [16] Howe, M. L. and M. H. Derbish (2010). "On The Susceptibility of Adaptive Memory To False Memory Illusions," *Cognition*, 115(2), 252-267.
- [17] Howe, M.L., S. R. Garner, M. Charlesworth, and L. Knott, (2011). "A Brighter Side To False Memory Illusions: False Memories Can Prime Children's And Adults' Insight-based Problem Solving," *Journal of Experimental Child Psychology*, 108, 383-393.
- [18] Kishida, K. T., B. King-Casas, and P. R. Montague. (2010). "Neuroeconomic Approaches to Mental Disorders," *Neuron*, 67, 543-554.
- [19] Koszegi, B. (2006). "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, 4(4), 673-707.
- [20] Laibson, D. (1997). "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 62, 443-477.

- [21] Loftus, E.F., and J.C. Palmer, (1974). "Reconstruction of Automobile Destruction: An example of the interaction between language and memory," *Journal of Verbal Learning and Verbal Behavior*, 13: 585-589.
- [22] Loftus, E.F. (1993). "The Reality of Repressed Memories," *American Psychologist*, 48(5):518-537.
- [23] McKay, R., R. Langdon and M. Coltheart, (2005). "Paranoia, Persecutory Delusions and Attributional Biases," *Psychiatry Research*, 136(2-3), 233-245.
- [24] McKay, R. T. and D. C. Dennett (2009). "The Evolution of Misbelief," *Behavioral and Brain Sciences*, 32, 493-561.
- [25] Meshi D., M.R. Drew, M. Saxe, M.S. Ansorge, D. David, L. Santarelli, C. Malapani, H. Moore, and R. Hen (2006). "Hippocampal Neurogenesis is Not Required for Behavioral Effects of Environmental Enrichment," *Nature Neuroscience*, 9, 729-731.
- [26] Pashler, H. (1998). *The Psychology of Attention*, MIT Press, Cambridge, MA.
- [27] Ramachandran, V. S. (1996). "The Evolutionary Biology of Self-deception, Laughter, Dreaming and Depression: Some Clues From Anosognosia," *Medical Hypotheses*, 47(5), 347-362.
- [28] Sims, C. A. (2003). "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665-690.
- [29] Strotz, R. (1955). "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23, 165-180.
- [30] Svenson, O. (1981). "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" *Acta Psychologica*, 94, 143-148.
- [31] Whitty, C. W. M. and W. A. Lewin. (1960). "Korsakoff Syndrome in the Post-cingulectomy Confusional State," *Brain*, 83, 648-653.
- [32] Williams, B.M., Y. Luo, C. Ward, K. Redd, R. Gibson, S.A. Kuczaj, J.G. McCoy (2001). "Environmental Enrichment: Effects on Spatial Memory and Hippocampal CREB Immunoreactivity," *Physiology and Behavior*, 73, 649-658.

- [33] Yang, J., E. Friedman, S. Mosher, G. Jian, R. MacFarquhar (2012). *Tombstone: The Great Chinese Famine, 1958-1962*, Farrar, Straus and Giroux.

Appendix

A Proof of Propositions

A.1 Proof of Proposition 1

(i) *PBE1* ($h_B^* = 1, h_\emptyset^* = 1$): We have $r^*(\emptyset) = 1$ and $r^*(G) = 1$. When $s = B$, define $\chi[r^*(\emptyset), \beta] = U_S(\theta_B) - U_T(\theta_B)$. This is given by:

$$\int_{\beta\theta_B V}^{\beta\theta^*(\emptyset)V} \{\theta_B V - c\} dF(c) = \int_{\beta\theta_B V}^{\beta\theta_\emptyset V} \{\theta_B V - c\} dF(c).$$

Notice that $\chi[r^*(\emptyset), 1] < 0$ and that $\chi[r^*(\emptyset), \beta] > 0$ for $\beta \in (0, \theta_B/\theta^*(\emptyset)]$. It follows that there exists $\beta' \in (\theta_B/\theta^*(\emptyset), 1)$ such that $\chi[r^*(\emptyset), \beta'] = 0$. Moreover, $\chi[r^*(\emptyset), \beta]$ is positive for $\beta \in (0, \beta')$ and negative for $\beta \in (\beta', 1)$, and

$$\frac{\partial \chi[r^*(\emptyset), \beta]}{\partial \beta} = \theta^*(\emptyset) V^2 [\theta_B - \beta\theta^*(\emptyset)] f[\beta\theta^*(\emptyset)V] - \theta_B V^2 [\theta_B - \beta\theta_B] f(\beta\theta_B V) < 0,$$

for $\beta \in [\theta_B/\theta^*(\emptyset), 1]$. Therefore, $h_B^* = 1$, if $\beta > \beta'$.

When $s = \emptyset$, we can similarly define $\Psi[r^*(\emptyset), r^*(G), \beta] \equiv U_F(\theta_\emptyset) - U_T(\theta_\emptyset)$ which is given by:

$$\int_{\beta\theta^*(\emptyset)V}^{\beta\theta^*(G)V} \{\theta_\emptyset V - c\} dF(c) = \int_{\beta\theta_\emptyset V}^{\beta\theta_G V} \{\theta_\emptyset V - c\} dF(c).$$

We have that $\Psi[r^*(\emptyset), r^*(G), 1] < 0$ and $\Psi[r^*(\emptyset), r^*(G), \beta] > 0$, for $\beta \in (0, \theta_\emptyset/\theta_G]$. Thus, $\Psi[r^*(\emptyset), r^*(G), \beta''] = 0$ for some $\beta'' \in (\theta_\emptyset/\theta_G, 1)$. Moreover, $\Psi[r^*(\emptyset), r^*(G), \beta]$ is positive for $\beta \in (0, \beta'')$ and negative for $\beta \in (\beta'', 1)$, and

$$\frac{\partial \Psi[r^*(\emptyset), r^*(G), \beta]}{\partial \beta} = \theta_G V^2 [\theta_\emptyset - \beta\theta_G] f(\beta\theta_G V) - \theta_\emptyset V^2 [\theta_{B\emptyset} - \beta\theta_\emptyset] f(\beta\theta_\emptyset V) < 0,$$

for $\beta \in [\theta_\emptyset/\theta_G, 1]$. Thus, $h_\emptyset^* = 1$, if $\beta > \beta''$.

It follows that a correct recall PBE1 ($h_B^* = 1, h_\emptyset^* = 1$) exists for $\beta > \max\{\beta', \beta''\}$.

