UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Self-Confidence and Motivated Memory Loss: Evidence from Schools

Vivek Roy-Chowdhury

Abstract

Motivated beliefs theory suggests the absorption of information may be biased, especially when it bears consequences for the ego. This paper finds empirical support for that hypothesis in the field, using longitudinal data on teenagers' memories of mathematics report card grades and administrative data on actual grades. Students: i) make more errors in recalling lower grades; ii) update their academic self-confidence in association with recalled grades rather than actual grades; and iii) have more flattering memories of grades when the survey was administered with a longer delay. The first two results bolster recent research in demonstrating that patterns of motivated recall are robust to within-individual estimation. The last result extends the field literature in showing that a large part of the mechanism for motivated information absorption is memory loss over time. A structural model is used to represent memories as the outcome of a subconscious choice problem, disentangling competing motives to enhance self-confidence and respect reality. The estimated model indicates that the costs of memory distortions decrease as time passes after information transmission, and students with low self-confidence had a greatly diminished preference for inflating self-confidence via memory distortions.

Reference Details

CWPE	2213
Published	21 February 2022
Key Words	motivated beliefs, education, ability, recall, selective memory
JEL Codes	D91, I21, J83
Website	www.econ.cam.ac.uk/cwpe

SELF-CONFIDENCE AND MOTIVATED MEMORY LOSS: EVIDENCE FROM SCHOOLS

Vivek Roy-Chowdhury*

February 2022

Abstract

Motivated beliefs theory suggests the absorption of information may be biased, especially when it bears consequences for the ego. This paper finds empirical support for that hypothesis in the field, using longitudinal data on teenagers' memories of mathematics report card grades and administrative data on actual grades. Students: i) make more errors in recalling lower grades; ii) update their academic self-confidence in association with recalled grades rather than actual grades; and iii) have more flattering memories of grades when the survey was administered with a longer delay. The first two results bolster recent research in demonstrating that patterns of motivated recall are robust to within-individual estimation. The last result extends the field literature in showing that a large part of the mechanism for motivated information absorption is memory loss over time. A structural model is used to represent memories as the outcome of a subconscious choice problem, disentangling competing motives to enhance self-confidence and respect reality. The estimated model indicates that the costs of memory distortions decrease as time passes after information transmission, and students with low self-confidence had a greatly diminished preference for inflating self-confidence via memory distortions.

JEL codes: D91, I21, J83

Keywords: motivated beliefs, education, ability, recall, selective memory

*Faculty of Economics, University of Cambridge; vr277@cam.ac.uk. I am grateful to Toke Aidt, Leonardo Bursztyn, Christopher Rauh, Julia Shvets, Jack Willis, Weilong Zhang and attendees of the European Winter Meeting of the Econometric Society for their useful comments.

I Introduction

Self-assessments are fundamental to economic decision making under uncertainty. Despite mounting evidence that beliefs appear overconfident in a number of domains, including financial markets (De Bondt and Thaler [1995]) and personal ability (Svenson [1981]), the ubiquity of the benchmark model, which implies that agents do not preferentially absorb or forget information, has remained mostly free from formal challenge until relatively recently. The implications of possible deviations from the workhorse model are likely to be significant. Malmendier and Tate (2005) find that CEOs of Forbes 500 companies overestimate the returns to their own investment projects, explaining large distortions in investment allocations; meanwhile, Oster et al. (2013) demonstrate excessive optimism in testing and behaviour among individuals at risk of Huntington's disease.

A significant body of work on theories of preferential information acquisition (Rabin and Schrag [1999], Bénabou and Tirole [2002], Brunnermeier and Parker [2005], and Köszegi [2006]) has arisen to reconcile economic theory with biases in belief updating. The most pertinent strand for this study is motivated beliefs theory, which encapsulates the idea that agents have an incentive to ignore or forget unfavourable information to protect their egos, and provides a compelling account of how overconfidence can persist even if confronted with contradictory new information. However, while laboratory studies on motivated beliefs are manifold, identification challenges mean there is currently scant empirical evidence on how this phenomenon may affect belief formation in the real world.

Using a longitudinal field dataset, this paper examines adolescents' memories of report card grades in mathematics, comparing them to administrative data on actual grades. It begins by using reduced form analysis to characterise biases in recall behaviour. This exercise generates three main findings. First, students made more errors in recalling low grades, suggesting that findings in recent field research (Huffman et al. [2019]) can be sustained when individual-level heterogeneity is neutralised. Second, memories of grades became more positive when the survey was administered with a longer delay. This result extends recent findings from the laboratory (Zimmermann [2020]) into a field context, in which signals may be more consequential for core beliefs about ability. Third, recalling higher grades is positively associated with changes in self-confidence over time, controlling for actual grades. Notably, students with higher prior self-confidence exhibit stronger self-enhancing biases in both overall recall errors and memory loss, in further alignment with theories of motivated beliefs. This may mean high self-confidence is self-sustaining through motivated information processing.

The second component of analysis involves estimation of a static discrete choice model which separates the subconscious decision to distort memory into a confidence-enhancing motive and a desire to respect reality. In each period (at the time of the survey), students maximise their hedonic utility by making a subconscious discrete choice over all possible memories of their most recent maths grade. The benefit of remembering a higher grade is that it reinforces self-confidence, which directly impacts utility. However, students also face direct costs from memory distortion which increase in the distance of memories from actual grades. The model unearths deeper conclusions about the subconscious decision-making process underlying attainment recall. First, it suggests that the costs of memory distortion decrease markedly with time, while the benefits of self-confidence are unchanged. Second, it suggests that students without a pre-existing stated preference for maths, and those with poor prior academic self-confidence, are those with the weakest interest in enhancing future self-confidence using memory distortions. Nonetheless, all students exhibit a positive preference for ego enhancement. These findings match the predictions of the model in Bénabou and Tirole (2002), but comply less well with alternative models of confirmation bias (Rabin and Schrag [1999]) in which students with low confidence should exhibit a preference for negative signals.

The dataset I use is the Beginning School Study (BSS). The BSS was an educational study of Baltimore City public schools in the 1980s, combining administrative data on attainment from schools with subjective data from surveys, primarily of children but also their parents and teachers. The fundamental observational prediction of motivated beliefs theory is an asymmetry in information acquisition. This kind of asymmetry is prohibitively difficult to separate from standard Bayesian updating processes using standard data on beliefs, especially without full subjective distributions of prior self-confidence and signals (Benoît and Dubra [2011]). The BSS stands apart from other educational field datasets in its direct elicitation of subjects' **perceptions** of objective report card grades. In collecting students' recollections of recent signals, it permits the empirical analysis to avoid the problematic practice of inferring how signals are perceived from how beliefs update.

Until recently, empirical research was silent on whether motivated signal processing is likely to occur at the time of signal transmission or afterwards, via memory loss. When imperfect recall is only observed after transmission of information, it is impossible to determine the time at which information was distorted.[1] Zimmermann (2020) provides one exception in the laboratory. One month after an IQ test administered during the experiment, poor results were recalled with lower accuracy; at the time of the test, both good and bad results were remembered with equal accuracy. Correspondingly, subjects' self-confidence recovered from negative signals after one month. While Zimmermann's evidence supports the existence of motivated memory loss, it does not guarantee that more credible signals about ability in the field, such as those received during school, will selectively decay over time in a similar fashion. The incentives to retain signals about ability may be very different when prior beliefs are weaker and signals contain more information, as is the case during schooling years.

As previously highlighted, an idiosyncrasy of the BSS's design offers me an opportunity to extend the findings in Zimmermann (2020) into the field: the Fall edition of the survey was administered a few months after report cards, while the Spring edition was generally delivered with just a month's delay. This creates large, systematic differences in recall behaviour between the largely identical surveys in Spring and Fall, with a comparison thus yielding the effects of the passage of time on memory. While, unsurprisingly, recall accuracy is markedly inferior in Fall, it is also more positively biased. This positive drift is greater for students with preferences for maths and those with higher prior academic self-confidence. These findings thus jointly suggest that the temporal decay in memory exhibits the same motivated patterns observed in the static recall data. The structural model significantly enriches these findings: when the parameters are allowed to differ across the Fall and Spring editions of the survey, the cognitive costs of memory distortion are substantially diminished in Fall, when more time has passed, but the marginal benefit of self-confidence is unchanged. Importantly, the results do not eliminate the possibility that a portion of information distortion could have also been contemporaneous, through directed cognition (Gabaix et al. [2006]) or information avoidance: recall is strongly biased even in the Spring edition of the survey, which was administered around a month after report cards. In the context of the BSS, where many students were weakly attached to their education, opportunities to selectively avoid receiving the information contained in report cards were likely to be plentiful, especially when a student could anticipate receiving a poor grade in a given subject. Notably, though, the evidence in Zimmermann (2020) suggests one month is sufficient to generate imperfect recall,

[1] Theoretical papers representing signal distortion as contemporaneous include Rabin and Schrag (1999) and Aydogan et al. (2017). On the other hand, Bénabou and Tirole (2002) and Köszegi (2006) explicitly consider the possibility of selective memory. While the foundational discussions on the plausibility and reach of memory distortion owe more to philosophical and psychological than economic theory, these authors, as well as Golman et al. (2017), provide some comment on the relevant research.

albeit in a laboratory context where the content of signals is less potent and prior beliefs may be less diffuse.

Given the paucity of field data directly capturing signal interpretation and recall, most of the existing empirical research on motivated beliefs has taken place in the laboratory. Eil and Rao (2011) and Mobius et al. (2014) find evidence that individuals are more receptive to good news than bad news when it concerns their ego, but also that this asymmetry disappears in neutral contexts, supporting motivated beliefs over confirmatory bias. While Ertac (2011) also finds evidence of asymmetry, it is in the opposite direction: subjects are more sensitive to negative information about their relative performance on a task. The experimental evidence in Coutts (2019) differs from both. Coutts finds evidence of equal asymmetry across both self-relevant and self-irrelevant settings, supporting the presence of confirmation bias rather than motivated beliefs.

