Open access • Journal Article • DOI:10.1007/S11166-018-9295-1

Subjective beliefs and confidence when facts are forgotten — Source link

Igor Kopylov, Joshua B. Miller

Institutions: University of California, Irvine, University of Alicante

Published on: 29 Dec 2018 - Journal of Risk and Uncertainty (Springer US)

Topics: Overconfidence effect, Forgetting and Ambiguity

Subjective Beliefs And Confidence When Facts Are Forgotten

Published 2018, Journal of Risk and Uncertainty

http://link.springer.com/article/10.1007/s11166-018-9295-1

Igor Kopylov[a] and Joshua Miller[b] *

November, 2018

Abstract

Forgetting a piece of decision-relevant information is a salient source of uncertainty that should influence one's beliefs, confidence, and ambiguity attitudes. To investigate this, we run several experiments where people bet on propositions (facts) that they cannot recall with certainty. We use betting preferences to infer subjects' *revealed beliefs* and their *revealed confidence* in these beliefs. Forgetting is induced via interference tasks and time delays (up to one year). We observe a natural *memory decay* pattern where beliefs become less accurate and confidence is reduced as well. Moreover, we find a form of *comparative ignorance* where subjects are more ambiguity averse when they cannot recall the truth rather than never having learnt it. In a different vein, we identify an *overconfidence* pattern: on average, subjects overpay for bets on propositions that they believe in, but underpay for the opposite bets. We formulate a two-signal behavioral model of forgetting that generates all of these patterns. It suggests new testable hypotheses that are confirmed by our data.

JEL Classification Numbers: D81; C91.

Keywords: Memory Decay; Overconfidence; Comparative Ignorance; Revealed Beliefs; Ambiguity Aversion;

---

*a: Department of Economics, University of California, Irvine, corresponding author (ikopylov@uci.edu, 949-8246182) b: Department Of Decision Sciences, Università Bocconi. The authors are grateful to Michele Peruzzi for excellent z-Tree programming and Iris Fuli for maintaining the code and assisting during the experimental sessions. We thank Mike McBride, Gary Charness, W. Kip Viscusi, Marc Machina, Larry Epstein, Massimo Marinacci, John Duffy, Stergios Skaperdas and an anonymous referee for their comments and discussions.

1 Introduction

Forgetting is manifested in many distinct contexts and behaviors. Misidentifications in eyewitness testimonies played a role in over 70% of 358 wrongful convictions overturned by DNA evidence in the Innocence Project.[1] Military officers forget how to identify threat vehicles (Rowan [33]), radiologists cannot recall details of processed exams even after short interruptions (Froehle and White [17]), and occupational first aiders exhibit a severe loss of skill with less than 20 percent being able to perform CPR one year after training (McKenna and Glendon [27]).

Besides the accuracy (or reliability) of memory, the *confidence* that people have in their recall is essential for their own actions and others' decisions. Memory failures are especially harmful when coupled with high confidence. For example, juries are more likely to believe witnesses who appear very confident and excuse inaccuracies in their testimony compared to witnesses who appear less confident but give accurate testimony[2] (Brewer and Burke [6], Lindsay, Wells, and Rumpel [26]). Various links between the accuracy and *self-reported* confidence of people's memories have been observed in legal and experimental contexts (see reviews in Wixted and Wells [22] and Roediger, Wixted and Desoto [32]).

In this paper we present new empirical evidence on the behavioral consequences of forgetting in decision making under uncertainty. We run several experiments where people bet on propositions (facts) that they cannot recall with certainty. Forgetting is induced via interference tasks and time delays of up to one year.

Our main methodological novelty is that we capture memory failures via *betting preferences* rather than via direct verbal queries. We hypothesize that this choice-based design should identify some forms of memory decay, over/under confidence, ambiguity attitudes, and provide additional insights on the evolution of beliefs and confidence of forgetful agents.

More formally, we elicit our subjects' monetary evaluations of *binary bets*. Any such bet

$$t_B = \begin{cases} Z & \text{if } B \text{ is true} \\ 0 & \text{if } B \text{ is false} \end{cases}$$

[1]These numbers are reported at http://www.innocenceproject.org/dna-exonerations-in-the-united-states/ for the time period of 1989–2018.

[2]In every one of the DNA exoneration cases involving eyewitness misidentification examined by Garrett [18], witnesses who mistakenly identified innocent defendants did so with high confidence in a court of law.

delivers a fixed prize of $Z = 11$ euros if the corresponding proposition B is true and nothing otherwise. Prior to evaluating t_B, each subject is instructed to *memorize* whether B is true or false. Assuming perfect memory, her evaluation of t_B should be $Z = 11$ if B is true, or zero if B is false. Naturally, forgetting should decrease the value of t_B below Z if B is true, but increase it above zero if B is false.

This methodology should be more suitable than verbal queries in various economic settings where incentives are monetary and forgetting can be significant.[3] Moreover, our framework allows us to define *revealed beliefs* as in de Finetti [11] and Savage [35], and then analyze the accuracy of these beliefs.

1.1 Our findings

It is well-known that memory becomes less accurate over time. In the extensive psychological literature (see Rubin and Wenzel [34], and Brown, Neath, and Chater [7] and references therein), this phenomenon is called *memory decay* and quantified with forgetting curves. The effect of time delays on confidence is more controversial, especially in sequential settings. In particular, Shaw and McClure [36] showed that repeated questioning increased witnesses confidence without increasing their accuracy.

In our experiments, beliefs become less accurate over time, and subjects' level of confidence in their beliefs decreases. To elaborate, all participants in our experiment first learn a list $\mathcal{M}$ of true propositions. The contents of this list are novel to all participants and do not pertain to their general knowledge, skills, or previous actions. We find that for true (false) propositions B, the average monetary value of t_B diminishes (increases) after interference tasks, and then diminishes (increases) further with a time delay of one year.

To analyze memory decay in more detail, we use revealed beliefs. Given any true proposition A in the list $\mathcal{M}$, a subject reveals a *positive* belief for A if she values the ticket t_A more than $t_{\neg A}$. Similarly, her revealed belief is called *negative* if she values $t_{\neg A}$ more than t_A, or *indeterminate* if she values t_A and $t_{\neg A}$ equally. Thus we partition the entire array $\mathcal{D}$ of data points into three disjoint parts $\mathcal{P}$, $\mathcal{N}$, and $\mathcal{I}$ that correspond to positive, negative, and indeterminate beliefs respectively.

[3] For example, agents behave as if they forget fees on credit card accounts (Agarwal, Driscoll, Gabaix, and Laibson [1]), the operating costs of mutual funds (Barber, Odean, and Zheng [3]), whether or not sales tax is included in stated prices (Chetty, Looney, Kroft [9]) etc. Note that binary bets in our experiments are analogous to *Arrow securities* used in the Arrow-Debreu equilibria.

We observe that the proportion of positive beliefs $\frac{|\mathcal{P}|}{|\mathcal{D}|}$ decreases after interference tasks and then decreases further one year later. Conversely, negative and indeterminate beliefs became more common over time. A more intriguing finding is that negative beliefs are highly transient as if most of them are based on random spontaneous hunches rather than some more permanent convictions.

Next, we find *overconfidence* on the subdomain $\mathcal{P} \cup \mathcal{N}$ where revealed beliefs can be either positive or negative, but not indeterminate. On this subdomain, subjects behave as if they overestimate the average accuracy of their strict beliefs $\frac{|\mathcal{P}|}{|\mathcal{P}|+|\mathcal{N}|}$. On average, they overpay (underpay) for tickets t_B that they find more (less) valuable than the opposite bets $t_{\neg B}$. In bookmaking terms, the monetary odds that are required to bet on B are respectively too short or too long.