(ii) *PBE2* ($h_B^* = 0, h_\emptyset^* = 1$): We have $r^*(\emptyset) = (1 - q)/(qp + 1 - q)$ and $r^*(G) = 1$. When $s = B$,

$$\chi[r^*(\emptyset), \beta] = \int_{\beta\theta_B V}^{\beta[\frac{1-q}{qp+1-q}\theta_\emptyset + \frac{qp}{qp+1-q}\theta_B]V} \{\theta_B V - c\} dF(c).$$

As with the proof of existence of PBE1, there exists $\bar{\beta}_2$ such that $\chi[r^*(\emptyset), \beta] > 0$ for $\beta \in (0, \bar{\beta}_2)$ and $\chi[r^*(\emptyset), \beta] < 0$ for $\beta \in (\bar{\beta}_2, 1)$. It follows that $h_B^* = 0$ if $\beta < \bar{\beta}_2$.

When $s = \emptyset$,

$$\Psi[r^*(\emptyset), r^*(G), \beta] = \int_{\beta(\frac{1-q}{qp+1-q}\theta_\emptyset + \frac{qp}{qp+1-q}\theta_B)V}^{\beta\theta_G V} \{\theta_\emptyset V - c\} dF(c).$$

We can similarly conclude that there is a $\underline{\beta}_2$ such that $\Psi[r^*(\emptyset), r^*(G), \underline{\beta}_2] = 0$ so that $h_\emptyset^* = 1$ for $\beta > \underline{\beta}_2$ since $\Psi[r^*(\emptyset), r^*(G), \beta] > 0$ for $\beta \in (0, \underline{\beta}_2)$ and $\Psi[r^*(\emptyset), r^*(G), \beta] < 0$ for $\beta \in (\underline{\beta}_2, \theta_\emptyset/\theta^*(\emptyset))$.

The existence of PBE2 requires $0 < \underline{\beta}_2 < \bar{\beta}_2 < 1$. While it is not necessary to solve explicitly for the thresholds in the other equilibria, under the assumption of uniform distribution these two thresholds can be solved explicitly as displayed below:

$$\bar{\beta}_2 = \frac{2(1 - q + qp)\theta_B}{(1 - q + p + qp)\theta_B + (1 - p)(1 - q)\theta_G}$$

and

$$\underline{\beta}_2 = \frac{2[1 + (p - 1)q][p(\theta_B - \theta_G) + \theta_G]}{2(1 - q)\theta_G + p[\theta_B + (2q - 1)\theta_G]}.$$

It follows that $1/2 < p < 1$ and $1 < \gamma < p(p - q + qp)/(1 - p)^2(1 - q)$, which can be satisfied for sufficiently large p . Thus PBE2 ($h_B^* = 0, h_\emptyset^* = 1$) exists if $\underline{\beta}_2 < \beta < \bar{\beta}_2$.

(iii) *PBE3a* ($h_B^* = 1, h_\emptyset^* = 0$): We have $r^*(G) = (q - qp)/(1 - qp)$ with $r^*(\emptyset)$ arbitrary since it is an off the equilibrium path belief. When $s = B$,

$$\chi[r^*(\emptyset), \beta] = \int_{\beta\theta_B V}^{\beta[r^*(\emptyset)\theta_\emptyset + (1 - r^*(\emptyset))\theta_B]V} \{\theta_B V - c\} dF(c).$$

Notice that $\chi[r^*(\emptyset), \beta] = 0$ and $h_B^* = 1$ when $r^*(\emptyset) = 0$. If $r^*(\emptyset) > 0$, we can show similarly that there exists a $\underline{\beta}_3$ such that $\chi[r^*(\emptyset), \underline{\beta}_3] = 0$ with $\chi[r^*(\emptyset), \beta]$ positive or negative depending on whether β is less than or greater than $\underline{\beta}_3$. Thus $h_B^* = 1$, if $\beta > \underline{\beta}_3$.

Similarly, when $s = \emptyset$,

$$\Psi[r^*(\emptyset), r^*(G), \beta] = \int_{\beta[r^*(\emptyset)\theta_0 + (1-r^*(\emptyset))\theta_B]V}^{\beta[\frac{q-qp}{1-qp}\theta_G + \frac{1-q}{1-qp}\theta_0]V} \{\theta_\emptyset V - c\} dF(c).$$

We can show that there is a $\bar{\beta}_3$ solving $\Psi[r^*(\emptyset), r^*(G), \bar{\beta}_3] = 0$ such that $\Psi[r^*(\emptyset), r^*(G), \beta]$ is positive or negative depending on whether β is less than $\bar{\beta}_3$ or greater than $\bar{\beta}_3$ but less than $\theta_0/\theta^*(\emptyset)$. Thus, $h_\emptyset^* = 0$ if $\beta < \bar{\beta}_3$. It follows that PBE3a ($h_B^* = 1$, $h_\emptyset^* = 0$) exists if $\underline{\beta}_3 < \beta < \bar{\beta}_3$.

Now we consider the intuitive criteria to refine this equilibrium. Note that this equilibrium is determined by the off-equilibrium belief $r^*(\emptyset)$. When $r^*(\emptyset) = 0$, self- B is indifferent between recall and amnesia. When $r^*(\emptyset) > 0$, he will choose to recall the bad signal for sufficiently large β . For type- \emptyset self, regardless of the value of $r^*(\emptyset)$, delusion is always strictly better than correct recall when $\beta < \bar{\beta}_3$. Thus after the equilibrium refinement under the intuitive criterion, self-1 knows that self-0 of type- \emptyset will not correctly recall not having received a signal. Receiving such an empty signal precludes being type- \emptyset , so that the off-equilibrium-path belief $r^*(\emptyset)$ can only be zero. In this case, the decision maker is indifferent between correct recall and amnesia and would not need a large β to remain self truthful. Thus, when β is small enough, this equilibrium prevails.

PBE3b ($h_B^* = 0$, $h_\emptyset^* = 0$): We have $r^*(\emptyset) = 0$ and $r^*(G) = (q - qp)/(1 - qp)$. When $s = B$,

$$\chi[r^*(\emptyset), \beta] = \int_{\beta\theta_B V}^{\beta\theta_B V} \{\theta_B V - c\} dF(c) = 0.$$

Thus $\chi[r^*(\emptyset), \beta] = 0$, i.e., self- B has no incentive to deviate from suppressing the bad signal for any β .