The experimental setting offers researchers supreme control of internal validity, but, depending on the context, may be limited in its ability to generate predictions of behaviour in realistic conditions.[2] As previously suggested, this limitation is likely to present especially significant challenges in a setting like this, where the subject of beliefs and the content of signals are of fundamental importance: the most basic feature of theoretical models of motivated beliefs is that distortions are more likely when the potential impact on an agent's ego is larger. For instance, the inconsistency across Eil and Rao (2011) and Mobius et al. (2014) relative to Ertac (2011) may be attributable to technical differences in the design, as the latter author suggests. But the disparate results may also be because Eil and Rao (2011) consider results on an IQ test, whereas Ertac (2011) examines a much more specialised assessment. The presence of ego in the former context seems intuitively more plausible than in the latter. Similarly, it is difficult to use experiments like Coutts (2019) as evidence contradicting motivated beliefs theory: similar updating behaviour in self-relevant and self-irrelevant domains can result from insufficient stakes in the self-relevant case. These discrepancies suggest room for field research to clarify what real-world stakes, if any, are sufficient to generate motivated beliefs in practice.

Oster et al. (2013) conduct a notable field study of information avoidance, distinct from biased recall, as a mechanism for belief protection. Their focus is on a case in which information is available at a low cost (through a test for Huntington's disease), but avoidable. In that setting, individuals avoid testing in order to eschew the pain of a positive diagnosis. Closer to the present research is Huffman et al. (2019), which presents field evidence associating biased recall with self-assessment, using cross-sectional data on

[2] See Levitt and List (2007) for a general but comprehensive discussion.

a workplace tournament. Much as in this paper, the first stage of their analysis involves demonstrating that recall errors are more likely when performance was poorer. The second stage demonstrates that those managers who have flattering memories of their performance are the ones most likely to overpredict their future performance relative to a model-based prediction.

My study builds on Huffman et al. (2019) in a few ways. The first is in demonstrating a positive drift in memory over time, indicating that biased memory loss accounts for a large part of the bias observed in ex post recall. The second is in my use of longitudinal data, which allows me to address potential concerns relating to unobserved cross-sectional heterogeneity. In the absence of within-individual variation, it would be difficult to preclude that individuals with lower ability and effort are not the ones most likely to receive poor performance signals and fail to recall or pay attention to them. Huffman et al. provide some evidence that unobserved ability does not account for their results: future performance in the tournament is uncorrelated with the probability of making a recall error. Nonetheless, persistent interpretive mistakes and exaggerations of performance could still play a role. My results offer some reassurance to the interpretation of their results since, at least in this context, individual heterogeneity is largely inconsequential for the observed relationship between recall accuracy and signal content.

In addition to populating a very scarce bank of empirical evidence on motivated beliefs, my research also contributes to a more recent research agenda within the economics of education. Economists have exerted considerable effort to understand gaps in educational attainment and why they widen over schooling years (Heckman et al. [2006]). This literature has begun to acknowledge the possibility of beliefs in ability detaching from reality, particularly in a recent focus on self-confidence as a non-cognitive skill (Cunha and Heckman [2007]; Cunha, Heckman, and Schennach 2010). However, relatively little is understood, especially empirically, about when and why this occurs. Heterogeneity in the use of motivated beliefs could be a viable candidate. To my knowledge, there is no research investigating whether motivated beliefs apply in schooling.[3] My results indicate that motivated protection of self-confidence occurs in early life, when beliefs about general ability are starting to be formed (Dweck [2002]). Most laboratory studies, and Huffman et al. (2019) in the field, concern narrower, domain-specific ability since credible signals about general ability are harder to provide later in life.

My results also enrich an emerging literature focusing more closely on feedback in schools. Dizon-Ross

[3]That being said, qualitative observations from education suggest that some children may be more "vulnerable" to negative signals than others (Dweck [2002]).

(2019) examines how uncertainty about children's academic abilities distorts investments in human capital in the context of a field setting. She demonstrates that poorer parents face greater informational frictions in assessing their children's abilities, including in a tangential examination of the dataset I use for this study. Kinsler, Pavan, et al. (2016) find evidence that parents' beliefs about ability are distorted by local grade averages. Other authors focus more closely on children's own beliefs: Alan et al. (2019) conduct a field intervention on "grit", demonstrating the malleability of children's beliefs on the relationship between effort and skill; Bursztyn et al. (2019) and Falk et al. (2020) illustrate the importance of the social environment in shaping children's attitudes towards self-assessment. My results, which suggest that motivated memory loss impacts the formation of self-confidence while it is at its most flexible, are consequential for how feedback should be conceptualised in this work. For instance, interventions which aim to increase information transmission regardless of its effect on self-confidence, such as in Dizon-Ross (2019), could have highly heterogeneous impacts if some students are more able to inure themselves to negative feedback than others. Furthermore, the role identified here for positively biased memory loss suggests that signal reinforcement may be required where confidence-deflating signals are socially valuable.

II Data: The Beginning School Study

This study uses survey data on students' perceptions of their academic ability and attainment, as well as administrative data on their actual attainment, from the Beginning School Study (BSS).[4] The BSS began in 1982, tracking a representative cohort of 790 children attending 20 Baltimore City public schools in the USA.

The backdrop of the BSS was one of striking educational disadvantage. When the study began, Baltimore was in the midst of a major economic decline which profoundly affected most of the students sampled (Alexander, Entwisle, and Olson [2014]). These challenges are visible in the demographic statistics in Table I: 61% of responding parents in the sample did not finish high school, 28% were non-employed sole earners, and 25% were the sole resident parent. In spite of the dispiriting context, parents were relatively optimistic about their children's educational prospects: in the first run of the survey, 98% expected their children to finish high school. Only 70% eventually did so without initially dropping out.

[4] Alexander and Entwisle (2003).

Table I: Sample demographics, Fall 1982 (first grade)

	Value
Child's race: Black	55%
Child's sex: Female	50%
Mother as parental respondent	86.3%
Parent school dropout	60.8%
Parent employed	60.4%
Parent not employed & sole earner	28.1%
Non-resident second parent	25.1%
Expect child not to finish high school	2%

The BSS conducted face-to-face interviews with students in almost every year of their education, sometimes twice, until leaving. It then made attempts to track respondents after they left school and began adult life, once in 1998 (with most respondents aged around 20) and again in 2006. It also surveyed parents and teachers, collecting subjective variables relating to students and other educational matters, but less frequently than students themselves.[5]

BSS survey sweeps repeatedly asked children what grade they remembered getting in their last quarterly report card. Survey waves were implemented twice a year, in Fall and Spring. In Fall (around November), children are asked to try to remember the grade they received at the end of the last academic year in June. In Spring (mostly in April or May), children were asked to remember the grade they got a few weeks earlier, in the third quarter. Recall accuracy is around 70% in Spring, but only around 50% in Fall (Table II). Notwithstanding the substantial seasonal variation, children's recall accuracy was largely stable over time, other than in fourth grade (which I thus omit from the analysis). The comparatively weak educational setting of the BSS likely weighed on the absorption of attainment information by students. In particular, imperfect recall observed in Spring suggests that there may have been ample opportunities for adolescents to avoid absorbing report card grade information in the moment.[6] However, recall accuracy is significantly diminished in Fall. This difference presents a valuable opportunity to explore the effects of the passage of time on memory, to be revisited in both the reduced form and structural analysis to follow.

[5] It would also be of interest to examine parents' attitudes, but less frequent sampling means sample sizes can become rather small when the econometric design involves a dynamic component. As such, the focus of this paper is on children's own beliefs about their ability.

[6] Accuracy of recall may still be somewhat better than in Dizon-Ross (2019), where 60% of parents in the Malawian sample say they did not know their child's last report card grade.

Table II: Summary statistics

Sweep	N	School grade (modal)	Mean maths mark (recalled)	Mean maths mark (actual)	% correct recall
Fall ’85	531	4	3	2.4	40.9
Fall ’87	496	6	2.9	2.4	49.6
Spring ’88	465	6	2.5	2.2	74
Fall ’88	184	6	2.8	2.2	47.8
Spring ’89	172	7	2.2	1.9	67.4
Fall ’89	381	8	2.6	2.1	51.2
Spring ’90	409	8	2.5	2.2	72.1
Fall ’90	444	9	2.7	2.2	48
Spring ’94	143	12	2.3	2	60.1

Note: Mean maths marks computed by assigning 4 to *Excellent*, 3 to *Good*, 2 to *Satisfactory*, and 1 to *Unsatisfactory*. Fall ’85 is excluded from all the ensuing analysis.

III Reduced Form Analysis

Using data from the BSS, I now present some reduced form evidence consistent with motivated beliefs theory. In what follows, t denotes the survey sweep, so that subscript it denotes the last available observation of the given variable for individual i when survey t was collected. Throughout the paper, s_{it} denotes report card grades in maths.[7] $\tilde{s}_{it}$ denotes recalled report card grades in mathematics. Grade recall was elicited using the following question: *“Remember the last report card you got when school ended for the summer? You could have gotten marks like E (Excellent), G (Good), S (Satisfactory), or U (Unsatisfactory). What mark did you get in Mathematics?”* As such, the domain for both s_{it} and $\tilde{s}_{it}$ is $\{E,G,S,U\}$; see Table III for a translation into numerical marks.[8] The other notable variable I use is D_{it}, a survey indicator of academic self-confidence; full coding can be found in Table IV.

III.A Recall Errors

It is already clear from the summary statistics in Table II that recall of grades was generally imperfect in the BSS. That is, very often, $\tilde{s}_{it} \neq s_{it}$. In principle, however, this event can result from either motivated or

[7]The BSS also collected information on grades in reading and, later, science. Since they were less consistently recorded, I focus solely on mathematics grades.

[8]Not all schools used the $\{E,G,S,U\}$ grading scale. I use a conversion table in the BSS documentation to map actual grades to that scale. Students were asked to recall grades on the $\{E,G,S,U\}$ scale, but they were also provided with the percentage score mapping of each grade category in Table III.