This pattern is roughly in line with many studies where subjects are found to overestimate their own knowledge (e.g. Fischoff, Slovic, and Lichtenstein [15], Keren [23]) ability, performance, and level of control (see Moore and Healy [28] for an overview). However, our design can be distinguished from this literature by the use of binary bets and two other features.

(i) The truthfulness of all relevant propositions is stated explicitly before evaluation stages rather than derived from general knowledge or intellectual efforts. Moreover, a subject's beliefs about the same proposition is elicited at several points in time. By contrast, Lichtenstein and Fischhoff [25] and other studies of overconfidence use questions about general knowledge and do not observe how subjects' responses change over time (presumably, because general knowledge is unlikely to vary in the short run).

(ii) Subjects evaluate propositions that pertain to objective facts, rather than to assessments of their own performance, ability, or actions. Thus we observe overconfidence without above-average effects studied by Camerer and Lovallo [8], Kirchler and Maciejovsky [24], Hoelzl and Rustichini [21], and others.

In a different vein, we use our data to identify a new kind of *ambiguity aversion*. Roughly speaking, our experiments suggest that people may be even less willing to bet when they cannot recall the truth rather than when they never have learnt it at all. We attribute this form of the Ellsberg Paradox to *comparative ignorance*. Fox and Tversky [16] use this term broadly to describe ambiguity aversion driven by a "feeling of incompetence" and a contrast with more knowledgeable decision makers (see also Heath and Tversky [19], Trautmann, Vieider, and Wakker [37]). Our subjects may exhibit comparative ignorance because of the contrast between having *forgotten* the truth and knowing it in the past.

1.2 A Two-Signal Model

Obviously, overconfidence and ambiguity aversion are inconsistent with the expected utility model and well-calibrated beliefs that are typical in game theory (e.g. Piccione and Rubinstein [30, 31]).

To accommodate the observed behavioral patterns, we formulate a two-signal model of forgetting in Section 4. We assume that agents can receive signals of two types: a memory signal that points to the true proposition with probability $\alpha \geq \frac{1}{2}$ and a noisy hunch that is objectively uninformative, but affects subjective evaluations nonetheless. We assume that the memory signal becomes less common over time. In the absence of the memory signal, subjects use their transient hunches, but become ambiguity averse. We run a simulation of the model that generates memory decay, overconfidence, and comparative ignorance. Moreover, our model suggests the following two hypotheses:

(H1) negative beliefs should be more transient than positive ones,

(H2) subjects should be more confident in positive rather than negative beliefs when their memory retains statistical accuracy.

We find both patterns in the data.

There are other models (e.g. Erev, Wallsten, and Budescu [12], Costello and Watts [10]) that introduce noise into probabilistic assessments to generate overconfidence and related biases. However, we are not aware of any such models that accommodate ambiguity aversion together with overconfidence. Moreover, our two-signal approach can be additionally motivated by the good fit that it achieves for our aggregate data and its testable hypotheses H1 and H2 that we confirm empirically.

1.3 Other Related Literature

Blavatskyy [5] studies overconfidence via evaluations of monetary bets on general knowledge questions. He also uses a different incentive compatible scheme where subjects make probabilistic rather than monetary evaluations.

He finds little overconfidence regardless of ambiguity attitudes. Erickson [13] observes a strong overconfidence about one's own future memory accuracy.[4]

Inattention provides another explanation for the failure to respond rationally to some fees and incentives (e.g. the operating costs of mutual funds in Barber et al. [3] and the sales tax in Chetty et al. [9]). In our experiment we did not have an independent measure of attention and therefore we cannot rule out the possibility of inattention explaining some of the heterogeneity in choices between subjects. Nevertheless, inattention cannot account for the key within-subject differences between treatments.[5]

In more abstract settings, memory limitations have been used to explain *dynamic inconsistency* (Piccione and Rubinstein [31], Battigalli [4]), *inertia* and impulsiveness in decisions (Hirshleifer and Welch [20]), availability heuristics and confirmation biases (Wilson [39], Mullainathan [29]), sunk cost fallacy (Baliga and Ely [2]) and other patterns. In such applications, memory failures are interpreted as a separate source of uncertainty that distorts previously available information and signals.

2 Design

We ran two sets of experimental sessions, main sessions and follow-up sessions, at the Bocconi Experimental Laboratory for the Social Sciences (BELSS) at Bocconi University using z-Tree experimental software [14]. In each type of session, upon entering the room, subjects were instructed to put on headphones and watch narrated video instructions, presented with closed captions. The precise procedures and design details can be found in Online Appendix A.2.[6]

We provide the basic details below.

2.1 Main Sessions: Learning and Partial Forgetting

Our first set of (within-subjects) sessions had a total of 98 participants. At the end of the instructions, subjects were shown a list of all true propositions (facts) that they could learn and later bet on. They were told that they could

[4]Only 53% of his subjects remembered to send an email to claim a twenty-dollar payment six months after the experiment. The beliefs revealed by their previous choices imply a forecast of a 76% claim rate.

[5]In particular, inattention cannot account for the persistent pattern of decreasing accuracy and confidence in beliefs over time.

[6]A version of the video instructions can be found at the following web address (lower quality than the original): http://www.youtube.com/watch?v=ufAsmQav7Qw

consult the sheet at the end of the experiment to confirm that they were paid accurately. The experimenter and two assistants circulated in order to assure that subjects could not write down or otherwise record the facts they were presented with.

Each session consisted of several tasks conducted within subjects in two stages, Learning and Partial Forgetting.

Learning (L): Memorization and Valuation tasks.

Subjects were presented with a sequence of 20 pairs of pictures of generic human faces (5 people) and animals (5 animals). For each "match-up", the "*winner*" was assigned. The subjects were asked to memorize this assignment. Then they were presented with the sequence again and asked to recall and identify the winner in each match-up. Incorrect answers were revealed at this stage, which provided an extra opportunity of memorization. The entire task required around 7 minutes to complete.

At the valuation stage, subjects were presented with a sequence of 30 propositions B asserting the identity of the winner in 15 out of the 20 previously memorized match-ups. Each of these match-ups generated two propositions in the sequence, one true and one false. For each B, the subjects were presented with a ticket t_B, which paid 11 euro if B was the real winner, 0 euro if it wasn't. The subjective valuation of the ticket was elicited via a procedure equivalent to a multiple price list BDM elicitation in which tickets were compared with a sure sum of k euros, for $k = 1, 2, \ldots, 11$.[7] The monetary payoffs were determined by randomly selecting one of the tickets t_A from the valuation task at the end of the study and executing the subjects' choice between t_A and a random number $R \in \{1, \ldots, 11\}$. The elicitation task was incentive compatible, and video instructions were pre-tested and explained to maximize subject comprehension, which was confirmed with control questions (see Online Appendix A.2). To minimize the influence of the memory task on the elicitation, the photos of the corresponding competitors were presented on different sides of the screen with respect to the memory task. In addition, each ticket t_A and its complement were presented in an order that was different from the order of the memory task, with a spacing of at least 11 decisions between them.

[7]A coarse BDM procedure such as this has an advantage over a fine procedure in that it is easier for subjects to understand. The disadvantage is that it has less statistical power with regard to detecting small differences in valuations.

Partial Forgetting (PF): Interference and Valuation Tasks.

To induce forgetting at the second stage, we used an interference task that engaged subjects in a meaningless computation task and learning another set of related propositions (see Online Appendix A.2). Then they completed another valuation task with a different order of match-ups. The same incentive-compatible scheme was used to determine monetary rewards.