When $s = \emptyset$,

$$\Psi[r^*(\emptyset), r^*(G), \beta] = \int_{\beta\theta_B V}^{\beta[\frac{q-qp}{1-qp}\theta_G + \frac{1-q}{1-qp}\theta_0]V} \{\theta_\emptyset V - c\} dF(c).$$

Similarly, we can show that there exists $\bar{\beta}_3$ such that $\Psi[r^*(\emptyset), r^*(G), \bar{\beta}_3] = 0$ and that $\Psi[r^*(\emptyset), r^*(G), \beta]$ is positive for $\beta \in (0, \bar{\beta}_3)$ and is negative for $\beta \in (\bar{\beta}_3, \theta_0/\theta_B]$. Therefore $h_\emptyset^* = 0$ if $\beta < \bar{\beta}_3$. it follows that PBE3b ($h_B^* = 0$, $h_\emptyset^* = 0$) exists if $\beta < \bar{\beta}_3$.

Q.E.D.

A.2 Proof of Proposition 3

Following the analysis of Proposition 1, we can obtain 16 PBE's in the absence of refinement by the intuitive criterion. The pattern of these 16 PBE's resembles Table 1 except for PBE3h in which we have amnesia. After applying the intuitive criterion, we can identify the 16 possible PBE's listed in Table 1. Let the reliability of signal \emptyset and signal G to self-0⁺ be $r^*(\emptyset)$ and $r^*(G)$ respectively, and that to self-1 be $r^{**}(\emptyset)$ and $r^{**}(G)$.

Take PBE3c as an example. This equilibrium is determined by the off-equilibrium belief $r^*(\emptyset)$. In this equilibrium, no matter what $r^*(\emptyset)$ is, delusion is always strictly better than correct recall for self-0 of type- \emptyset ; while for self-0 of type- B , correct recall and amnesia are indifferent when $r^*(\emptyset) = 0$. After the refinement, when self-1 receives an empty signal, he will know that he is type- B , and thus the amnesia becomes fake.

PBE3e offers another example. Here, self-0⁺ of type- \emptyset strictly prefers delusion rather than correct recall, while self-0⁺ of type- B is indifferent between correct recall and amnesia when $r^{**}(\emptyset) = 0$. Thus, after refinement, the belief can only be $r^{**}(\emptyset) = 0$.

We can similarly refine the pre-intuitive-criterion equilibria corresponding to PBE3a, PBE3f, PBE3h, PBE4a and PBE4b and obtain the results summarized in Table 1.

Q.E.D.

B. Experimental Instruments

B.1 Memory Test on IQ Performance

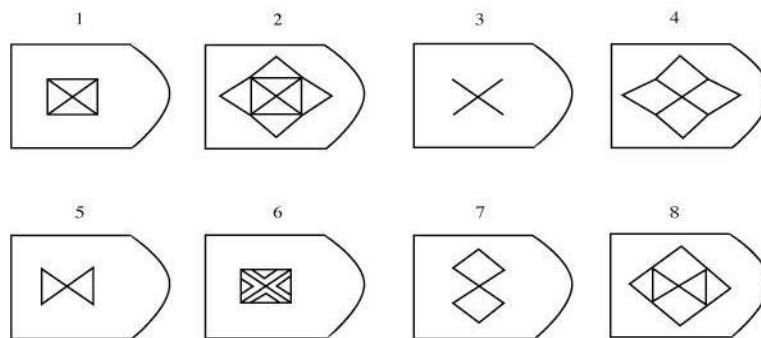
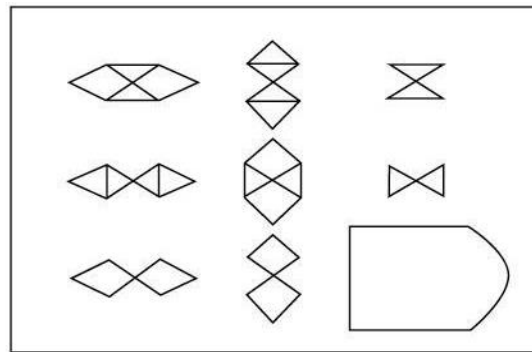
Each subject comes across 6 memory tasks. Each memory task consists of a Ravens IQ question together with the correct answer. Subjects are asked to choose one of options relating to their recall of their performance of the specific question. The specific wording used for each task is appended below. The actual Ravens questions used are displayed in the subsequent pages.

You are asked to recall your performance on some questions you may have attempted in Wave 1 of our study. After being presented with a question together with its correct answer, you can choose:

- a. My response was correct.
- b. My response was incorrect.
- c. I didn't see this question.
- d. I don't remember.

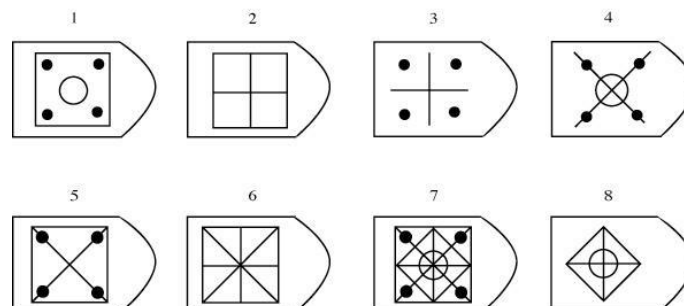
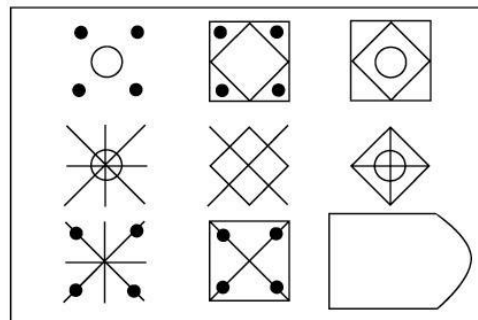
You will receive \$1 for a correct response or lose \$1 for an incorrect response if you choose (a), (b), or (c). If you choose (d), you will receive \$0

Question 1



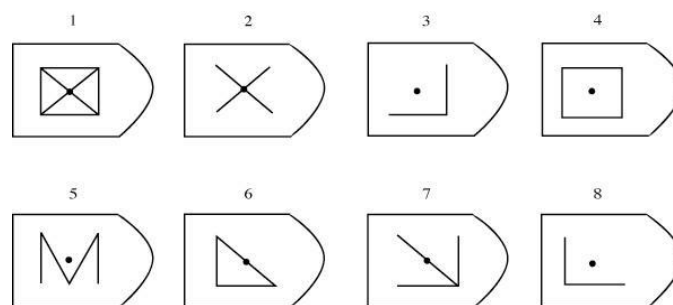
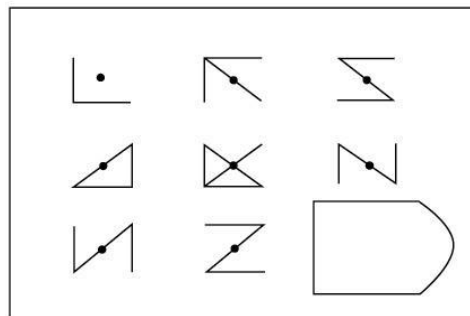
The correct answer is 3.