Table III: Grade interpretation

s_{it}	Mark
Excellent	90–100%
Good	80–89%
Satisfactory	70–79%
Unsatisfactory	<70%

Table IV: Coding for D_{it}

How smart do you think you are compared to other kids in your school this year?	D_{it}
One of the smartest	5
Smarter than most kids	4
About as smart as everybody else	3
Not as smart as most kids	2
Not very smart at all	1

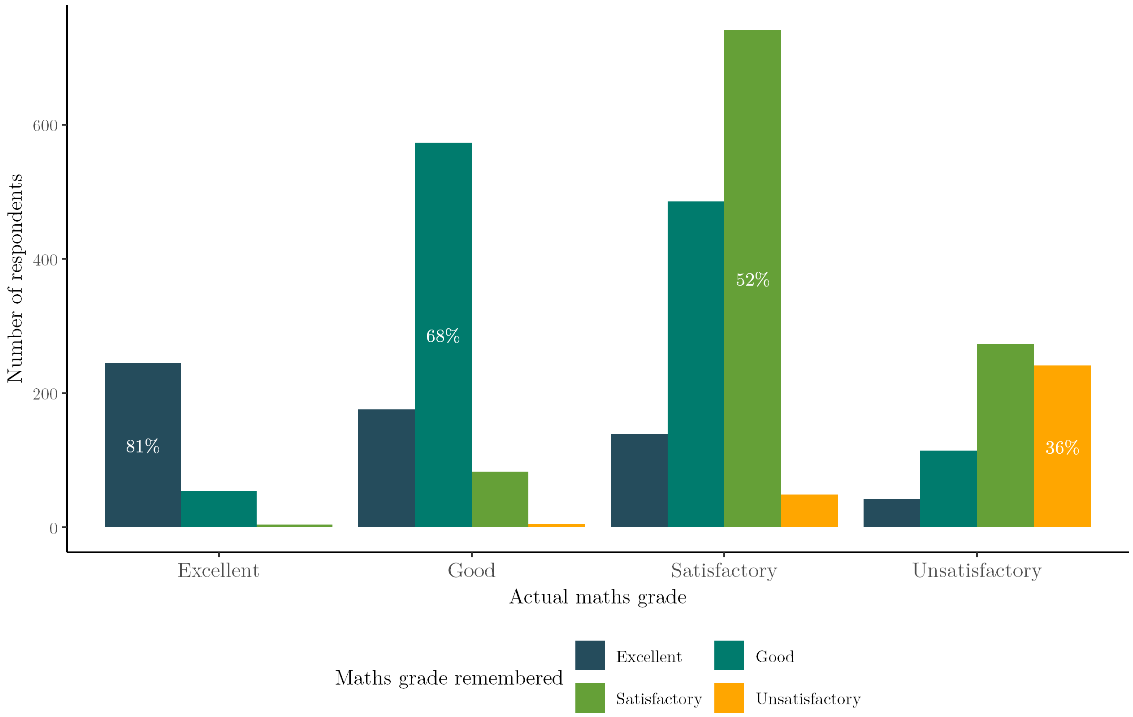
unmotivated errors of recall.

Theories of motivated beliefs imply that information distortion is used as a tool to gain, or maintain, self-confidence. In this context, the event that a grade is forgotten, $\mathbb{1}(\tilde{s}_{it} \neq s_{it})$, would be the outcome of an optimisation problem. On one side of the trade-off is the incentive to preserve positive self-image by distorting or ignoring incoming bad news. Empirically, academic self-confidence has been associated with greater happiness in adolescents (Cheng and Furnham [2002]): it embodies a hedonic benefit. More recent work in economics also outlines an instrumental benefit of higher self-confidence (Cunha and Heckman [2007], Cunha, Heckman, and Schennach [2010], and Alan et al. [2019]): it motivates students to work harder.

On the other side of the trade-off are the costs of distortion: first, information suppression or avoidance may incur a psychological and practical burden, and second, students face material incentives to hold correct beliefs. In the context of schooling, correct beliefs help students avoid making overly optimistic task and effort choices, as well as more significant decisions on whether to remain in education. The presence of this trade-off means information suppression should be used selectively, when it has a more negative impact on the individual. Before attempting a more complex analysis which separates these underlying incentives in Section IV, I follow most of the empirical literature in simply examining whether the probability of making a recall error differs across good and bad news.

Concretely, the course of events leading to recall of a grade ensues can be envisaged as follows. The opportunities for information avoidance and suppression begin when report cards are transmitted to students, who could ignore their report card if anticipating its poor content via other informal signals, or even avoid fully digesting information while seeing it (Gabaix et al. [2006]). After that initial assimilation of information comes the subconscious decision to retain it, and it is at this point that the decay of information via memory loss becomes more important. Both of these mechanisms, contemporaneous and retrospective, culminate in measured recall errors. While I examine their combined effects in this section, I demonstrate that biased memory loss plays a distinct and significant role in Section III.B.

Figure I: Remembered grades vs. actual grades, split by survey edition



Note: Aggregated over survey sweeps. White percentages are accurate recall rates by grade.

Figure I provides a clear indication of biased recall in the BSS data. It plots actual grades received in maths and splits them by the grade the child remembered. The first observation is that recall errors are positively skewed, other than for the highest grade (for which upward errors are impossible): children generally remembered getting better grades than they actually did. Secondly, errors become increasingly prevalent the lower the actual grade is.

In order to formalise this analysis, I estimate a regression whose dependent variable is the event that a student incorrectly recalls their grade, $\mathbb{1}(\tilde{s}_{it} \neq s_{it})$.[9] The aim is to check whether the relationship between recall errors and the qualitative content of grades is robust to confounding variables. Arguably the most problematic of those here are cognitive ability and engagement with schooling, which one might expect

[9]This dependent variable also permits negative recall errors. I allow both positive and negative errors because of boundary conditions for the top and bottom grades: it is not possible to remember a better grade than *Excellent* and a worse one than *Unsatisfactory*.

to be correlated with both grades, s_{it}, and the event that they are correctly recalled, $\mathbb{1}(\tilde{s}_{it} \neq s_{it})$. Since the BSS observed students longitudinally, I can make major progress in addressing these concerns by including individual and grade-level fixed effects in the analysis. Intuitively, that means I examine whether the same individuals are more likely to forget lower grades than higher ones. I exclude Fall '85, when most students were in 4^{th} grade and recall was very poor (Table II), from the sample. The regression sample therefore largely follows the cohort while in their early adolescence (mostly, aged 11–15), passing through school grades 6–9 and 12. The empirical specification is a linear probability model with robust standard errors,

$$\mathbb{1}(\tilde{s}_{it} \neq s_{it}) = \beta_0 + \sum_{z \in \{G,S,U\}} \beta_{1z} \mathbb{1}(s_{it} = z) + \beta_2 Fall_t + \alpha_i + \gamma_t + \varepsilon_{it}. \tag{1}$$

As in most of the existing empirical literature, $\sum_{z \in \{G,S,U\}} \beta_{1z} \mathbb{1}(s_{it} = z)$ in Equation (1) should capture the motivated component of $\mathbb{1}(\tilde{s}_{it} \neq s_{it})$: they measure the relationship between the probability of incorrectly recalling a grade and the quality of the grade itself. Like in the raw data, the estimated coefficients under column (1) in Table V provide clear evidence that adolescents are more likely to fail to recall poor grades. This effect is very large: students were almost 50 percentage points (pp) more likely to correctly recall getting the highest grade, *Excellent* (marks above 90%), than the lowest one, *Unsatisfactory* (marks lower than 70%). The three lower grades were less likely to be recalled than the highest one. As already indicated, recall deteriorates substantially in the Fall edition of the survey, which was delivered much longer after report card grades than the Spring edition.

These regressions vindicate the pattern in the raw data, visible in Figure I. They suggest that adolescents deploy motivated cognition when interpreting grades in general: on the whole, they are more likely to ignore or forget bad signals than good ones. However, the analysis can deepen its support of the theory by considering what kind of students are especially likely to exhibit bad-news avoidance in order to maintain their self-confidence. An auxiliary prediction from Bénabou and Tirole (2002)'s model is that individuals with higher initial self-image have a greater incentive to ignore or forget information: they have more to lose, driven by a non-linearity in the benefit of high self-confidence. Since the BSS provides us with a measure of prior academic self-confidence in $D_{i,t-1}$, an immediate test of this prediction is that the relationship between $\mathbb{1}(\tilde{s}_{it} \neq s_{it})$ and s_{it} is stronger when $\mathbb{1}(D_{i,t-1} > 3)$. That is, students with higher prior self-confidence exhibit a greater tendency to forget bad grades more than good ones.

This hypothesis finds support from column (5) in Table V, which include an interaction of s_{it} with

Table V: Reduced form models for recall of Maths grades

	Dependent variable:					
	$1(\tilde{s}_{it\|t} \neq s_{it})$					
	Base	Base	Prior belief	Prior belief	Subject preference	Subject preference
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(s_{it} = Good)$	0.105**	0.124***	−0.007	0.044	−0.021	−0.001
	(0.041)	(0.031)	(0.063)	(0.050)	(0.074)	(0.058)
$\mathbb{1}(s_{it} = Satisfactory)$	0.264***	0.247***	0.127**	0.141***	0.147**	0.102*
	(0.042)	(0.029)	(0.062)	(0.048)	(0.074)	(0.056)
$\mathbb{1}(s_{it} = Unsatisfactory)$	0.512***	0.457***	0.406***	0.378***	0.377***	0.289***
	(0.048)	(0.032)	(0.067)	(0.051)	(0.079)	(0.059)
Maths favoured			−0.166**	−0.085		
			(0.068)	(0.053)		
$\mathbb{1}(D_{i,t-1} > 3)$					−0.158*	−0.165***
					(0.084)	(0.062)
Fall	0.230***	0.230***	0.230***	0.227***	0.206***	0.221***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.025)	(0.024)
$\mathbb{1}(s_{it} = Good)$*Maths favoured			0.184**	0.130**		
			(0.076)	(0.063)		
$\mathbb{1}(s_{it} = Satisfactory)$*Maths favoured			0.240***	0.198***		
			(0.075)	(0.060)		
$\mathbb{1}(s_{it} = Unsatisfactory)$*Maths favoured			0.163**	0.154**		
			(0.082)	(0.067)		
$\mathbb{1}(s_{it} = Good)*\mathbb{1}(D_{i,t-1} > 3)$					0.168*	0.184**
					(0.092)	(0.072)
$\mathbb{1}(s_{it} = Satisfactory)*\mathbb{1}(D_{i,t-1} > 3)$					0.117	0.192***
					(0.092)	(0.070)
$\mathbb{1}(s_{it} = Unsatisfactory)*\mathbb{1}(D_{i,t-1} > 3)$					0.217**	0.269***
					(0.100)	(0.077)
Individual FE	Yes	No	Yes	No	Yes	No
Academic year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,659	2,659	2,605	2,605	2,145	2,145
R^2	0.372	0.133	0.368	0.139	0.394	0.125
Adjusted R^2	0.194	0.129	0.195	0.134	0.194	0.118

Note: *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. $\mathbb{1}()$ is the indicator function. $\mathbb{1}(D_{i,t-1} > 3)$ = high self-confidence. s_{it} = actual maths grade. $\tilde{s}_{it}$ = recalled maths grade.