2.2 Follow-up Sessions: Complete Forgetting and Complete Ignorance

This second set of (between-subjects) sessions were conducted one year after the first one and consisted of valuation tasks only. The two types of subjects participated and were presented with identical instructions.

Complete Forgetting (CF).

The subjects who participated in the first sessions were invited again without the knowledge that they there were being invited to a similar study. With a one year delay, there was a considerable attrition rate (~ 70 percent), but the 28 subjects who returned were remarkably similar to the subjects who did not return in terms of their behaviors in treatments L and PF (see Online Appendix A.1 for details.) These subjects were given the same video instructions that they had been presented with one year before and were asked to complete the valuation task for the match-ups of the previous year. The same incentive-compatible scheme was used to determine monetary rewards.

Complete Ignorance (CI).

Subjects who had not participated before were separately recruited (33 subjects). The experiment was conducted exactly as in treatment CF. Subjects received identical video instructions for the valuation task, but were informed that they would have no opportunity to first learn the match-up winners.

3 Results

For each experimental treatment $s \in \{L, PF, CF, CI\}$, let $\mathcal{D}(s)$ be the set of all pairs (A, i) such that A is a true proposition, and agent i evaluates tickets t_A and $t_{\neg A}$ in stage s. Her monetary evaluations are written as

$$V(A, i, s) \quad \text{and} \quad V(\neg A, i, s)$$

respectively. Note that the array $\mathcal{D}(s)$ is the same in L and PF, but varies in CF and CI because of the distinct pools of subjects in these sessions.

The evaluations $V(A, i, s)$ and $V(\neg A, i, s)$ determine *revealed beliefs* for any pair of alternatives A and $\neg A$. If $V(A, i, s) > V(\neg A, i, s)$, then the true proposition A is viewed as more likely than the false alternative $\neg A$ by agent i in stage s. This belief is called *positive*. Conversely, if $V(A, i, s) < V(\neg A, i, s)$, then $\neg A$ is perceived as more likely than A by agent i in stage s. This belief is called *negative*. Finally, the equality $V(A, i, s) = V(\neg A, i, s)$ captures an *indeterminate* belief where A and $\neg A$ are viewed as equally likely by agent i in stage s.

Accordingly, the entire data set $\mathcal{D}(s)$ is partitioned into three disjoint parts

$$\begin{aligned}
\mathcal{P}(s) &= \{(A, i) \in \mathcal{D}(s) : V(A, i, s) > V(\neg A, i, s)\} \\
\mathcal{N}(s) &= \{(A, i) \in \mathcal{D}(s) : V(A, i, s) < V(\neg A, i, s)\} \\
\mathcal{I}(s) &= \{(A, i) \in \mathcal{D}(s) : V(A, i, s) = V(\neg A, i, s)\}
\end{aligned}$$

based on the beliefs revealed in pairs $(A, i) \in \mathcal{D}(s)$.

We observe several patterns that can be explained by forgetting.

3.1 Memory Decay

Take any pair $(A, i) \in \mathcal{D}(s)$ for $s \in \{L, PF, CF\}$. During the memorization task, agent i learns that A is true. Given this information, the tickets t_A and $t_{\neg A}$ should have values $Z = 11$ and zero euros respectively. However, forgetting can make agent i express $V(A, i, s) < Z$ and $V(\neg A, i, s) > 0$.

Table 1 specifies the average willingness to bet on true propositions,

$$V_T(s) := \frac{\sum_{(A,i)\in\mathcal{D}(s)} V(A, i, s)}{|\mathcal{D}(s)|},$$

and the average willingness to bet on false propositions,

$$V_F(s) := \frac{\sum_{(A,i)\in\mathcal{D}(s)} V(\neg A, i, s)}{|\mathcal{D}(s)|}$$

across subjects in each of the four treatments s.

Table 1: Average willingness to bet, standard errors in parentheses.

Willingness to bet	s =L	PF	CF	CI
Bets on true propositions, $V_T(s)$	7.55	5.88	4.63	5.41
	(0.14)	(0.16)	(0.35)	(0.27)
Bets on false propositions, $V_F(s)$	3.56	4.17	4.33	5.16
	(0.13)	(0.15)	(0.28)	(0.25)
Number of Observations	1464	1464	420	495
Number of Subjects	98	98	28	33

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

Table 2: Proportions of revealed beliefs, standard errors in parentheses

Metric	s =L	PF	CF	CI
Positive beliefs, $\frac{\lvert\mathcal{P}(s)\rvert}{\lvert\mathcal{D}(s)\rvert}$	.67	.52	.37	.38
	(0.02)	(0.02)	(0.03)	(0.03)
Negative beliefs, $\frac{\lvert\mathcal{N}(s)\rvert}{\lvert\mathcal{D}(s)\rvert}$	.22	.31	.33	.36
	(0.01)	(0.01)	(0.03)	(.03)
Indeterminate beliefs, $\frac{\lvert\mathcal{I}(s)\rvert}{\lvert\mathcal{D}(s)\rvert}$	.11	.17	.30	.26
	(0.01)	(0.02)	(0.05)	(0.06)
Number of Observations	1464	1464	420	495
Number of Subjects	98	98	28	33

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

Table 2 summarizes the proportions of positive, negative, and indeterminate beliefs in each stage s. We find the comparisons

$$\begin{aligned} &V_T(L) > V_T(PF) > V_T(CF) \quad \text{and} \quad V_F(L) < V_F(PF) \\ &\frac{|\mathcal{P}(L)|}{|\mathcal{D}(L)|} > \frac{|\mathcal{P}(PF)|}{|\mathcal{D}(PF)|} > \frac{|\mathcal{P}(CF)|}{|\mathcal{D}(CF)|} \qquad \frac{|\mathcal{N}(L)|}{|\mathcal{D}(L)|} < \frac{|\mathcal{N}(PF)|}{|\mathcal{D}(PF)|} \end{aligned} \tag{1}$$

to be statistically significant ($p < .01$, Wilcoxon Sign-Rank test[8]).

[8]The comparison between L and PF is within-subject, whereas comparing CF with L or PF is within-subject with attrition, and comparing CF with other treatments is between-subjects. While each subject has repeated measures, for statistical tests, unless otherwise stated, each subject contributes one observation consisting of his/her average/proportion across proposition pairs $(A, \neg A)$. We report p-values for the Wilcoxon Sign-Rank test unless otherwise stated.

These inequalities are analogous to the standard *memory decay* pattern where the accuracy of memory diminishes over time (see Rubin and Wenzel [34]) for a review of psychological evidence). In our data, this pattern holds both for average monetary metrics and revealed belief proportions.

Note that the inequalities $V_F(PF) < V_F(CF)$ and $\frac{|\mathcal{N}(PF)|}{|\mathcal{D}(PF)|} < \frac{|\mathcal{N}(CF)|}{|\mathcal{D}(CF)|}$ hold as well, but are not statistically significant.

The memory decay patterns are further illustrated at the individual level by two histograms in Figures 1 and 2 that report the distribution of the within-subject average willingness to bet on true propositions $V(A, i, s)$ and the average proportion of positive beliefs.

3.2 Forgetting and Confidence

In addition to revealed beliefs, monetary evaluations allow us to measure subjective confidence in these beliefs.

For each subject i and session s, consider sets

$$\mathcal{P}(i, s) := \{A : (A, i) \in \mathcal{P}(s)\}$$
$$\mathcal{N}(i, s) := \{A : (A, i) \in \mathcal{N}(s)\}$$

that consist of propositions A for which agent i in session s reveals positive and negative beliefs respectively. If $A \in \mathcal{P}(i, s)$ or $A \in \mathcal{N}(i, s)$, then the proposition $B = A$ or $B = \neg A$ respectively is viewed as *more likely* than the (less likely) alternative $\neg B$. Of course, this comparison is purely subjective because A is objectively true and $\neg A$ is objectively false.