Question 2



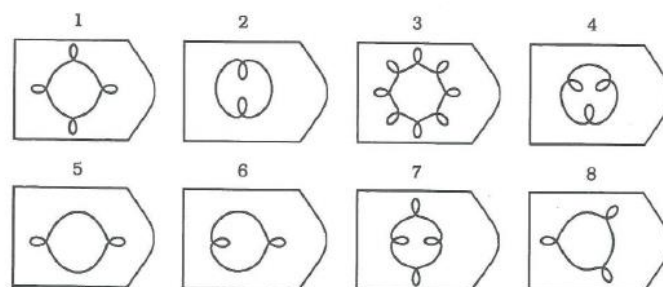
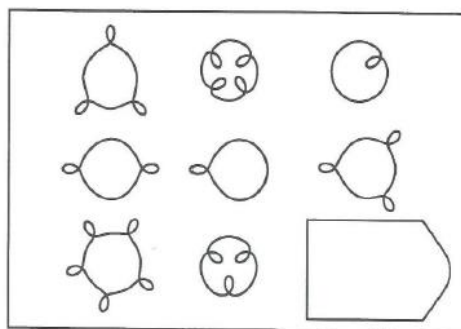
The correct answer is 2.

Question 3



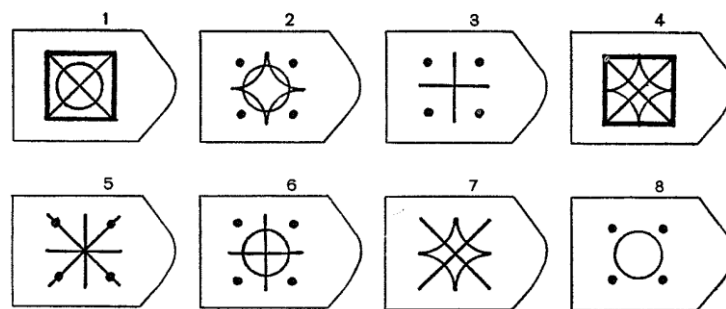
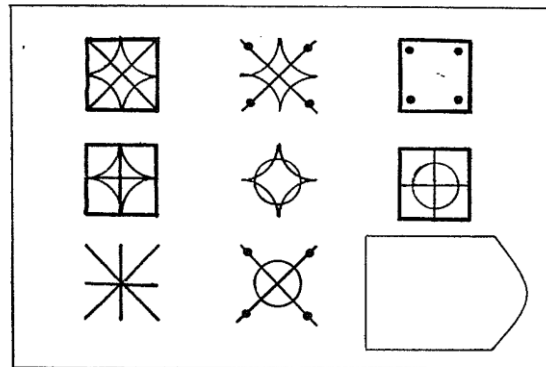
The correct answer is 4.

Question 4



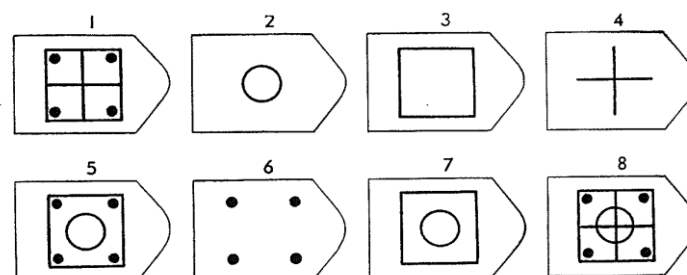
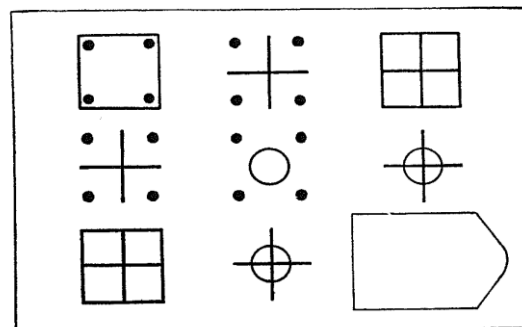
The correct answer is 5.

Question 5



The correct answer is 6.

Question 6



The correct answer is 7.

B.2 Elicitation of Risk Attitude through a Portfolio Choice Task

The degree of risk aversion is assessed based on the subject's decisions on how much of a given amount cash to allocate to purchasing a number of even-chance lottery which pays on average a positive amount. The actual task (appended) elicits one of 11 levels of risk aversion from holding all cash (coded as 0) to holding no cash (coded as 10).

In this task, you are endowed with \$27. You have the option to invest an amount on an experimental stock constructed from a deck of 20 cards comprising 10 black cards and 10 red cards. For every dollar invested, you receive \$2.5 if you guess the color of a randomly drawn card correctly. Otherwise, you receive \$0 and lose your investment. The following table displays your investment options which consist of investing between \$0 and \$27 in steps of \$3 in this experimental stock and keeping the rest as cash. The last two columns indicate your Total Earnings given by Cash + Investment Return for the cases of correct and incorrect guesses respectively.

DECISION: For the following 10 options, please indicate your decision with a tick (✓).

Investment				Total Earnings	
	✓	Cash	Invest	Correct Guess	Incorrect Guess
1		\$27.00	\$0.00	\$27.00	\$27.00
2		\$24.00	\$3.00	\$31.50	\$24.00
3		\$21.00	\$6.00	\$36.00	\$21.00
4		\$18.00	\$9.00	\$40.50	\$18.00
5		\$15.00	\$12.00	\$45.00	\$15.00
6		\$12.00	\$15.00	\$49.50	\$12.00
7		\$9.00	\$18.00	\$54.00	\$9.00
8		\$6.00	\$21.00	\$58.50	\$6.00
9		\$3.00	\$24.00	\$63.00	\$3.00
10		\$0.00	\$27.00	\$67.50	\$0.00

B.3 Elicitation of Attitude towards Ambiguity

The degree of ambiguity aversion, AA, is assessed using the difference in certainty equivalents elicited for an even-chance lottery (deck of 10 red and 10 black cards) and for an ambiguous lottery (deck of 20 cards with unknown numbers of red and black cards). Using the switch point for each lottery as proxy for its certainty equivalent, we can elicit a 21-level measure (from -10 to 10) of ambiguity aversion by subtracting the switch point for the ambiguous lottery from the switch point of the even-chance lottery.