$\mathbb{1}(D_{i,t-1} > 3)$.[10] The effect of high prior self-confidence seems to be to widen the gap between recall of *Excellent* grades and all the others: these students were more likely to correctly recall getting *Excellent* grades and less likely to correctly recall the others. Very similar patterns can be observed for students who stated that maths was their favourite subject in the last available period prior to t, in column (3).[11] In the absence of higher stakes (when a preference for maths was not expressed, or when initial beliefs in ability were lower) students were equally likely to remember *Good* and *Excellent* grades.

The results in column (5) also help to differentiate between models of motivated beliefs, in which information is selected in order to maintain self-confidence, and alternatives. The most prominent of those alternatives is confirmation bias (Rabin and Schrag [1999]), in which individuals are more likely to misinterpret signals which do not accord with their existing belief, regardless of whether those signals convey good or bad news about the ego. Distinguishing between these models is generally difficult: in the experimental literature, the approach usually taken is to compare signal interpretation in conditions in which stakes are high and low, but that method may be problematic when stakes are not adequately replicated in the laboratory.

Under a model of confirmation bias, the expectation would be for students with poor initial beliefs about their ability to prefer further negative signals. The large magnitude of the effect in column (1) provides a first indication that this is unlikely to be the case: the tendency to preferentially recall better signals is extremely strong across the sample. This assertion is confirmed by the results in column (5). The coefficients on $\mathbb{1}(s_{it} = z)$ for $z \in \{Satisfactory, Unsatisfactory\}$ demonstrate that even students who believe they are of average or below average ability are much more prone to forgetting bad grades than good ones. Even greater insight into this matter is provided by the analysis to come in Section IV.D, which decomposes this same heterogeneity into the opposing incentives underlying memory distortions.

Each of the even-numbered columns in Table V omits individual fixed effects from the specification to its left. In all cases, the results undergo little qualitative change other than some small differences in relative magnitudes. This conclusion could provide some reassurance to existing field research lacking in longitudinal variation (Huffman et al. [2019]), since it indicates that fixed unobserved individual characteristics, such as cognitive ability or engagement with schooling, are not an important joint determinant of the quality of signals and the ability or willingness to recall them correctly.

[10] $\mathbb{1}(D_{i,t-1} > 3)$ is measured in the last available period prior to t for each student to minimise data loss.

[11] While being a less direct measure of self-confidence, a preference for maths may also capture a greater hedonic motive for, or consumption value of, self-confidence. This is another incentive for motivated beliefs explored in Bénabou and Tirole (2002).

III.B The Role of Memory Loss

As noted by theoretical work (Bénabou and Tirole [2002]), observed biases in recall alone cannot be used to infer biases in memory loss. The findings in Table V, much like those in Huffman et al. (2019), could reflect selective signal avoidance: since students receive a stream of signals on their performance through the academic quarter, they could avoid reading a report card grade if anticipating its unfavourable contents. And even if report cards were unavoidable, students could likely leverage mechanisms like directed cognition (Gabaix et al. [2006]) to avoid absorbing unfavourable information in the moment. However, the design of the BSS presents a compelling opportunity, mirroring the approach taken by Zimmermann (2020) in the laboratory, to discern whether biased memory loss serves the cause of motivated beliefs.

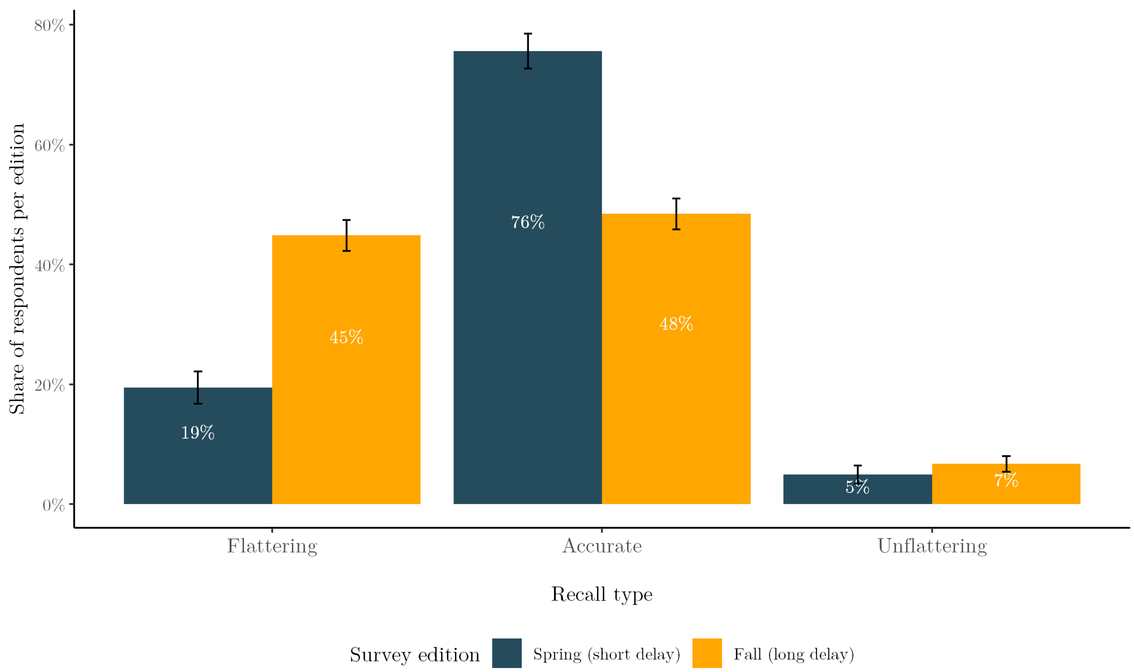
In both editions of the survey within each year, Spring and Fall, adolescents were asked to remember their most recent grades. In Spring those grades were likely from just over a month earlier, but in Fall, the most recent grades were likely received more than three months ago. It is therefore likely that differences in recall between Fall and Spring reflect the effects of memory loss. The analysis in this section establishes that, since patterns of recall are markedly more motivated (positively biased) in Fall than in Spring, a large part of the information selection underlying motivated beliefs in this context is enabled by biased memory loss.

Figure II plots types of recall by survey edition for the middle two grades, *Good* and *Satisfactory*. For these grades, both positive ('flattering') and negative ('unflattering') recall errors are possible. 'Accurate' recall occurs when the correct grade is remembered. As previously noted, recall is less accurate in Fall than in Spring: the probability of remembering the correct grade falls from 76% to 48%. However, almost all of this additional error is concentrated in a redistribution towards flattering errors in Fall compared to Spring. By comparing the difference between flattering and unflattering errors in Fall and Spring, we can reach the reasonable conclusion that $^2/_3$ of the bias towards flattering errors in Fall is explained solely by memory loss over time after the first month since information transmission. Assuming truthful reporting, the remaining $^1/_3$ could be explained by either memory loss during the first month or information avoidance.[12]

The next step is to verify that this positive drift in memory is retained in a regression approach where unobserved cross-sectional heterogeneity can be neutralised. In this case, within-student variation can be

[12] It should be noted that in the raw data (see Figure A4 in Appendix A) is that even in the Spring edition, recall is worse for the lower grades, *Satisfactory* and especially *Unsatisfactory*. Given the relatively short delay for the Spring survey, of just over a month, this could support some role for selective attention or information avoidance. That being said, Zimmermann (2020) successfully uses a delay of one month to measure the effects of memory loss in his laboratory test of the same phenomenon.

Figure II: Recall type, split by survey edition



Note: Only *Good* and *Satisfactory* grades included, since it is possible to remember either a better ('flattering') or a worse ('unflattering') grade. Error bars depict 95% confidence intervals.

used to verify that when the same student is sampled in Fall, after a longer delay relative to grades, their memory tends to be more flattering. Since the objective is now to examine whether memory becomes more optimistic over a longer horizon, given the actual signal received, I use $\mathbb{1}(\tilde{s}_{it} = Excellent)$ — the event that the student recalls getting an *Excellent* grade — as the dependent variable, controlling for the actual grade they received, s_{it}.[13]

$$\mathbb{1}(\tilde{s}_{it} = E) = \delta_0 + \sum_{z \in \{G,S,U\}} \delta_{1z} \mathbb{1}(s_{it} = z) + \delta_2 Fall_t + \alpha_i + \gamma_t + \varepsilon_{it}. \tag{2}$$

The statistically significant and positive estimate for δ_2, the coefficient on Fall in column (1) of Table VI, implies that being surveyed in Fall has a positive impact on students' propensity to remember getting the highest grade. This formalises the positive drift in memory illustrated in raw data in Figures II and A4, and demonstrates its robustness to a within-individual identification strategy. That is, the same students are more likely to remember higher grades when surveyed in Fall than in Spring.

As in Section III.A, I can check whether the generally positive drift in memory over time is different for children who express a greater preference for maths, or have higher prior self-confidence. This is achieved by interacting $Fall_t$ with those two variables, as in specifications (2) and (3) in Table VI. Once again, there is highly compelling evidence that both groups of students exhibited stronger motivated information processing: in this case, they were more prone to an upward drift in their memory of grades as time elapses. Combined with the heterogeneity in general recall in Table V, these results suggest that those with higher initial confidence and enthusiasm were the ones most likely to inure themselves to incoming negative feedback, and that selective memory may have been an important tool in facilitating this defence mechanism. Notably, though, even students with lower prior beliefs in their ability experienced a significant positive drift in memory over time, demonstrated by the positivity of the coefficient on $Fall_t$ in column (3). This corroborates a similar finding in Zimmermann (2020), and, matching his interpretation, further strengthens the suggestion that the results reflect motivated memory loss rather than, for example, a drift of memories towards prior beliefs.

[13] Another option would be to examine whether the probability of making a positive recall error increases in Fall, as in Figure II. However, as in Section III.A, this strategy would only permit me to examine cases where the middle two grades, *Good* and *Satisfactory*, were received. It would involve discarding a large quantity of useful information on how memory loss affects students receiving the top and bottom grades.