For any subject i, compute her average willingness to bet on more likely propositions and less likely alternatives:

$$V_{ML}(i, s) = \frac{\sum_{A \in \mathcal{P}(i,s)} V(A, i, s) + \sum_{A \in \mathcal{N}(i,s)} V(\neg A, i, s)}{|\mathcal{P}(i, s)| + |\mathcal{N}(i, s)|}$$
$$V_{LL}(i, s) = \frac{\sum_{A \in \mathcal{P}(i,s)} V(\neg A, i, s) + \sum_{A \in \mathcal{N}(i,s)} V(A, i, s)}{|\mathcal{P}(i, s)| + |\mathcal{N}(i, s)|}.$$

This definition requires $\mathcal{P}(i, s) \cup \mathcal{N}(i, s) \neq \emptyset$, which holds for all i, except for two subjects in CF, and four subjects in CI.

For each session s, Table 3 reports the average metrics

$$V_{ML}(s) := \frac{\sum_i V_{ML}(i, s)}{|\{i : \mathcal{P}(i, s) \cup \mathcal{N}(i, s) \neq \emptyset\}|}$$
$$V_{LL}(s) := \frac{\sum_i V_{LL}(i, s)}{|\{i : \mathcal{P}(i, s) \cup \mathcal{N}(i, s) \neq \emptyset\}|}$$

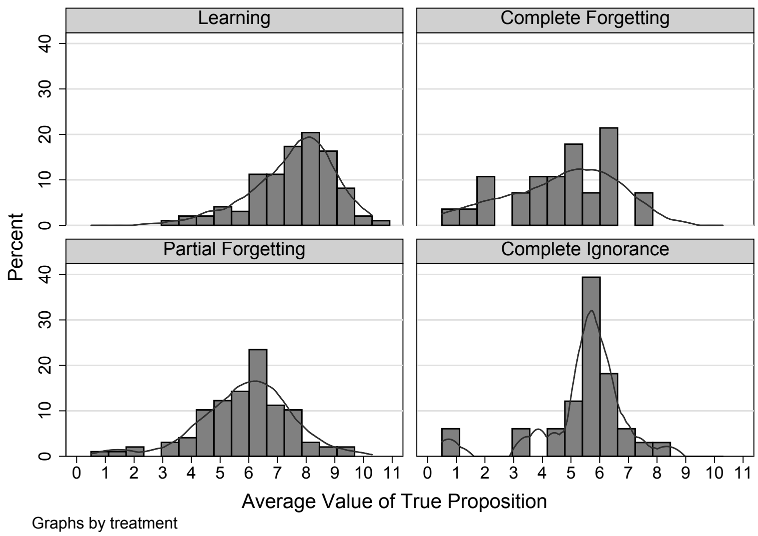

Figure 1: Histogram for the average value of true proposition, within subjects, by treatment (with kernel density)

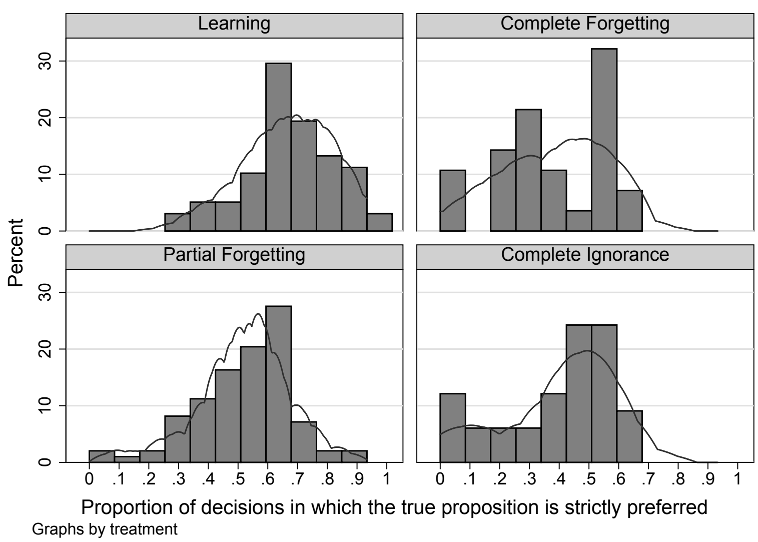

Figure 2: Histogram for the average proportion of positive beliefs, within subjects, by treatment (with kernel density)

Table 3: Average willingness to bet on more likely (ML) and less likely (LL) propositions, standard errors in parentheses

Metric	$s=$L	PF	CF	CI
Average subjective value $V_{ML}(s)$	9.27	8.44	6.98	8.11
	(0.13)	(0.16)	(0.42)	(0.36)
Average subjective value $V_{LL}(s)$	2.13	2.14	2.68	2.95
	(0.11)	(0.11)	(0.32)	(0.36)
Number of Observations	1464	1464	390	435
Number of Subjects	98	98	26	29

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

across all subjects i such that $\mathcal{P}(i,s) \cup \mathcal{N}(i,s) \neq \emptyset$. We find that the inequalities

$$V_{ML}(L) > V_{ML}(PF) > V_{ML}(CF) \tag{2}$$

are statistically significant ($p < 0.01$), but $V_{LL}(L) < V_{LL}(PF) < V_{LL}(CF)$ are not.

Inequality (2) shows that forgetful subjects become less confident in their revealed beliefs over time. Thus they show a partial awareness of their memory decay. Note that the statistic V_{LL} is essentially unchanged between L and PF and increases in CF at a statistically insignificant level. Thus our subjects can be still reluctant to bet against their beliefs even when they lose confidence in these beliefs. This reluctance can result from ambiguity aversion that we discuss in Section 3.4 below.

3.3 Overconfidence

Even though subjects exhibit some awareness about the decay of their memory, their calibration is not perfect and exhibits some *overconfidence*.

To show this, compute the objective expected values for bets on subjectively more likely and less likely propositions. For each i and s, let

$$R_{ML}(i,s) = \frac{Z \cdot |\mathcal{P}(i,s)|}{|\mathcal{P}(i,s)| + |\mathcal{N}(i,s)|}$$

$$R_{LL}(i,s) = \frac{Z \cdot |\mathcal{N}(i,s)|}{|\mathcal{P}(i,s)| + |\mathcal{N}(i,s)|} = Z - R_{ML}(i,s).$$

Table 4: Average difference between subjective evaluation and objective return for bets on more likely (ML) and less likely (LL) propositions, with standard errors in parentheses

Metric	s =L	PF	CF	CI
The discrepancy $V_{ML}(s) - R_{ML}(s)$ for more likely propositions	1.03 (0.17)	1.44 (0.24)	1.22 (0.44)	2.48 (0.38)
The discrepancy $V_{LL}(s) - R_{LL}(s)$ for less likely propositions	-.63 (0.14)	-1.85 (0.17)	-2.55 (0.43)	-2.41 (0.36)
Number of Observations	1464	1464	390	435
Number of Subjects	98	98	26	29

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

The corresponding averages across all subjects i in session s are

$$R_{ML}(s) := \frac{\sum_i R_{ML}(i,s)}{|\{i : \mathcal{P}(i,s) \cup \mathcal{N}(i,s) \neq \emptyset\}|}$$
$$R_{LL}(s) := \frac{\sum_i R_{LL}(i,s)}{|\{i : \mathcal{P}(i,s) \cup \mathcal{N}(i,s) \neq \emptyset\}|}.$$

Table 4 reports the discrepancies $V_{ML}(s) - R_{ML}(s)$ and $V_{LL}(s) - R_{LL}(s)$. We find that the inequalities

$$V_{ML}(s) > R_{ML}(s) \quad \text{and} \quad V_{LL}(s) < R_{ML}(s) \tag{3}$$

hold in all s in a statistically significant way.