Even-chance Lottery¹

This situation involves your guessing the color – red or black – of a card drawn randomly from a deck of 20 cards, comprising **10 black** cards and **10 red** cards.

Option A: You guess the color-black or red, and then draw a card from the 20 cards. If you make a correct guess, you receive \$60; otherwise, you receive nothing. That is: **50% chance of receiving \$60 and 50% chance of receiving \$0.**

The **Option B** column lists the amounts you will receive for sure if you choose this option.

DECISION: For each of the 10 rows, please indicate your decision in the final column with a tick (✓).

	Option A	Option B	Decision
1	50% of receiving \$60, 50% of receiving \$0	Receiving \$15 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
2	50% of receiving \$60, 50% of receiving \$0	Receiving \$19 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
3	50% of receiving \$60, 50% of receiving \$0	Receiving \$23 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
4	50% of receiving \$60, 50% of receiving \$0	Receiving \$25 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
5	50% of receiving \$60, 50% of receiving \$0	Receiving \$27 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
6	50% of receiving \$60, 50% of receiving \$0	Receiving \$29 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
7	50% of receiving \$60, 50% of receiving \$0	Receiving \$30 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
8	50% of receiving \$60, 50% of receiving \$0	Receiving \$31 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
9	50% of receiving \$60, 50% of receiving \$0	Receiving \$33 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
10	50% of receiving \$60, 50% of receiving \$0	Receiving \$35 for sure	A <input type="checkbox"/> B <input type="checkbox"/>

¹This lottery can also be used to elicit risk aversion. While it has been used to measure ambiguity aversion, in order to prevent the multicollinearity in regression, we apply a different lottery for risk attitude.

Ambiguous Lottery

This situation involves your drawing randomly one card from a deck of 20 cards with unknown proportions of red and black cards.

Option A: Guess the color of a card to be drawn randomly by you from a deck of 20 cards with unknown proportions of red and black cards. You will receive \$60 if your guess is correct; and \$0 otherwise.

The **Option B** column lists the amounts you can receive for sure if you choose this option.

DECISION: For each of the 10 rows, please indicate your decision in the final column with a tick (✓).

	Option A	Option B	Decision
1	Betting on the color of a card drawn	Receiving \$15 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
2	Betting on the color of a card drawn	Receiving \$19 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
3	Betting on the color of a card drawn	Receiving \$23 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
4	Betting on the color of a card drawn	Receiving \$25 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
5	Betting on the color of a card drawn	Receiving \$27 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
6	Betting on the color of a card drawn	Receiving \$29 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
7	Betting on the color of a card drawn	Receiving \$30 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
8	Betting on the color of a card drawn	Receiving \$31 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
9	Betting on the color of a card drawn	Receiving \$33 for sure	A <input type="checkbox"/> B <input type="checkbox"/>
10	Betting on the color of a card drawn	Receiving \$35 for sure	A <input type="checkbox"/> B <input type="checkbox"/>

B.4 Elicitation of Time Preference through a Multiple List Task

The degree of time preference discount rates are elicited from comparisons between their tradeoffs in a proximal task (next day versus 30 days later) and a distal task (351 days versus 381 days later). The switch point in distal task elicits normal discount rate, while the ratio of the switch points in two tasks elicits the present bias.

This task involves your choosing between receiving a sum of money on a specific day and another sum of money on another specific day. There are **20 choices to make**. The first ten pairs of choices are about receiving \$100 tomorrow versus receiving a larger amount 31 days later; the second ten pairs of choices are about receiving \$100 in 351 days versus receiving a larger amount of money in 381 days.

For this task, we will pay one randomly selected **participant** in this room at the end of today's study. For this participant, we will choose randomly **one** out of the 20 choices and pay him/her accordingly. Specifically, we will give him/her a **cheque** with the specified date at the end of **today's experiment**. Under Singapore banking practices, a cheque can be cashed only on or within 6 months of the date of the cheque.

DECISION: For each of the 20 rows, please indicate your decision in the final column with a tick (✓).

	A	Tomorrow	B	31 days later
1	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$101
2	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$104
3	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$107
4	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$110
5	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$113
6	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$116
7	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$119
8	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$122
9	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$125
10	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$128
	A	351 days later	B	381 days later
11	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$101
12	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$104
13	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$107
14	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$110
15	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$113
16	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$116
17	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$119
18	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$122
19	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$125
20	<input type="checkbox"/>	\$100	<input type="checkbox"/>	\$128

C. Statistical Results on Memory Patterns

Our statistical results on memory patterns are described in two sections. Section C.1 comprises several tables. Table C1 displays the overall memory patterns in terms of the frequencies of each of four choices – a (*I did it right*), b (*I did it wrong*), c (*I did not see it before*), and d (*I don't remember*) – corresponding the 3 possible initial performance – correct, incorrect, and absent. Table C2 compares the proportion of positive memory biases with the corresponding proportion of negative memory bias and shows that in each case, the frequency of positive memory bias is significantly higher than the corresponding frequency of negative memory bias. Table C3 compares the proportions of positive delusion for Q5 and Q6 with the rates of correct response for Q1 – Q4 and shows that the proportion of positive delusion for each of Q5 and Q6 is significantly higher than each of the proportions of correct responses for Q1 to Q4. Table C4 displays the relation between positive delusion and positive confabulation.

Section C.2 reports the results of probit regression of the various memory biases with respect to a number of factors including present bias, IQ, risk aversion, ambiguity aversion and gender.

C.1 Patterns of recall, delusion and confabulation

			a - did it right				b - did it wrong				c - didn't see it				d - don't remember				Total																																												
Did appear	2804	s = G	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	2076	74.0%																																											
			396	366	305	281	30	35	25	22	40	37	15	9	142	143	125	105																																													
			9.4%	8.7%	7.3%	6.7%	0.7%	0.8%	0.6%	0.5%	1.0%	0.9%	0.4%	0.2%	3.4%	3.4%	3.0%	2.5%																																													
			1348				112				101				515																																																
		s = B	32.0%				2.7%				2.4%				12.2%				728	26.0%																																											
			Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4																																													
			24	39	57	66	24	21	60	101	6	3	8	9	39	57	106	108																																													
0.6%				0.9%				1.4%				1.6%				0.6%				0.5%				1.4%				2.4%				0.1%				0.1%				0.2%				0.2%				0.9%				1.4%				2.5%				2.6%			
186				206				26				310																																																			
4.4%				4.9%				0.6%				7.4%																																																			
Didn't appear	1402	s = φ	Q5		Q6		Q5		Q6		Q5		Q6		Q5		Q6		1402	33.3%																																											
			394		439		30		36		60		34		217		192																																														
			9.4%		10.4%		0.7%		0.9%		1.4%		0.8%		5.2%		4.6%																																														
			833		66		94		409																																																						
			19.8%		1.6%		2.2%		9.7%																																																						
	4206	Total	2367				384				221				1234				4206																																												
	100%		56.3%				9.1%				5.3%				29.3%				100%																																												
Seen it before - 2768 (65%)																					Didn't see it or forgot - 1478 (35%)																																										

Table C1. Overall memory patterns.