Table VI: Spring/Fall recall comparison

	Dependent variable:					
	$1(\tilde{s}_{it\|t} = \text{Excellent})$					
	Base	Base	Subject preference	Subject preference	Prior belief	Prior belief
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(s_{it} = Good)$	-0.598^{***}	-0.601^{***}	-0.599^{***}	-0.606^{***}	-0.576^{***}	-0.631^{***}
	(0.036)	(0.030)	(0.035)	(0.030)	(0.045)	(0.035)
$\mathbb{1}(s_{it} = Satisfactory)$	-0.683^{***}	-0.725^{***}	-0.679^{***}	-0.717^{***}	-0.637^{***}	-0.733^{***}
	(0.036)	(0.027)	(0.036)	(0.028)	(0.045)	(0.033)
$\mathbb{1}(s_{it} = Unsatisfactory)$	-0.698^{***}	-0.743^{***}	-0.693^{***}	-0.727^{***}	-0.650^{***}	-0.758^{***}
	(0.037)	(0.028)	(0.037)	(0.028)	(0.047)	(0.033)
Fall	0.073^{***}	0.077^{***}	0.040^{***}	0.048^{***}	0.051^{***}	0.064^{***}
	(0.012)	(0.012)	(0.014)	(0.013)	(0.017)	(0.017)
Maths favoured			-0.002	0.041^{***}		
			(0.019)	(0.013)		
Fall*Maths favoured			0.070^{***}	0.061^{***}		
			(0.023)	(0.023)		
$\mathbb{1}(D_{i,t-1} > 3)$					0.007	0.028^{**}
					(0.020)	(0.014)
Fall*$\mathbb{1}(D_{i,t-1} > 3)$					0.062^{**}	0.051^{*}
					(0.028)	(0.029)
Individual FE	Yes	No	Yes	No	Yes	No
Academic year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,659	2,659	2,605	2,605	2,145	2,145
R^2	0.551	0.337	0.546	0.352	0.563	0.340
Adjusted R^2	0.424	0.334	0.422	0.348	0.419	0.335

Note: *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. $\mathbb{1}()$ is the indicator function. $\mathbb{1}(D_{i,t-1} > 3)$ = high self-confidence. s_{it} = actual maths grade. $\tilde{s}_{it}$ = recalled maths grade.

III.C Recall and Self-Confidence

Sections III.A and III.B have outlined evidence that adolescents in the BSS exhibit recall behaviour consistent with motivated beliefs. I now extend the reach of that finding in demonstrating that this information distortion may affect how students update their self-confidence in their academic ability over time. I do so by investigating sequential updating of $\mathbb{1}(D_{it} > 3)$, the survey indicator of academic self-confidence (Table IV). Given that I need to include a lagged dependent variable $\mathbb{1}(D_{i,t-1} > 3)$ to capture sequential belief updating, but I also wish to control for individual heterogeneity α_i, I use Bartolucci and Nigro (2010)'s Quadratic Exponential model. The model permits joint estimation of the marginal effects of the following factors: $\mathbb{1}(D_{i,t-1} > 3)$, an individual level effect, and the covariates of interest, s_{it} and $\tilde{s}_{it}$ (actual and recalled mathematics grades). I also include $\hat{s}_{it}$, students' $t-1$ predictions of the grade they would receive in period t, in order to permit grade surprises to impact beliefs. In Equation (3), I write $y_{it} = \mathbb{1}(D_{i,t-1} > 3)$ for brevity, and x_{it} is a vector with the full set of dummy variables for s_{it}, $\tilde{s}_{it}$, and $\hat{s}_{it}$ over the domain $\{G,S,U\}$, leaving the omitted category E. The model from Bartolucci and Nigro (2010) then implies a conditional likelihood of the following form for y_{it}:

$$p(y_{it}|\alpha_i, X_i,) = \frac{\exp(y_{it}[\alpha_i + x'_{it}\beta + y_{i,t-1}\gamma + e^*_t(\alpha_i, X_i)])}{1+\exp(\alpha_i + x'_{it}\beta + y_{i,t-1}\gamma + e^*_t(\alpha_i, X_i))}, \tag{3}$$

where, for $t < T$,

$$e^*_t(\alpha_i, X_i) = \log \frac{1+\exp(\alpha_i + x'_{i,t+1}\beta + e^*_{t+1}(\alpha_i, X_i) + \gamma)}{1+\exp(\alpha_i + x'_{i,t+1}\beta + e^*_{t+1}(\alpha_i, X_i))}, \tag{4}$$

and

$$e^*_T(\alpha_i, X_i) = \phi + x'_{i,t+1}\beta^*. \tag{5}$$

It is easier to understand the model in its latent index form. I thus introduce a latent continuous variable for self-confidence, $\tilde{\theta}_{it}$, satisfying $y_{it} = \mathbb{1}(D_{it} > 3) = \mathbb{1}(\tilde{\theta}_{it} > 0)$. Then, the model specified above is equivalent to estimating

$$\tilde{\theta}_{it} = \alpha_i + x'_{it}\beta + y_{i,t-1}\gamma + e^*_{t+1}(\alpha_i, X_i) + \varepsilon_{it}, \tag{6}$$

where ε_{it} are independent with a logistic distribution. The main aspect of Equation (6) which requires comment is the error correction term $e^*_{t+1}(\alpha_i, X_i)$. This term implies current self-confidence is bolstered by the anticipated impact of high self-confidence, $\mathbb{1}(D_{it} > 3)$, on the likelihood of future high self-confidence, as detailed by Equation (4). This specification has a relatively intuitive interpretation here: current confidence may be bolstered by the knowledge that future high self-confidence may positively influence outcomes like school continuation and college attendance.

The estimated coefficients within Equation (6) are reported in Table VII. The first observation is that self-confidence exhibits substantial autocorrelation, even when static individual heterogeneity is controlled for. In spite of this persistence, remembered grades can substantially impact self-confidence. Remembering a $Satisfactory$ grade, compared to an $Excellent$ one, implies a deterioration in self-confidence which swamps the positive effect of being relatively confident in the previous period: the corresponding decrease in the odds of being relatively confident is about 60%. On the other hand, actual grades are not statistically important for the updating self-confidence. The relative effect being larger for remembered grades is unsurprising: actual grades should have no impact on beliefs if adolescents do not remember them correctly. However, it is notable that sequential changes in beliefs are materially associated with distorted memories. Concretely, a teenager who received *Satisfactory* but remembers getting *Good* appears to update their belief about their ability in the same way as one who actually got a *Good* grade and correctly recalled it. This appears to vindicate students' deployment of selective self-deception: recalling a low grade has material consequences for the evolution of self-confidence. I can also utilise this behavioural tendency in Section IV.A when combining belief updating and memory distortion into a single framework.

III.D Reduced Form: Summary

The reduced form results described above combine to develop a coherent narrative on dynamic signal perception and beliefs in the BSS. Firstly, students in this dataset demonstrate a clear preference for overly positive recollections of grades, and recall errors are more likely the lower the grade is. Secondly, a large part of this information distortion occurs *ex post*: when more time elapses between the report card and the survey, memories of grades become markedly more optimistic. Finally, remembered grades matter for the updating of self-confidence while actual grades do not.

Particularly notable is the heterogeneity in information distortions across students. In particular, the analysis establishes that students with higher initial beliefs in their ability, and those with a greater preference for maths, were more likely to use signal and memory distortion in a way that benefited their egos. Higher prior beliefs about ability seem to be associated with individuals being more able to armour themselves against bad news. Conversely, this implies that those with lower self-confidence — in the context of

Table VII: Regression for belief updating — Quadratic Exponential with school year fixed effects and conditional individual heterogeneity

	Dependent variable:
	$\mathbb{1}(D_{it} > 3)$
Actual grades	
$\mathbb{1}(s_{it} = Good)$	0.619
	(0.464)
$\mathbb{1}(s_{it} = Good)$	0.573
	(0.489)
$\mathbb{1}(s_{it} = Unsatisfactory)$	0.440
	(0.557)
Recalled grades	
$\mathbb{1}(\tilde{s}_{it} = Good)$	-0.342
	(0.292)
$\mathbb{1}(\tilde{s}_{it} = Satisfactory)$	-0.856^{**}
	(0.351)
$\mathbb{1}(\tilde{s}_{it} = Unsatisfactory)$	-1.474^{***}
	(0.525)
Expected grades	
$\mathbb{1}(\hat{s}_{it} = Good)$	0.024
	(0.253)
$\mathbb{1}(\hat{s}_{it} = Satisfactory)$	0.202
	(0.315)
$\mathbb{1}(\hat{s}_{it} = Unsatisfactory)$	0.295
	(0.686)
Lagged self-confidence	
$\mathbb{1}(D_{i,t-1} > 3)$	0.505^{***}
	(0.195)
Observations	1,678
R^2	0.550
Adjusted R^2	0.356
Residual Std. Error	0.380 (df = 1171)

Note: *p<0.1; **p<0.05; ***p<0.01. Quadratic exponential model; coefficients are log-odds. Standard errors in brackets. $\mathbb{1}()$ is the indicator function. $\mathbb{1}(D_{it} > 3)$ = high self-confidence. s_{it} = actual maths grade. $\tilde{s}_{it}$ = recalled maths grade. $\hat{s}_{it}$ = expected maths grade.

$\mathbb{1}(D_{i,t-1} > 3)$, those who consider themselves to be of average or below-average ability — may be more vulnerable to bad signals when they do receive them. This finding, as well as those relating to the implications of time for motivated memory loss, is developed further in the analysis to follow.

IV Structural Model

I now proceed to explore the motives underlying self-confidence maintenance and memory distortion jointly in a single framework. The basic foundation of the model is a representation of memory distortion as the outcome of a subconscious optimisation program. In each period (at the time of the survey being administered), students maximise their periodic utility by making a discrete choice over possible memories of their most recent maths grade. In the model, the benefit of remembering a higher grade is assumed to be that it reinforces self-confidence. Students are likely to be protective of their self-confidence because doing so embodies both a hedonic benefit (Cheng and Furnham [2002]) and an instrumental one (Cunha and Heckman [2007], Cunha, Heckman, and Schennach [2010], and Alan et al. [2019]). However, students also face direct costs from memory distortion, structured so that the truth acts as a centre of gravity for memories. By exploring the relative role of these opposing motives, and especially how they differ across students and over time (Section IV.D), I can greatly refine the conclusions from the reduced form analysis in Section III.[14]

The model's representation of the subconscious decision-making process underlying memory distortions is intentionally parsimonious. Its simplicity is a boon in providing a bridge from the reduced form results to the estimated structural parameters: the model brings together the findings that students do more to distort negative memories (Section III.A) and that memories appear to materially impact the evolution of self-confidence over time (Section III.C) by casting the latter as the motive for the former. By measuring the importance of actual attainment for memory choices, it then separates the self-confidence-enhancing motive from the "reality constraints" faced by students.