This pattern is a form of *overconfidence*: subjects who believe that B is more likely than $\neg B$ overestimate the average accuracy of this belief and hence, overestimate $R_{ML}(s)$. Indeed, people commonly overestimate their own abilities and performance (see Keren [23], Moore and Healy [28] for an overview). The same pattern can apply to evaluation of memory as well. In one study with monetary incentives, Ericson [13] observes that half of subjects forget to send an email to claim a twenty-dollar payment six months after the experiment. The same subjects initially reveal a belief that they will claim the prize with probability of around 70%. Unlike Ericson, we find biases in *current* rather than anticipated future memory. Moreover, we deliberately use propositions that do not mention subjects' own performance, ability, or actions. In this way, we alleviate concerns about self-evaluative emotions. Of course, it is still possible that some subjects can reinterpret

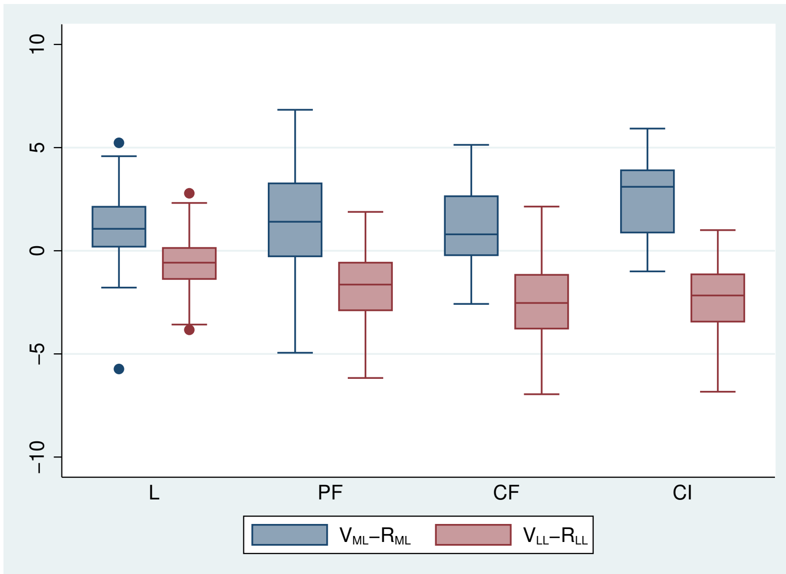

Figure 3: Overconfidence in treatments $s = L, PF, CF, CI$

tickets t_B as bets that their beliefs are correct or not rather than proposition B is true or false.

Note that the inequalities $V_{ML}(s) > R_{ML}(s)$ and $V_{LL}(s) < R_{LL}(s)$ cannot be explained by the standard expected utility model with well-calibrated beliefs. Instead, the overconfidence pattern in our data can emerge if risk-neutral subjects mix their memory signals with objectively irrelevant hunches. We develop such a two-signal model in Section 4 below.

3.4 Complete Forgetting and Comparative Ignorance

In our data, overconfidence coexists with ambiguity aversion.

Note that the comparisons

$$V_T(s) > V_F(s) \quad \text{and} \quad \frac{|\mathcal{P}(s)|}{|\mathcal{D}(s)|} > \frac{|\mathcal{N}(s)|}{|\mathcal{D}(s)|}$$

are significant ($p < .01$) in sessions L and PF, but become insignificant in $s = CF$. Therefore, the complete forgetting treatment is effective in erasing previous memorization to a statistically irrelevant signal. Moreover, we observe no significant differences in the proportions of revealed beliefs (positive, negative, or indeterminate) between the two sessions CF and CI.

However, there is a significant ($p < .05$) difference between CF and CI in the average willingness to bet across all propositions

$$\hat{V}(s) = \frac{\sum_{(A,i)\in\mathcal{D}(s)} V(A,i,s) + V(\neg A,i,s)}{2\,|\mathcal{D}(s)|} = \frac{V_T(s) + V_F(s)}{2}.$$

We find $\hat{V}(CF) = 4.48$ to be lower than $\hat{V}(CI) = 5.28$. In other words, subjects behave as if they are more ambiguity averse when they have forgotten than when they never knew.

We explain this pattern in terms of *comparative ignorance*—a broad term that Fox and Tversky [16] use to describe ambiguity aversion driven by a "feeling of incompetence" and a contrast with more knowledgeable decision makers. Indeed, a subject's ambiguity aversion in session CF can be increased by the contrast between her total memory failure with the more informed cognitive states she had in the past sessions L and CF. This argument suggests that people can feel less competent when they *forget* whether B or $\neg B$ is true rather than when they have always been ignorant of this identity.

3.5 Indeterminate Beliefs and Comparative Ignorance

Indeterminate beliefs between A and $\neg A$ in pairs $(A,i) \in \mathcal{I}(s)$ can result from the use of the *principle of insufficient reason* where two propositions A and $\neg A$ are taken as equiprobable if there is no evidence in favor of A or $\neg A$. Thus the proportion $\frac{|\mathcal{I}(s)|}{|\mathcal{D}(s)|}$ can roughly approximate the prevalence of this principle in each treatment s.

Indeed, we observe indeterminate beliefs are significantly ($p < 0.01$) more common in the CF treatment than in L and PF sessions:

$$\frac{|\mathcal{I}(CF)|}{|\mathcal{D}(CF)|} > \frac{|\mathcal{I}(PF)|}{|\mathcal{D}(PF)|} \quad \text{and} \quad \frac{|\mathcal{I}(CF)|}{|\mathcal{D}(CF)|} > \frac{|\mathcal{I}(L)|}{|\mathcal{D}(L)|}.$$

This finding is in line with Voorhoeve, Binmore, and Stewart [38] where the principle of insufficient reason is observed to be more prevalent for less familiar events.

Moreover, we hypothesize that ambiguity aversion on the domain $\mathcal{I}(s)$ should be stronger in the forgetting treatments $s = L, PF, CF$ than in $s = CI$. To check this hypothesis, consider an individual metric

$$V_{PIR}(i,s) = \frac{\sum_{A:(A,i)\in\mathcal{I}(s)} V(A,i,s)}{|\{A : (A,i) \in \mathcal{I}(s)\}|}$$

Table 5: Average metrics with simulated values in square brackets

Metric	s =L	PF	CF	CI
$V_T(s)$	7.55[7.27]	5.88[5.83]	4.63[4.38]	5.41[5.35]
$V_F(s)$	3.56[3.49]	4.17[3.94]	4.33[4.38]	5.16[5.35]
$\frac{\lvert\mathcal{P}(s)\rvert}{\lvert\mathcal{D}(s)\rvert} * 100\%$	66.7 [67.5]	52 [51.8]	36.7 [36]	38.2 [36]
$\frac{\lvert\mathcal{N}(s)\rvert}{\lvert\mathcal{D}(s)\rvert} * 100\%$	22.1 [21.5]	31.1 [28.8]	33.5 [36]	36.2 [36]
$\frac{\lvert\mathcal{I}(s)\rvert}{\lvert\mathcal{D}(s)\rvert} * 100\%$	11.2 [10.9]	16.9 [19.4]	29.8 [28]	25.6 [28]
$V_{ML}(s)$	9.27 [9.2]	8.44 [8.28]	6.98 [7.15]	8.11 [8.8]
$R_{ML}(s)$	8.23 [8.34]	7.00 [7.07]	5.75 [5.5]	5.63 [5.5]
$V_{LL}(s)$	2.13 [1.67]	2.14 [1.91]	2.68 [2.2]	2.95 [2.2]
$R_{LL}(s)$	2.76 [2.66]	4.00 [3.93]	5.24 [5.5]	5.37 [5.5]
$V_{PIR}(s)$	5.27 [5.01]	4.17 [4.42]	4.02 [3.67]	5.42 [4.95]

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

and compute the average $V_{PIR}(s)$ across all $V_{PIR}(i, s)$ such that the set $\{A : (A, i) \in \mathcal{I}(s)\}$ is not empty. Then $V_{PIR}(CI) = 5.42$ is significantly higher than $V_{PIR}(CF) = 4.02$ or $V_{PIR}(PF) = 4.17$. The inequality $V_{PIR}(CI) > V_{PIR}(L)$ is not significant.[9]

Thus the principle of insufficient reason motivates more ambiguity aversion in subjects when they forget whether B or $\neg B$ is true rather than when they never knew the truth at all.