Q#	Valence	Rate of Amnesia	Std. Err.	Std. Dev.	t	p	Rate of Confabulation	Std. Err.	Std. Dev.	t	p
1	Positive	65.22%	0.0578	0.4798	5.677	0.0000	50.00%	0.0729	0.5053	9.706	0.0000
	Negative	31.49%	0.0193	0.4649			7.04%	0.0124	0.2562		
2	Positive	74.07%	0.049	0.441	7.395	0.0000	65.00%	0.0621	0.481	12.9	0.0000
	Negative	32.97%	0.0201	0.4705			8.73%	0.0141	0.2826		
3	Positive	65.52%	0.0361	0.4767	8.135	0.0000	48.72%	0.0464	0.502	11.15	0.0000
	Negative	31.46%	0.022	0.4649			7.58%	0.0146	0.265		
4	Positive	53.67%	0.0339	0.4998	6.248	0.0000	39.52%	0.0379	0.4904	9.324	0.0000
	Negative	28.86%	0.0228	0.4537			7.26%	0.0149	0.2599		
Overall	Positive	61.99%	0.0209	0.4859	13.46	0.0000	47.45%	0.0253	0.5	21.21	0.0000
	Negative	31.36%	0.0105	0.4641			7.67%	0.007	0.2662		

Table C2. Memory error proneness conditional on valence. The t-values and p-values across consecutive rows indicate the level of significance of the difference between the rate of a positive bias and the rate of a negative bias.

		Rate of Correct Response			
		Question 1 86.73%	Question 2 82.88%	Question 3 67.05%	Question 4 59.49%
Rate of Positive Delusion	Question 5 92.92%	0.0006	0.0000	0.0000	0.0000
	Question 6 92.42%	0.0011	0.0000	0.0000	0.0000

Table C3. Significance of positive delusion relative to rates of correct response

Question	1	2	3	4	Question	5	6
Rate of Positive Confabulation with No Positive Delusion	26.67%	42.86%	36.67%	25.71%	Rate of Positive Delusion with Positive Confabulation	73.19%	78.36%
Rate of Unconditional Positive Confabulation	50.00%	65.00%	48.72%	39.52%	Rate of Unconditional Positive Delusion	58.72%	66.02%
p-value	0.058	0.065	0.1202	0.0627	p-value	0.0007	0.0026

Table C4. Relation between positive delusion and positive confabulation

C.2 Probit Regression on Memory Patterns

Positive Amnesia

Question 1

	1	2	3	4	5	6	7
beta	0.7941675 (0.658)	0.8082918 (0.652)	0.9394467 (0.592)	1.264741 (0.469)	1.871065 (0.332)	1.967551 (0.308)	1.826994 (0.352)
delta		0.3886214 (0.849)	0.3782758 (0.853)	0.7180619 (0.727)	0.5519284 (0.790)	0.5969196 (0.774)	0.587779 (0.779)
IQ			0.0192332 (0.513)	0.0201985 (0.485)	0.0261768 (0.378)	0.0259101 (0.383)	0.0224352 (0.459)
RA				-0.0572618 (0.372)	-0.0433793 (0.511)	-0.0462668 (0.499)	-0.0527257 (0.444)
AA					-0.0480567 (0.375)	-0.0483265 (0.372)	-0.05008 (0.361)
Gender						-0.086747 (0.801)	-0.0995153 (0.774)
Duration							-0.0012986 (0.475)

Question 2

	1	2	3	4	5	6	7
beta	0.8557124 (0.611)	0.696621 (0.679)	0.8145705 (0.616)	1.007389 (0.538)	0.9710023 (0.557)	0.9745587 (0.551)	0.7396312 (0.669)
delta		-1.421397 (0.479)	-1.495394 (0.457)	-1.137025 (0.566)	-1.130668 (0.567)	-1.136647 (0.570)	-0.8442994 (0.672)
IQ			-0.0139588 (0.636)	-0.0102958 (0.731)	-0.0105721 (0.723)	-0.0105327 (0.722)	-0.0127513 (0.672)
RA				-0.0639856 (0.306)	-0.0645767 (0.315)	-0.0647504 (0.310)	-0.0570249 (0.382)
AA					0.0071694 (0.890)	0.0071223 (0.891)	0.0060058 (0.908)
Gender						-0.00613 (0.985)	0.037209 (0.908)
Duration							-0.0024316 (0.269)

Question 3

	1	2	3	4	5	6	7
beta	-0.8249069 (0.481)	-0.8564155 (0.471)	-0.8896165 (0.454)	-0.7696302 (0.519)	-0.8650621 (0.471)	-0.9276477 (0.441)	-0.9154441 (0.448)
delta		1.071534 (0.424)	1.221687 (0.365)	1.342288 (0.324)	1.296452 (0.343)	1.345893 (0.325)	1.384385 (0.310)
IQ			0.0423272* (0.054)	0.0433028** (0.046)	0.0424696** (0.050)	0.0434698** (0.045)	0.0445658** (0.036)
RA				-0.0504879 (0.233)	-0.0514472 (0.227)	-0.048922 (0.256)	-0.0490838 (0.255)
AA					0.0122553 (0.693)	0.0134586 (0.666)	0.0146636 (0.642)
Gender						0.126143 (0.535)	0.1259161 (0.536)
Duration							0.0002993 (0.786)

Question 4

	1	2	3	4	5	6	7
beta	-0.2620316 (0.804)	-0.3599861 (0.740)	-0.3679186 (0.734)	-0.3275821 (0.763)	-0.3235199 (0.764)	-0.2349377 (0.829)	-0.1905093 (0.862)
delta		1.870027 (0.108)	1.83408 (0.116)	1.862345 (0.112)	1.940154* (0.093)	1.8922 (0.103)	1.880756 (0.106)
IQ			0.008094 (0.662)	0.0078576 (0.672)	0.0103571 (0.580)	0.0101407 (0.588)	0.0105812 (0.572)
RA				-0.0123731 (0.737)	-0.001487 (0.968)	-0.0017145 (0.964)	-0.0006019 (0.987)
AA					-0.0507406* (0.053)	-0.0529455** (0.044)	-0.0510307* (0.052)
Gender						-0.1416088 (0.428)	-0.1489755 (0.405)
Duration							0.0008321 (0.320)