The model provides highly intuitive estimates of the parameters governing both elements of the subconscious decision-making process, and notably how those parameters vary across circumstances and student characteristics. It suggests that the "reality constraints" on memory distortion loosen with the passage

[14]Due to the greater computational demands of this analysis, which takes place in a multinomial logit model with latent variables appearing in utility functions, I no longer control for individual fixed effects. This approach can be justified on the basis that the results relating to recall in Section III are largely unaffected by their exclusion.

of time, but the motive to enhance self-confidence does not change. Conversely, students with low self-confidence and weaker preferences for maths exhibit diminished preferences for protecting self-confidence using memory distortions, but no reduction in the strength of reality constraints.

IV.A Model Description

I drop the i subscript throughout this section because individual fixed effects are no longer in use. Students' utility functions $U_{z,t}$ are defined over the set of remembered grades $\tilde{s}_t = z \in \{E = \mathit{Excellent}, G = \text{Good}, S = \text{Satisfactory}, U = \text{Unsatisfactory}\}$,

$$\begin{aligned} U_{z,t} &= \lambda \tilde{\theta}(\tilde{s}_t = z, \Omega_{t-1}) + c(\tilde{s}_t = z, s_t) + \varepsilon_{z,t} \\ &= V_{z,t} + \varepsilon_{z,t}. \end{aligned} \tag{7}$$

In Equation (7), Ω_{t-1} is the student's psyche in period $t-1$. If true ability in period t is θ_t, $\tilde{\theta}_t(.)$ is a function describing the student's expectation of θ_t implied by Ω_{t-1} and the latest grade the student recalls getting, $\tilde{s}_t$. More simply, $\tilde{\theta}(\tilde{s}_t = z, \Omega_{t-1}) = E(\theta_t | \tilde{s}_t = z, \Omega_{t-1})$, where $\tilde{\theta}_t(.)$ is continuous in Ω_{t-1} conditional on $\tilde{s}_t$. Meanwhile, $c(\tilde{s}_t = z, s_t)$ increases symmetrically in the distance between $\tilde{s}_t$ and s_t, and measures reality constraints: the costs of memory decoupling from reality. The key empirical tool provided to the model is thus that actual grades are defined to only be relevant for choices insofar as they impact reality constraints; the benefit of remembering a higher grade is its effect on self-confidence, for which only recalled information (not actual information) matters. As indicated in Section III.C, actual grades have no statistical association with the updating of self-confidence over time.

In a departure from most standard models of choice, the choice problem represented by Equation (7) cannot be interpreted as an explicit decision. It is better seen as a representation of a largely subconscious process intending to protect the individual's psychological welfare.[15] This interpretation also lends some credibility to the assumption that all variables in Equation (7) are observed by the (implicit) decision maker, including $\varepsilon_{z,t}$. On that basis, the ensuing analysis assumes that the student's subconscious simply selects the memory which maximises periodic utility.[16]

In Equation (7), it is not essential that students pursue ego-protective signal processing. If some stu-

[15] See Bénabou and Tirole (2002) for more discussion on optimisation and the "psychological immune system".

[16] Strictly speaking, the model could also capture some conscious element of decision making, especially in the case of information avoidance: students may knowingly avoid reading report cards when they anticipate their poor contents.

dents preferred memories which negatively affected their self-esteem, for instance, their utility functions should feature $\lambda < 0$. Such a tendency could arise from the kind of confirmation bias posited by authors in both psychology (Kwang and Swann Jr [2010]) and economics (Rabin and Schrag [1999]). While I have already obtained evidence that such a tendency is unlikely in this context, given that even students with low self-confidence tend to forget bad memories more often (Section III), estimating the structural parameters heterogeneously over the sample offers more precise insight.

IV.B Estimation

In order to estimate the model using the BSS data, I need to introduce additional assumptions. The first is a standard one in random utility models, which is that $\varepsilon_{z,t}$ has a Type-1 Extreme Value distribution. The second is to define $\tilde{\theta}(.)$, recycling the notation in Section III.C, as a latent variable underlying the discrete survey indicator D_t (*"How smart do you think you are compared to other kids in your school this year?"*). These two assumptions bring the model into the known class of discrete choice problems with latent elements of utility, generally known as "Hybrid Choice Models" (Ben-Akiva et al. [2002]). I also assume that D_{t-1}, the lagged survey indicator of self-confidence, is a sufficient statistic for Ω_{t-1}.[17] Then, for parsimony, I impose two restrictions on $c(\tilde{s}_t = z, s_t)$: firstly, $c(\tilde{s}_t = s_t, s_t)$ is set to 0 as a normalisation: the cost of remembering the correct grade is 0. Furthermore, $c(.)$ is set to depend only on the absolute number of discrete steps between remembered and actual grades, say $w(\tilde{s}_t, s_t)$. For example, $w(G,S) = w(S,G) = w(E,G)$. That means $c(.)$ can be fully characterised by three parameters, γ_j for $j \in \{1,2,3\}$, measuring the range of possible absolute distances of actual grades from each chosen alternative. For example, if a student got G on their last report card, the utility associated with remembering E is

$$U_{E,t} = \lambda \tilde{\theta}(E, D_{t-1}) + \gamma_1 + \varepsilon_{E,t}. \tag{8}$$

Equation (8) reiterates that actual grades are assumed to affect only the cognitive cost of memory distortion, not the impact on self-confidence, as suggested by the reduced form analysis.

The self-confidence term $\tilde{\theta}(.)$ needs to be defined at all possible choice alternatives $\tilde{s}_t \in \{E = Excellent, G = \text{Good}, S = \text{Satisfactory}, U = \text{Unsatisfactory}\}$, including counterfactual ones. The survey indicator of self-confidence observed in period t, D_t, is a function of actual memories in period t (see Table IV for coding).

[17] Empirical tests vindicate this decision in that other variables, such as students' expectations of what grade they will receive, do not play a statistical role in belief updating.

As such, identification of $\tilde{\theta}(.)$ across all choice alternatives requires some additional steps. The first of these is to define four separate latent variables, $\tilde{\theta}_{z,t} = \tilde{\theta}(z, D_{t-1})$ for $z \in \{E, G, S, U\}$, representing the four possible paths for latent self-confidence $\tilde{\theta}$ conditional on the survey indicator of prior self-confidence, D_{t-1},

$$\begin{aligned} \tilde{\theta}_{z,t} &= \beta_1^z + \sum_{r=2}^{5} \beta_r^z \mathbb{1}(D_{t-1} = r) + \varepsilon_t^{\tilde{\theta}}; \\ \varepsilon_t^{\tilde{\theta}} &\sim N(0, \sigma_{\tilde{\theta}}^2); \\ z &\in \{E, G, S, U\}. \end{aligned} \tag{9}$$

The four equations for $\tilde{\theta}_{z,t}$, outlined in Equation (9), provide a complete description of how students' mean beliefs about their ability update sequentially given their perceptions of the latest grades. Crucially, $\beta_r^z, r \in \{1,2,3,4,5\}$ all depend on the choice of grade recall, z. Thus, recall choice affects $\tilde{\theta}_{z,t}$ because it has both an additive effect on the level of self-confidence (the average of β_r^z for a given z) and a multiplicative one with prior self-confidence (the relative sizes of β_r^z for a given z). This component of the model is fundamental, as it defines how the benefit of remembering a grade differs across the possible grade memories and thus forms the basis for estimating λ in the utility function. In simpler terms, it implies that the four potential levels of self-confidence are estimated using the empirical relationship between self-confidence and grade memories. It thus forms an analogue to Equation (6), with the main difference being that I drop the redundant covariates (actual and expected grades) to ease the computational load.

Since D_t is a discrete indicator of $\tilde{\theta}_{z,t}$, I can then define the observational rule for $\tilde{\theta}_{z,t}$ as in Equation (10), where I_t^z is shorthand for $\mathbb{1}(\tilde{s}_t = z)$, a dummy variable for each possible actual grade. Equation (10) imposes that D_t follows an ordered logistic distribution conditional on $\tilde{\theta}_{z,t}$, where $\tilde{s}_t = z$. Note that I restrict the linear coefficient on all four latent variables to take the same value, ζ_1. This imposes the restriction that the survey indicator has the same relationship with underlying ability regardless of which grade memory is chosen,[18]

$$\begin{aligned} x_t'\zeta &= \zeta_0 + \zeta_1(I_t^E \tilde{\theta}_{E,t} + I_t^G \tilde{\theta}_{G,t} + I_t^S \tilde{\theta}_{S,t} + I_t^U \tilde{\theta}_{U,t}); \\ P(D_t = k) &= \frac{e^{\tau_k - x_t'\zeta}}{1 + e^{\tau_k - x_t'\zeta}} - \frac{e^{\tau_{k-1} - x_t'\zeta}}{1 + e^{\tau_{k-1} - x_t'\zeta}}; \\ k &\in \{1,2,3,4\}. \end{aligned} \tag{10}$$

Each latent variable $\tilde{\theta}_{z,t}$ enters the relevant utility outcome $U_{z,t}$, facilitating computation of λ, the

[18]The threshold parameters τ_k can be estimated internally as part of the maximum likelihood procedure.

marginal impact of self-confidence on utility, after integrating over the distribution of $\varepsilon_t^{\tilde{\theta}}$. Importantly, since $\tilde{\theta}_{z,t}$ are constructed using the empirical relationship between D_{it}, $\tilde{s}_t$, and D_{t-1}, the presence of $\tilde{\theta}_{z,t}$ in $U_{z,t}$ now introduces the additional assumption that students also subconsciously recognise these empirical relationships when trading off between possible memory choices. That is, a student's subconscious is aware of how self-confidence would likely evolve if each of the possible grades were chosen and selects grade memories on that basis, maximising utility as defined in Equation (11),

$$U_{z,t} = \lambda\,\tilde{\theta}_{z,t} + \sum_{j=1}^{3} \gamma_j \mathbb{1}(w(z,s_{it}) = j) + \varepsilon_{z,t}. \tag{11}$$

As emphasised, the fundamental aim of the model is to decompose memory choice probabilities into two competing incentives. The first is the ego benefit of remembering a higher grade. The second is a "reality constraint" in the sense outlined by Bénabou and Tirole (2002): remembering a different grade from that actually received incurs cognitive and practical costs. Having transmuted Equation (7) into its estimable form in Equation (11), it is clear that identification of these two factors relies on two essential features of the model. The first is that actual grades only affect the cost of memory distortion, not the benefit. The second is that the benefit of higher recalled grades is their effect on the sequential updating of self-confidence. λ is then measured as the correlation between choice probabilities and those potential marginal impacts on self-confidence: if grades with a greater partial impact on self-confidence are chosen more often, then λ will be greater than 0. As will be seen in Section IV.D, this is true to varying extents for students in the sample.