4 Two-Signal Model of Forgetting

In this section we present a two-signal model of memory that accommodates all of the above patterns.

Consider an agent who evaluates each pair of bets t_A and $t_{\neg A}$ based on two signals.

- A memory signal $M \in \{A, \neg A\}$ is received with probability μ_s that depends on the treatment $s \in \{L, PF, CF\}$. The signal M points

[9]Indeterminate beliefs in stage L are rare and harder to explain exclusively by complete forgetting.

to the true proposition A with probability $\alpha > 0.5$ and to the false negation $\neg A$ with probability $1-\alpha$.

- A hunch (heuristic, intuition) signal $H \in \{A, \neg A\}$ is received with probability ξ. The signal H is independent of memory and points to A and $\neg A$ with equal probabilities $\frac{1}{2}$. Even though this signal is uninformative, the agent believes that H is true with probability $\beta \in [\frac{1}{2}, 1]$ where β is randomly drawn from some distribution on $[\frac{1}{2}, 1]$ for each pair $\{A, \neg A\}$.

In stages $s \in \{L, PF, CF\}$, the agent processes these signals into monetary evaluations $V(A, s)$ and $V(\neg A, s)$ as follows.

(i) If both signals M and H are received, then she obtains a subjective probability π_M that M is true via the Bayesian formula:

$$\pi_M = \begin{cases} \frac{\alpha\beta}{\alpha\beta+(1-\alpha)(1-\beta)} & \text{if } M = H \\ \frac{\alpha(1-\beta)}{\alpha(1-\beta)+\beta(1-\alpha)} & \text{if } M \neq H, \end{cases}$$

and computes $V(M, s) = \pi_M Z$ and $V(\neg M, s) = (1 - \pi_M)Z$ via the expected value criterion.

(ii) If only M arrives, then $V(M, s) = \alpha Z$ and $V(\neg M, s) = (1 - \alpha)Z$ comply with expected value again.

(iii) If only H arrives, then the agent becomes ambiguity averse because she can perceive that her memory fails. Let $V(H, s) = \frac{\beta+0.5}{2} Z$ and $V(\neg H, s) = (1-\beta)Z$. These evaluations are consistent with the multiple priors model on the binary state space $\{H, \neg H\}$ and linear utility for money. The set of priors is taken to be the interval $\left[\frac{\beta+0.5}{2}, \beta\right]$.

(iv) If no signals arrive, then the agent relies on the principle of insufficient reason. Her ambiguity aversion applies to both t_A and $t_{\neg A}$ in this case. Let $V(A, s) = V(\neg A, s) = 0.33Z$.

(v) In stage CI, the agent is ambiguity and risk neutral if she receives the hunch signal: $V(H, CI) = \beta Z$ and $V(\neg H, CI) = (1 - \beta)Z$. If she receives no signal, let $V(A, CI) = V(\neg A, CI) = 0.45Z$.

To illustrate, assume that

- $\alpha = 0.85$, $\mu_L = 0.9$, $\mu_{PF} = 0.45$, $\mu_{CF} = 0$, and $\xi = 72\%$,

- β equals 0.65, 0.75, 0.85, 0.95 with probabilities 25% each.

Table 5 shows that this choice of parameters generates all of the above patterns: memory decay, diminishing confidence, overconfidence, and comparative ignorance.

We can further use the two-signal model to generate some other testable predictions about subjective beliefs and confidence.

4.1 Evolution of Positive and Negative Beliefs

In the two-signal model with $\alpha > \frac{1}{2}$, the proportion of beliefs that conform to the memory signal should be higher among positive beliefs than among negative beliefs. For example, for the above parameters, 48% of all negative beliefs and only 8% of positive beliefs are expressed without a memory signal. Accordingly, positive beliefs should be more likely to be preserved together with the memory signal, and negative beliefs are more likely to vary together with the random hunch.

We have enough data to track the evolution of subjective beliefs between treatments L and PF. We can also simulate this evolution with the two-signal model. Assume that if the memory signal arrives both in L and PF, then it has the same value over $\{A, \neg A\}$, but the hunch signal and its intensity β get a new random drawing in each stage.

Table 6 summarizes the proportions of all nine combinations of revealed beliefs in the two time periods, both in real and simulated data (in brackets). Each cell of this table reports the percentage of observations in the overlap of the row and column sets. The last column and row of Table 6 report the aggregate percentages of positive, negative, and indeterminate beliefs in stages L and PF.

Note that negative beliefs expressed in stage L are highly transient both in real and simulated data. Conditional on $\mathcal{N}(L)$, the rates of positive and negative beliefs in stage PF are close: $\frac{|\mathcal{N}(L)\cap\mathcal{P}(PF)|}{|\mathcal{N}(L)|} = 45\%[44\%]$ and $\frac{|\mathcal{N}(L)\cap\mathcal{P}(PF)|}{|\mathcal{N}(L)|} = 38\%[37\%]$ respectively. In other words, a negative belief in stage L is almost as likely to be replaced by a positive belief in PF as to be preserved. By contrast, positive beliefs are more sticky. Conditional on $\mathcal{P}(L)$, the frequencies of positive and negative beliefs are 59%[57%] and 26%[24%] respectively. Thus positive beliefs in stage L are mostly preserved in PF.

Table 6: Evolution of Revealed Beliefs in L and PF, simulated values in square brackets

	$\mathcal{P}(PF)$	$\mathcal{N}(PF)$	$\mathcal{I}(PF)$	$\mathcal{D}(PF)$
$\mathcal{P}(L)$	39.4 [38.6]	17.4 [16.1]	9.9 [12.8]	66.7 [67.5]
$\mathcal{N}(L)$	8.4 [7.9]	9.9 [9.4]	3.8 [4.3]	22.1 [21.6]
$\mathcal{I}(L)$	4.2 [5.3]	3.8 [3.3]	3.2 [2.3]	11.2 [10.9]
$\mathcal{D}(L)$	52 [51.8]	31.1 [28.8]	16.9 [19.4]	100

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

4.2 Confidence in Positive and Negative Beliefs

The two-signal model also suggests that negative beliefs should be more often affected by ambiguity aversion because they are more likely to be expressed without a memory signal. Moreover, it is plausible that the memory accuracy α should be higher than the average β, which is the subjective weight attached to the random hunch. Thus the average confidence should be weaker for negative beliefs than for positive beliefs as long as memory retains some statistical accuracy.