Overall

	1	2	3	4	5	6	7
beta	-0.2404396 (0.713)	-0.2518129 (0.703)	-0.2676429 (0.685)	-0.145776 (0.826)	-0.0348341 (0.959)	-0.0192549 (0.977)	-0.0104931 (0.988)
delta		0.9512304 (0.203)	0.9558642 (0.200)	1.086129 (0.147)	1.111577 (0.138)	1.102749 (0.141)	1.10358 (0.141)
IQ			0.023752** (0.038)	0.0240326** (0.034)	0.0252597** (0.026)	0.025206** (0.026)	0.0253568** (0.026)
RA				-0.0395967* (0.088)	-0.0358421 (0.127)	-0.0361704 (0.125)	-0.0361528 (0.126)
AA					-0.0234285 (0.175)	-0.0236604 (0.172)	-0.0233487 (0.179)
Gender						-0.0240491 (0.834)	-0.0246167 (0.830)
Duration							0.000111 (0.848)

Positive Delusion

Question 5

	1	2	3	4	5	6	7
beta	1.371485** (0.024)	1.350161** (0.027)	1.433774** (0.022)	1.463545** (0.019)	1.463108** (0.019)	1.46458** (0.019)	1.480271** (0.018)
delta		-0.9075226 (0.162)	-0.9859913 (0.137)	-0.9758679 (0.141)	-0.9750857 (0.142)	-0.9764553 (0.141)	-0.983469 (0.139)
IQ			-0.082087*** (0.000)	-0.082096*** (0.000)	-0.081514*** (0.000)	-0.081923*** (0.000)	-0.082068*** (0.000)
RA				-0.0131843 (0.530)	-0.0142241 (0.500)	-0.0146752 (0.489)	-0.014933 (0.481)
AA					0.0147352 (0.324)	0.0145041 (0.332)	0.0157217 (0.293)
Gender						-0.0221522 (0.827)	-0.0175577 (0.862)
Duration							0.0005618 (0.255)

Question 6

	1	2	3	4	5	6	7
beta	1.261377** (0.044)	1.257677** (0.044)	1.327283** (0.035)	1.335134** (0.034)	1.33564** (0.034)	1.35134** (0.032)	1.35519** (0.032)
delta		-0.7214664 (0.286)	-0.8191609 (0.229)	-0.8141165 (0.232)	-0.811294 (0.233)	-0.8161903 (0.230)	-0.827643 (0.224)
IQ			-0.094623*** (0.000)	-0.094592*** (0.000)	-0.094546*** (0.000)	-0.095883*** (0.000)	-0.095434*** (0.000)
RA				-0.0035093 (0.868)	-0.0037154 (0.862)	-0.0049557 (0.817)	-0.0050673 (0.813)
AA					0.0022557 (0.882)	0.0014793 (0.922)	0.0024039 (0.874)
Gender						-0.0692957 (0.511)	-0.0648083 (0.539)
Duration							0.0004184 (0.404)

Question 5 + Question 6

	1	2	3	4	5	6	7
beta	1.420578** (0.019)	1.396377** (0.021)	1.524013** (0.013)	1.553447** (0.011)	1.555384** (0.011)	1.563221** (0.011)	1.580907*** (0.010)
delta		-0.979881* (0.098)	-1.010873* (0.089)	-1.001527* (0.092)	-0.9947964* (0.094)	-0.9978245* (0.093)	-1.021137* (0.086)
IQ			-0.091721*** (0.000)	-0.091685*** (0.000)	-0.091290*** (0.000)	-0.091963*** (0.000)	-0.092255*** (0.000)
RA				-0.0120496 (0.551)	-0.0128175 (0.529)	-0.0135206 (0.507)	-0.0138294 (0.497)
AA					0.0079421 (0.565)	0.007542 (0.584)	0.0085902 (0.533)
Gender						-0.0393304 (0.680)	-0.0329615 (0.730)
Duration							0.0006065 (0.215)

Positive Confabulation

Question 1

	1	2	3	4	5	6	7
beta	5.282607** (0.035)	5.461641** (0.027)	5.259943** (0.040)	5.248981** (0.040)	5.259458** (0.043)	5.507947** (0.027)	5.484241** (0.028)
delta		3.293258 (0.134)	3.32157 (0.125)	3.573243 (0.108)	3.596333 (0.106)	4.178757* (0.053)	4.587787** (0.041)
IQ			-0.100614** (0.011)	-0.10136*** (0.009)	-0.10164*** (0.009)	-0.11063*** (0.005)	-0.110871*** (0.006)
RA				-0.0470985 (0.569)	-0.0481662 (0.559)	-0.0635196 (0.466)	-0.0688103 (0.437)
AA					0.0053849 (0.926)	0.0213925 (0.720)	0.0226309 (0.705)
Gender						-0.4257042 (0.361)	-0.384236 (0.414)
Duration							0.0010538 (0.634)

Question 2

	1	2	3	4	5	6	7
beta	5.238254* (0.065)	5.186679* (0.068)	6.078295* (0.093)	5.951469 (0.107)	5.937576 (0.102)	5.551745 (0.132)	13.28676*** (0.000)
delta		-1.0902 (0.618)	-2.144963 (0.355)	-1.759248 (0.458)	-1.815778 (0.449)	-1.604267 (0.508)	-2.161814 (0.487)
IQ			-0.1165*** (0.001)	-0.1162*** (0.001)	-0.1158*** (0.001)	-0.11706*** (0.001)	-0.18646*** (0.000)
RA				-0.0401482 (0.635)	-0.0484782 (0.553)	-0.0508861 (0.548)	-0.0314005 (0.731)
AA					-0.040097 (0.483)	-0.0316545 (0.585)	0.0000585 (0.999)
Gender						0.3802115 (0.311)	0.3107527 (0.487)
Duration							-0.01154*** (0.002)