Finally, the joint likelihood for each observation can be written as

$$L(\tilde{s}_t = z, D_t = k|x_t) = \int_{\varepsilon_t^{\tilde{\theta}}} \frac{e^{V_{z,t}}}{\sum_{j\in\{E,G,S,U\}} e^{V_{jt}}} \left(\frac{e^{\tau_k - x_t'\zeta}}{1+e^{\tau_k - x_t'\zeta}} - \frac{e^{\tau_{k-1} - x_t'\zeta}}{1+e^{\tau_{k-1} - x_t'\zeta}} \right) d\varepsilon_t^{\tilde{\theta}}. \tag{12}$$

The model is estimated using numerical integration and maximum likelihood procedures in Apollo (Hess and Palma [2019]). I also make use of Apollo's built-in functionality for calculating standard errors clustered at the individual level, estimated using the sandwich estimator outlined in Huber et al. (1967).

IV.C Results for the Basic Model

Parameter estimates for the basic model can be located in Table VIII. Before interpreting the parameters of primary interest, λ and γ_j, it is important to establish that the latent potential self-confidence measures

$\tilde{\theta}_{z,t}$ are being plausibly estimated. As in Equation (9), $\beta_{r,z}$ are the relevant parameters. In Table VIII, I report cumulative coefficients for ease of interpretation. That is, I report $\beta_{1,z} + \beta_{r,z}$ for $r > 1$.[19] This sum is linearly related to the expected level of self-confidence $\tilde{\theta}(\tilde{s}_t = z, D_{t-1} = r)$ and thus permits statements to be made with relative ease about how different grade memory choices z affect potential self-confidence at various prior values for self-confidence r.

The significant degree of persistence in self-confidence, mirroring Section III.C, is visible in the fact that $\beta_{r,z}$ materially increases in r for each given z. That is, starting self-confidence $D_{t-1} = r$ has a major impact on $\tilde{\theta}_{z,t}$ and thus D_t, and one that is large compared to the impact of the remembered grade z for any given r. Set against this persistence is the impact of remembered grades, visible from a comparison of $\beta_{1,z} + \beta_{r,z}$ across z for each given $r > 1$. Regardless of prior self-confidence, grade memories have a monotonically positive relationship with self-confidence. However, the marginal impact differs across starting values of self-confidence. For instance, remembering a failing grade has a substantially larger negative impact on self-confidence at the lowest and highest prior levels ($r = 1, 5$) than the intermediate ones. On the other hand, the marginal impact of remembering *Excellent* relative to *Good* is generally small. However, it is almost 0 for the lowest level of prior self-confidence and the third lowest ($r = 1, 3$), and a little larger for other levels. The coefficients also map, albeit non-linearly, to the discrete categories for the recorded variable D_t. For instance, the impact on $\tilde{\theta}$ of remembering U rather than E ranges from 2–3 depending on prior self-confidence. From Equation (9), it can be seen that this $\tilde{\theta}$ maps to D_t via the linear coefficient η_1 and then the threshold parameters τ_k^{θ}. These indicate that $\tilde{\theta}$ need increase by around 5 ($3/0.6$) units to generate a higher predicted response category (or less than half that to pass from the second-highest to the highest category). Remembering a failing grade rather than an *Excellent* one is roughly sufficient to reduce self-confidence from the highest to the second-highest category in a single period.

In short, the belief updating parameters suggest strong persistence in self-confidence, but that grade memories also have a significant effect on its evolution. They also indicate that the effect of grade memories depends non-linearly on prior self-confidence. This justifies the choice to allow all belief updating parameters to update across choices of grade memory, and establishes that the latent quantities $\tilde{\theta}_{z,t}$ are being plausibly estimated. Most importantly, as desired, $E(\tilde{\theta}_{Et}) > E(\tilde{\theta}_{Gt}) > E(\tilde{\theta}_{St}) > E(\tilde{\theta}_{Ut})$ for any given value of D_{t-1}. This conclusion is crucial in lending meaning to λ. Since variation in $\tilde{\theta}_{z,t}$ over $z \in \{E, G, S, U\}$ captures the potential increase in self-confidence corresponding to remembering a higher grade, λ measures

[19] For *Excellent*, $\beta_{1,E}$ is constrained to 0.

the partial effect of those potential increases on memory choice probabilities. λ is positive and significant: students more often have grade memories which enhance self-confidence than those which do not.

Finally, γ_j for $j \in \{1,2,3\}$ have both the expected sign and relative magnitude. That is, increased memory distortions are estimated to impose larger costs to the student. As previously emphasised, these parameters capture the "reality constraints" faced by the pupil: empirically, they reflect how much actual grades matter for memory choices. Notably, the marginal cost of memory distortions appears to be stable over increasing distortion sizes. A memory distortion of distance 1 (say, from an actual *Satisfactory* grade to a memory of a *Good* one) has a cost measured by γ_1, about 1 util. A memory distortion of distance 2 adds a further cost of around 1.4, and another 1.5 is added at the maximum distance of 3.

IV.D Parameter Heterogeneity

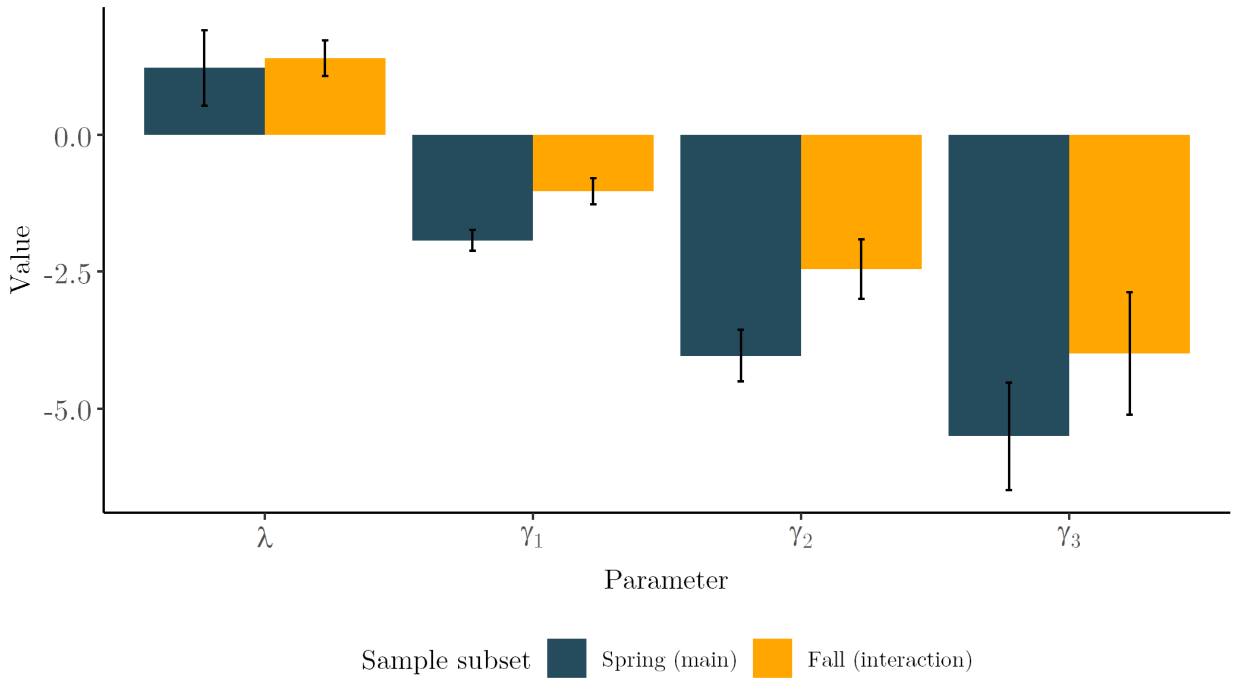
Having now established that the model produces plausible estimates of the parameters of interest, I can turn to a particularly interesting set of extensions: in allowing parameters to be heterogeneous across observable variables, the model provides deep insight into the source of differences in recall behaviour across students and editions of the survey. For simplicity and computational ease, heterogeneity is only permitted for the main parameters of interest, λ and γ_j; the belief updating parameters remain homogeneous across the sample. In order to generate heterogeneity while permitting statistical comparisons across groups, I estimate the model using interaction terms between the relevant elements of $U_{z,t}$ and various covariates of interest.

The first source of heterogeneity I consider is not across students, but the Spring and Fall editions of the survey. Recalling the analysis in Section III.B, it has already been established that memory distortions are accentuated in Fall compared to Spring, which is reassuring given the greater time delay (3 months or more) for the former version of the survey. However, the structural model permits a deeper interpretation of the motivations for this difference.