To check this hypothesis, we compute the average value V_{PT} placed on true propositions that are correctly perceived as more likely (positive), and the average value V_{NF} of false propositions that are incorrectly perceived as more likely (negative):

$$V_{PT}(i,s) = \frac{\sum_{A\in\mathcal{P}(i,s)} V(A,i,s)}{|\mathcal{P}(i,s)|}$$
$$V_{NF}(i,s) = \frac{\sum_{A\in\mathcal{N}(i,s)} V(\neg A,i,s)}{|\mathcal{N}(i,s)|}.$$

Table 7: Average confidence metrics, with simulated values in square brackets, and standard errors in parentheses

Metric	$s =$L	PF	CF	CI
$V_{PT}(s)$	9.57 [9.34]	8.62 [8.58]	7.02 [7.15]	8.29 [8.8]
	(0.13)	(0.17)	(0.43)	(0.36)
$V_{NF}(s)$	8.34 [8.77]	7.95 [7.75]	6.91 [7.15]	8.01 [8.8]
	(0.20)	(0.21)	(0.43)	(0.38)

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

The corresponding averages across all subjects i in session s are

$$V_{PT}(s) := \frac{\sum_i V_{PT}(i,s)}{|\{i : \mathcal{P}(i,s) \neq \emptyset\}|}$$
$$V_{NF}(s) := \frac{\sum_i V_{NF}(i,s)}{|\{i : \mathcal{N}(i,s) \neq \emptyset\}|}.$$

Table 7 shows that the inequalities $V_{PT}(s) > V_{NF}(s)$ are statistically significant in $s = L$ and $s = PF$, but are insignificant in $s = CF$ and $s = CI$. Our simulated data satisfies the same inequalities.

5 Conclusion

Our data identifies several empirical patterns in betting preferences when memory is imperfect. Most importantly, we find that

- both the accuracy and subjective confidence in *revealed beliefs* decay over time,
- subjects are more ambiguity averse under complete forgetting than under complete ignorance,
- subjects are overconfident in their beliefs.

We propose a stylized two-signal model of forgetting that accommodates all of these patterns by assuming that subjects combine valid memory signals with statistically irrelevant hunches. We use the model to find more testable patterns in our data.

References

[1] S. Agarwal, J. C. Driscoll, X. Gabaix, and D. Laibson. Learning in the credit card market. Mimeo, SSRN: http://ssrn.com/abstract=109162, 2013.

[2] S. Baliga and J. Ely. Mnemonomics: The sunk cost fallacy as a memory kludge. *American Economic Journal: Microeconomics*, 3:35–67, 2011.

[3] B. Barber, T. Odean, and L. Zheng. Out of sight, out of mind: The effects of expenses on mutual fund flows. *The Journal of Business*, 78:2095–2120, 2005.

[4] P. Battigalli. Dynamic consistency and imperfect recall. *Games and Economic Behavior*, 20(1):31–50, 1997.

[5] P. Blavatskyy. Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty*, 38:39–49, 2009.

[6] N. Brewer and A. Burke. Effects of testimonial inconsistencies. *Law and Human Behavior*, 30:31–50, 2002.

[7] G. Brown, I. Neath, and N. Chater. A temporal ratio model of memory. *Psychological Review*, 114(3):539–576, 2007.

[8] C. Camerer and D. Lovallo. Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1):306–318, 1999.

[9] R. Chetty, A. Looney, and K. Kroft. Salience and taxation: Theory and evidence. *American Economic Review*, 99:1145–1177, 2009.

[10] F. Costello and P. Watts. Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, 121:463–480, 2014.

[11] B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institute Henri Poincare*, 7:1–68, 1937. Translated and reprinted in Kyburg and Smokler, 1964.

[12] I. Erev, T. Wallsten, and D. Budescu. Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, 101:519–527, 1994.

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

[14] U. Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

[15] B. Fischhoff, P. Slovic, and S. Lichtenstein. Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4):552–564, 1977.

[16] C. Fox and A. Tversky. Ambiguity aversion and comparative ignorance. *Quaterly Journal of Economics*, 110:585–603, 1995.

[17] C. Froehle and D. White. Interruption and forgetting in knowledge-intensive service environments. *Production and Operations Management*, 23(4):704–722, 2014.

[18] B. Garrett. *Convicting the innocent*. Harvard University Press, Cambridge, MA, 2011.

[19] C. Heath and A. Tversky. Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, 4(1):5–28, 1991.

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

[21] E. Hoelzl and A. Rustichini. Overconfident: Do you put your money on it? *The Economic Journal*, 115(503):305–318, 2005.

[22] W. J. and G. Wells. The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, 18:10–65, 2017.

[23] G. Keren. Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, 77:217–273, 1991.

[24] E. Kirchler and B. Maciejovsky. Simultaneous over- and underconfidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty*, 25:65–85, 2002.

[25] S. Lichtenstein and B. Fischhoff. Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, 20(2):159–183, 1977.

[26] R. Lindsay, G. Wells, and C. Rumpel. Can people detect eyewitnesses identification accuracy within and across situations? *Journal of Applied Psychology*, 66:79–89, 1981.

[27] S. McKenna and I. Glendon. Occupational first aid training: Decay in cardiopulmonary resuscitation (cpr) skills. *Journal of Occupational and Organizational Psychology*, 58(2):109–117, 1985.

[28] D. Moore and P. Healy. The trouble with overconfidence. *Psychological Review*, 115(2):502–517, 2008.

[29] S. Mullainathan. A memory-based model of bounded rationality. *Quarterly Journal of Economics*, 117(3):735–774, 2002.

[30] M. Piccione and A. Rubinstein. The absent-minded drivers paradox: synthesis and responses. *Games and Economic Behavior*, 20(1):121–130, 1997.

[31] M. Piccione and A. Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20(1):3–24, 1997.

[32] H. Roediger, J. Wixted, and K. A. Desoto. The curious complexity between confidence and accuracy in reports from memory. In L. Nadel and W. Sinnott-Armstrong, editors, *Memory and Law*. Oxford University Press, New York, 2012.

[33] C. Rowan. Assessing memory decay rate: What factors are the best predictors of decrements in training proficiency in a threat vehicle identificaiton task. In *Proceedings of the Human Factors and Ergonomics Society, 59th Annual Meeting*, 2015.

[34] D. Rubin and A. Wenzel. One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103(4):734–760, 1996.

[35] L. J. Savage. *The Foundations of Statistics*. Dover Publications Inc., New York, second revised edition, 1972.

[36] J. S. Shaw and K. A. McClure. Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, 20:629–653, 1996.

[37] S. Trautmann, F. Vieider, and P. Wakker. Causes of ambiguity aversion: Known versus unknown preferences. *Journal of Risk and Uncertainty*, 36:225–243, 2008.

[38] A. Voorhoeve, K. Binmore, and L. Stewart. How much ambiguity aversion? Finding indifferences between Ellsberg's risky and ambiguous bets. *Journal of Risk and Uncertainty*, 45(3):215–238, 2012.

[39] A. Wilson. Bounded memory and biases in information processing. *Econometrica*, 82(6):2257–2294, 2003.

A Online Appendix for "Subjective Beliefs And Confidence When Facts Are Forgotten" by Kopylov and Miller

A.1 Supplementary Graphs

The 28 subjects who returned, participating in the Complete Forgetting did not differ substantially from the 70 subjects who did not return. In Table 8, for the first set of sessions, we present summary statistics for subjects who didn't return and subjects that returned . As can be seen the values are comparable. No difference is significant.

Table 8: The 28 subjects that returned did not differ substantially from the 70 subjects that did not return

	didn't return	returned
Learning Treatment:		
"prefer" true proposition	0.68	0.63
"prefer" false proposition	0.21	0.24
indifference to truth	0.10	0.13
value of true proposition	7.55	7.54
value of false proposition	3.42	3.88
total value	10.99	11.43
Partial Forgetting Treatment:		
"prefer" true proposition	0.52	0.50
"prefer" false proposition	0.31	0.32
indifference to truth	0.17	0.17
value of true proposition	5.92	5.79
value of false proposition	4.17	4.18
total value	10.08	9.97

Notation: L (Learning), PF (Partial Forgetting), CF (Complete Forgetting), CI (Complete Ignorance).