Question 3

	1	2	3	4	5	6	7
beta	-0.474619 (0.697)	-0.5399677 (0.661)	-0.344863 (0.789)	-0.1312921 (0.919)	-0.1326745 (0.918)	-0.0059975 (0.996)	-0.0324705 (0.980)
delta		0.6887943 (0.660)	0.7712858 (0.636)	1.098409 (0.508)	1.096085 (0.508)	1.051822 (0.527)	0.8146951 (0.631)
IQ			-0.10865*** (0.008)	-0.10989*** (0.008)	-0.10989*** (0.008)	-0.11539*** (0.005)	-0.12014*** (0.004)
RA				-0.121626** (0.028)	-0.121714** (0.028)	-0.122541** (0.025)	-0.123878** (0.024)
AA					-0.0048389 (0.891)	-0.0061372 (0.862)	-0.010477 (0.771)
Gender						-0.1550712 (0.542)	-0.1695216 (0.507)
Duration							-0.0012682 (0.294)

Question 4

	1	2	3	4	5	6	7
beta	1.948055 (0.123)	2.099393* (0.098)	2.156297* (0.097)	2.351999* (0.064)	2.423644* (0.056)	2.364749* (0.062)	2.314937* (0.070)
delta		-1.067181 (0.476)	-1.399379 (0.349)	-1.236319 (0.412)	-1.084547 (0.470)	-1.070195 (0.475)	-0.9127672 (0.544)
IQ			-0.066326** (0.013)	-0.067124** (0.013)	-0.062925** (0.019)	-0.06248** (0.019)	-0.060466** (0.023)
RA				-0.0423789 (0.304)	-0.0392943 (0.340)	-0.0415939 (0.313)	-0.0419776 (0.307)
AA					-0.0418525 (0.153)	-0.0410506 (0.161)	-0.0434662 (0.133)
Gender						0.1151931 (0.572)	0.1256694 (0.537)
Duration							0.0013184 (0.196)

Overall

	1	2	3	4	5	6	7
beta	1.725033** (0.026)	1.698349** (0.029)	1.821775** (0.023)	1.956423** (0.014)	1.96093** (0.014)	1.922626** (0.016)	1.928131** (0.015)
delta		0.2805378 (0.745)	0.1511403 (0.863)	0.4529767 (0.608)	0.4741457 (0.590)	0.4822222 (0.584)	0.4645362 (0.600)
IQ			-0.0723814*** (0.000)	-0.072832*** (0.000)	-0.0719473*** (0.000)	-0.0712574*** (0.000)	-0.0714856*** (0.000)
RA				-0.0705285** (0.011)	-0.0697685** (0.012)	-0.0703107** (0.012)	-0.0702163** (0.012)
AA					-0.0301654 (0.115)	-0.0299036 (0.119)	-0.0300007 (0.119)
Gender						0.0600647 (0.652)	0.0581832 (0.664)
Duration							-0.0001113 (0.867)

Unconditional Recall (Q1 - Q4)

Recall having seen a question before (a or b vs c or d)

Question	beta	delta	IQ	RA	AA	Gender	Duration
1	-0.804336 (0.187)	0.3034358 (0.651)	0.0660972*** (0.000)	-0.0132892 (0.525)	-0.0188609 (0.212)	-0.043549 (0.669)	-0.0007232 (0.131)
2	-0.470225 (0.436)	0.0954357 (0.885)	0.0592278*** (0.000)	-0.0122586 (0.556)	-0.0087269 (0.557)	-0.090477 (0.368)	0.0002017 (0.676)
3	-0.053894 (0.930)	0.2226187 (0.738)	0.0973882*** (0.000)	0.0123747 (0.555)	0.0030649 (0.837)	-0.008689 (0.931)	0.0004736 (0.340)
4	-0.455937 (0.450)	0.677774 (0.301)	0.0600005*** (0.000)	0.0240122 (0.258)	-0.026188* (0.088)	-0.2442** (0.016)	-0.000012 (0.980)

Recall of performance (a if G and b if B)

Question	beta	delta	IQ	RA	AA	Gender	Duration
1	-0.148014	0.4859601	0.0839283***	-0.0067675	-0.0106143	-0.027418	-0.0005093
	(0.806)	(0.454)	(0.000)	(0.739)	(0.475)	(0.784)	(0.289)
2	0.1593956	0.1726147	0.0969918***	-0.0117573	-0.0197312	-0.093358	-0.000034
	(0.792)	(0.791)	(0.000)	(0.558)	(0.176)	(0.353)	(0.944)
3	-0.090765	0.0883392	0.1173619***	-0.0159512	-0.019625	-0.071429	-0.0000216
	(0.880)	(0.892)	(0.000)	(0.430)	(0.182)	(0.467)	(0.966)
4	0.4880106	-0.082445	0.0693887***	0.0160242	-0.0223612	-0.2102**	0.000368
	(0.408)	(0.897)	(0.000)	(0.428)	(0.130)	(0.033)	(0.446)

Recall of incorrect performance (b if B)

Question	beta	delta	IQ	RA	AA	Gender	Duration
1	0.7186604	0.5561472	-0.1071***	-0.0235089	0.0113897	0.1612998	-0.0008283
	(0.510)	(0.686)	(0.000)	(0.629)	(0.714)	(0.442)	(0.454)
2	2.155642***	-0.248457	-0.1155***	-0.0287729	0.012025	0.205057	-0.00298**
	(0.009)	(0.860)	(0.000)	(0.625)	(0.736)	(0.368)	(0.023)
3	-0.2127811	0.3220304	-0.0887***	-0.0375528	0.0175866	0.1520616	-0.0005887
	(0.809)	(0.731)	(0.000)	(0.193)	(0.405)	(0.297)	(0.378)
4	0.448777	1.439895	-0.0919***	-0.0182529	-0.029973	0.1075975	0.0006709
	(0.524)	(0.103)	(0.000)	(0.495)	0.113	(0.392)	(0.253)

Unconditional Delusion (a or b vs c or d for Q5 and Q6)

Question	beta	delta	IQ	RA	AA	Gender	Duration
5	-1.3828**	0.8945127	0.0576371***	0.010944	-0.015836	0.048389	-0.0004277
	(0.022)	(0.166)	(0.000)	(0.595)	(0.280)	(0.624)	(0.373)
6	-1.2047**	0.6696395	0.083026***	0.0050407	-0.002950	0.0657217	-0.0004058
	(0.046)	(0.313)	(0.000)	(0.811)	(0.843)	(0.522)	(0.405)

The number of d choice

Duration	RA	AA	beta	delta	IQ	gender
0.0000399 (0.919)	-0.0235254 (0.215)	0.0130634 (0.327)	0.557472 (0.270)	-0.5467876 (0.332)	-0.078864*** (0.000)	0.075925 (0.379)