A.2 Experimental Sessions & Procedures

All experiments discussed herein were conducted at the Bocconi Experimental Laboratory for the Social Sciences (BELSS) at Bocconi University using z-Tree experimental software. Participants were Bocconi students who had previously registered to the subject pool via the commercial online recruiting platform Sona-Systems. Email invitations were sent to a random subset of the subject pool who self-reported to have an advanced level of English in listening comprehension, reading comprehension, writing and speaking. The study information presented in the invitation was minimal; importantly, as the study involved ambiguity, expected payments were not communicated.[10] Each experimental session lasted roughly 90-minutes, including payment. We ran two sets of experimental sessions. We first describe the initial "main" sessions. Further below we briefly describe the the follow-up sessions, which differ minimally and were conducted one year later.

Overview of Main Sessions Upon arrival to the lab, students were randomly seated at individual computer carrel booths that isolated their screens from the view of other subjects. Subjects were instructed to put on headphones and watch the 17-minute carefully narrated video instructions, presented with closed captions (see Footnote 6).[11] After the instructional video was complete, subjects participated in a comprehension check task in which they answered at least six questions to make sure they understood the connection between their decisions, the random selection, and their payoffs. After each question from the comprehension check task, the individual performance of each subject was broadcast to the experimenter computer. Subjects who had difficulty were personally attended to (typically $5-10\%$ of subjects with the video).[12] Once subjects completed at least six questions and no subject made a mistake for at least two questions in a row, we proceeded to the experiment.

[10]The title of the study was "Memory and Incentives (in English)". The description of the study was "There is a brief memorization task and then you make choices. The better your memory performance, the more you will make, on average. The maximum you can make is 27 euros, and many participants will earn this much. Please bring your codice fiscale and ID in order to be paid." (The final set of sessions involved 18 euros.) The minimum payment (7 euros) and the (expected) duration (75 minutes) were reported in separate fields.

[11]The first 2 sessions were read aloud by the experimenter. This mode of delivery was not as effective in making sure each subject understood the task because it did not control what subjects attended to. In these sessions, each subject understood the task eventually, as they were personally explained.

[12]Without the video, in the first two sessions, 40% of subjects needed assistance.

The experiment involved memorizing the winner in a match-up between two competitors A and B in a "game", in which each competitor was represented by picture.[13] There were 10 competitors total so there were $\binom{10}{2} = 45$ possible match-ups and thus 45 outcomes to memorize. Five of the competitors were represented by pictures of animals and five by pictures of a person's face.[14] Most competitors won between 44% and 55% of the time.[15] The experiment involved two blocks of four experimental tasks, with an interference task performed in between. After the experiment was over, subjects completed a questionnaire and then were paid.

The experiment In the first experimental task of a block, the *passive learning* task, subjects learned the winners of 20 of the 45 match-ups.[16] In each brief period of the task, the pictures of each competitor in the match-up were presented along side each other on the screen, after two seconds a green box appeared around the winner of the game, and two seconds later the pictures disappeared and the next match-up was presented. The task completed in 6×20 seconds. A video of the task can be found within the complete video instructions in Footnote 6.

The second experimental task, the *active learning* task, involved subjects actively learning the winners of the same 20 match-ups, presented in the same order with the competitors on the same side of the screen as in the *passive learning* task. The pictures of both competitors in the match-up were presented along side each other on the screen, subjects clicked to the picture of the competitor they believed was the winner, and they received feedback if they were correct (green box) or incorrect (red "X"). If they did not respond within 3 seconds, a green box appeared around the winner of the game, and two seconds later the pictures disappeared and the next match-up was presented. The task completed in 6×20 seconds. A video of the task can be found within the complete instructions in Footnote 6 of the

[13] The game had no meaning, see video instructions for details (link in Footnote 6).

[14] The pictures of faces were obtained from `http://faceresearch.org` while the pictures of animals were obtained from publically availability photos under the creative commons licence.

[15] The outcomes of matches were not chosen randomly. Care was taken to make sure there was no association between the identity of the winner and the species, size of animal, or gender. In order for the task to not appear to be too difficult at first, one of the competitors was selected to win 77% of the time, and another was selected to win 66% of the time. Note, while each competitor has either a $p = 1$ or $p = 0$ probability of winning, the degree of uncertainty is not so high because each competitor wins at least 44% of the time.

[16] In the first two sessions they learned the outcomes of 30 matchups.

paper.

The third experimental task, the *memory check* task, subjects viewed the same 20 match-ups with the photos on the same side of the screen, but were presented with the match-ups in random order, and were asked to click on who they believed to be the winner and report their confidence as "Low", "Medium", or "High", without any direct financial incentives. The subjects proceeded at their own pace and were allotted 6×20 seconds to complete the task.[17]

The fourth and final experimental task of a block, the *get value* task, involved the subjects indirectly valuing, with incentive compatible choices, a contract which paid them 11 euros if an indicated competitor won in a given match-up, and 0 euros if the indicated competitor lost in that match-up.[18] A subset of 15 of the presented match-ups were presented to them. For each match-up, let C_1 and C_2 be the labels for each competitor. Subjects performed the valuation task twice, once for the contract that paid 11 euros if competitor C_1 was the winner and once for the *mirror contract* which paid 11 euros if competitor C_2 was the winner. Each contract and its mirror were presented with a minimum of 11 decisions between them, with an average of 14 decisions between them.[19] The subjects proceeded at their own pace and were allotted 7 minutes to complete the task.

After the completion of the first block, subjects engaged in an *interference task* that presumably would interfere with their memory and lead to forgetting. Subjects were told that the task was designed to test their numeracy. They were to calculate the change due after a cash payment for a restaurant bill. They were to complete as many as possible in the time allotted 7 minutes, 40 seconds.

The second block of four tasks followed the interference task. The *passive learning* task and the *active learning* task involved the remaining 25 match-ups.[20] The *memory check* task followed presenting the match-ups in random

[17]Concerns about order effects relating to the ensuing *get value* task should not be too strong as it is unlikely subjects remember their exact reported confidence, given the time delay and the shuffling in the order of matchups. The case of total valuation $\hat{U} = \hat{V}_T + \hat{V}_F$, it should be emphasized that $\hat{V}_T$ and $\hat{V}_F$ are elicited as separate decisions.

[18]For each contract the subjects made 11 decisions between the contract and a certain k euro payment for $k = 1, \ldots, 11$. Their value for the contract was the unknown threshold value between where they switched from certain payment to the contract.

[19]This design feature proved important as during focus group studies, subjects reported a strong desire to be consistent between a contract and its mirror, where consistent means that their valuations sum to 11.

[20]After the second session additional match-ups from "Game 2" were memorized with the aim of interfering with the memory of of the match-ups from the first game.

order. Finally the identical *get value* task from the first block was performed, with additional decisions involving contracts for the new match-ups.[21]

When the second block was complete, subjects filled out a questionnaire. Next, two randomly selected decisions were selected to count for payment. Finally, subjects were called individually to be paid.

Overview of Follow-up Sessions The follow-up sessions sessions were conducted one year after the first one. The subjects who participated in the first sessions were invited again without the knowledge that they there were being invited to a similar study (Complete Forgetting Treatment). A new set of subjects were separately recruited (Complete Ignorance Treatment). Subjects were presented with the same video instructions as in the main sessions but were instructed that they would not have an opportunity to memorize the match-ups, and would only participate in the *get value* task. The procedures were otherwise identical to that of the main sessions.

[21]In the sessions with "game 1" and "game 2", we focus on the 15 match-ups that were valued in both blocks.