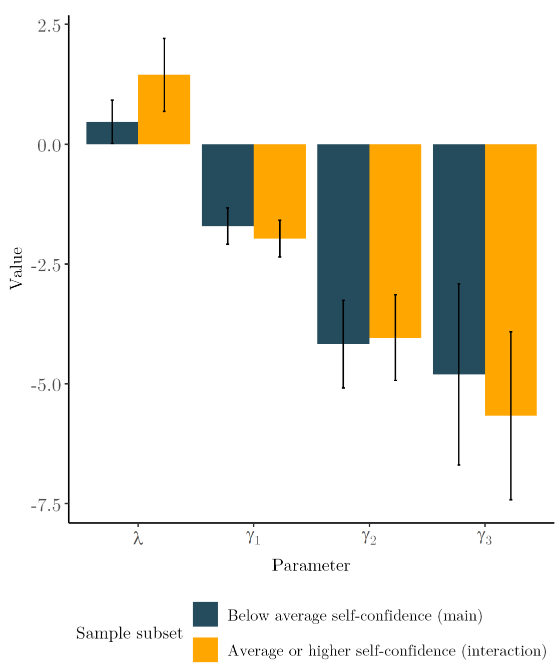
Figure III and Table IX indicate no evidence that λ is statistically different across Spring and Fall. Conversely, the differences in γ_j are statistically and practically significant, indicating that lower recall in Fall is driven by a considerable reduction in the costs of memory distortion. This is intuitive: while there should be no reason to expect that students value self-confidence more across editions of the survey administered with different time delays, it would be reasonable to expect that the measured costs of memory distortion should be smaller when a greater time period has elapsed. This result helps greatly to characterise

Figure III: Utility parameter heterogeneity

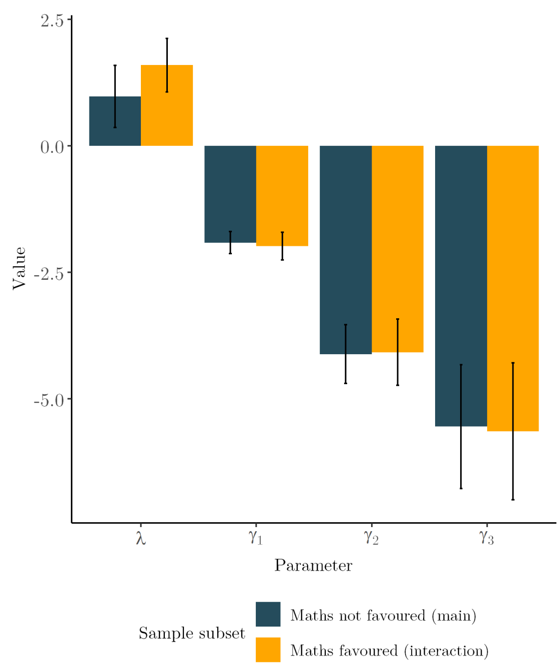
(a) Spring vs Fall



(b) Prior self-confidence



(c) Preference for maths



Error bars depict 95% confidence intervals. Interaction term (orange) confidence intervals are for comparisons relative to main group (navy), while main effect confidence intervals are for differences from 0. λ is the marginal utility of self-confidence; γ_1, γ_2 and γ_3 are the costs of recall distortions of sizes 1, 2 and 3 respectively.

the results in Section III.B. While there is a major upward drift in memory errors increase in Fall relative to Spring, that self-enhancing drift is not the result of a greater demand for self-confidence when more time has elapsed, but simply a weaker pull towards reality acting against the same demand for self-confidence. While all of the cost parameters are smaller in magnitude in Fall, an examination of their relative magnitude unravels the conclusion obtained from the whole sample (Table VIII) that costs increase steadily with the size of the memory distortion. Rather, it appears that the pattern of marginal costs is non-linear in both editions of the survey, but in opposing directions. In Spring, the largest marginal costs appear to be incurred by distortions of distances 1 and 2 (2 units each). In Fall, there is relatively little marginal cost of distorting by distance 1, but a greater cost from further distortions. This appears to indicate that for recent memories (in the Spring edition), the most strenuous effort is exerted in creating any memory distortion at all. For more distant memories (in the Fall edition), small memory distortions become much less strenuous, but larger ones are more costly.

Having established that the model's estimates plausibly reflect the difference in memory choice between Fall and Spring, I now proceed to estimate separate utility parameters for a multidimensional vector of demographic covariates of interest. Namely, while controlling for the Fall dummy variable, I also include dummies for sex (male = 1), race (black = 1), and parental education (1 if the responding parent finished high school). While the conclusion that the Fall survey results in different structural parameters is maintained in this specification, as one might expect since demographic characteristics are balanced across editions of the survey, there appears to be no evidence that the model's structural parameters vary significantly across demographic factors (Table A1 in Appendix A). While a lack of a gender gap may be particularly surprising, it matches the reduced form finding in Huffman et al. (2019) relating to gender differences in overconfidence.

Another dimension of interest is whether or not students' preferred subject is mathematics. In Section III, I demonstrated that this variable is linked to a greater bias in memory. Figure IIIc and Table IX indicates strong evidence that (in contrast to the effects of the greater time delay between Spring and Fall), the difference in the reduced form is driven by a substantially higher λ (by about 60%) for students who previously stated that their favourite subject was maths. That is, if students express a greater preference for maths, they value the confidence-enhancing effect of remembering better grades much more than other students. Meanwhile, the parameter estimates for γ_j barely differ across the two groups of students. These two results jointly have a highly intuitive interpretation. They indicate that when making the implicit choice of how much to distort their memories, students with a greater preference for maths are more concerned with the impact of memories of maths grades on their self-confidence, but are still equally sensitive to the set of reality constraints.

The final dimension of heterogeneity I consider permits a slightly closer test of motivated beliefs theory against its competitors. In both psychology (Kwang and Swann Jr [2010]) and economics (Bénabou and Tirole [2002]; Rabin and Schrag 1999), the question of whether individuals seek to enhance or reinforce their self-beliefs is somewhat unresolved. Confirmation bias theory would imply a negative marginal utility of self-confidence ($\lambda < 0$) for individuals whose self-confidence is already low. Meanwhile, models of motivated beliefs generally suggest that individuals have an interest in enhancing their egos regardless of the circumstances ($\lambda > 0$), albeit possibly to varying extents. While the analysis in Section III has already suggested the latter is true in this context, I can strengthen that conclusion by checking how the structural parameters vary along prior self-confidence. I do so by exploring whether the parameters of the utility function are different when $\mathbb{1}(D_{t-1} < 3)$; that is, when students' previous answers to the self-confidence question fell in one of the lowest two discrete categories. Once again, the only parameter affected is λ rather than γ_j. In this case, the effect of low prior self-confidence is to reduce the magnitude of λ by around $2/3$. This suggests that the confidence-enhancing motive for memory distortion applies much more feebly to students whose self-confidence is already low. However, these students do not exhibit a preference for grade memories which negatively enhance their self-confidence, as would be predicted by confirmation bias.

IV.E Structural Model: Summary

The model's parsimony lends itself to a rather transparent parameter interpretation. The cost parameters γ_j identify the dependence of memories on actual grades, while λ measures the responsiveness of memory choices to their estimated impact on ability beliefs. It is reassuring that the parameter estimates comply with and deepen the narrative underlying the results in Section III, which are able to rely on an identification strategy more robust to individual-level heterogeneity. Further reassurance is provided by the invariance of the reduced form results to the omission of individual fixed effects.

The model decomposes the decision to distort memory into a symmetric cost component, pulling memories towards reality, and an ego-enhancing motive to remember better grades than were actually achieved. The significance of both factors in predicting choice probabilities vindicates the conceptualisation of imperfect recall in the BSS as a subconsciously strategic tool rather than an error. Notably, the presence of increasing memory distortion costs implies that students do deploy memory distortion selectively rather than uniformly. Investigations of parameter heterogeneity further validate the conceptual framework, demonstrating firstly that the estimated costs of memory distortions are substantially reduced after a longer time delay, but the demand for self-confidence is unaffected. They also indicate that differences in student characteristics may beget differences in the subjective benefits of information distortion. Students with a greater prior preference for maths are much more sensitive to the self-confidence implications of memory choices, while those with low prior self-confidence are substantially less sensitive to them. This form of heterogeneity may have durable impacts on beliefs and outcomes if, as indicated by the analysis in this section and Section III.C, memories impact self-confidence more than actual grades.

V Discussion and Conclusion

This paper finds support for the idea that individuals distort incoming information to protect their egos, and that these distortions are enabled by biased memory loss. Both findings have implications for understandings of information transmission and retention, especially when that information concerns the self. Principally, they indicate that in contexts where negative signals convey socially valuable information, additional effort may be required to ensure that they are properly digested and retained. The role I find for biased memory loss, matching Zimmermann (2020) in the laboratory, is particularly notable, since it highlights that measuring the instantaneous absorption of information is inadequate in understanding long-term effects on beliefs. It also suggests that greater effort may be required to bolster the retention of negative signals as time passes. In finding stronger empirical support for motivated beliefs than models of confirmation bias, my analysis also contributes to the longer-standing debate in psychology on whether the "self-enhancing" or "self-verifying" motive for signal distortions is dominant, and in what circumstances (Kwang and Swann Jr [2010]).

A perennial concern with the use of survey data is that it is generally impossible to establish that reporting is truthful. Since working memory and its interactions with the psyche are inherently subliminal, incentivisation (as deployed in Huffman et al. [2019]) does not provide a panacea: it could produce a misleadingly accurate recall measure, since students would exert more effort in trying to remember their grades than in quotidian settings. However, two pieces of evidence provide reassurance that dishonesty is unlikely to play a major role in the results. The first is that biases in memory loss increase greatly in the Fall edition of the survey, which was administered after a longer delay. This allows me to draw the conclusion that a

Table VIII: Parameter estimates for discrete choice model

	Estimate
Self-confidence motive	
λ	1.273^{***} (0.349)
Reality constraints	
γ_1	-1.544^{***} (0.068)
γ_2	-3.253^{***} (0.162)
γ_3	-4.816^{***} (0.347)
Belief updating	
$\beta_{1,G}$	-0.015 (1.106)
$\beta_{1,S}$	-0.789 (1.119)
$\beta_{1,U}$	-2.761^{*} (1.449)
$\beta_{2,E}$	1.399 (1.501)
$\beta_{1,G}+\beta_{2,G}$	1.132 (1.459)
$\beta_{1,S}+\beta_{2,S}$	0.455 (1.293)
$\beta_{1,U}+\beta_{2,U}$	-0.262 (1.107)
$\beta_{3,E}$	3.071^{*} (1.604)
$\beta_{1,G}+\beta_{3,G}$	3.070^{*} (1.612)
$\beta_{1,S}+\beta_{3,S}$	2.605^{*} (1.459)
$\beta_{1,U}+\beta_{3,U}$	1.413 (1.415)
$\beta_{4,E}$	5.067^{***} (1.962)
$\beta_{1,G}+\beta_{4,G}$	4.780^{**} (1.958)
$\beta_{1,S}+\beta_{4,S}$	4.127^{**} (1.832)
$\beta_{1,U}+\beta_{4,U}$	2.938 (1.845)
$\beta_{5,E}$	5.945^{***} (2.214)
$\beta_{1,G}+\beta_{5,G}$	5.653^{**} (2.196)
$\beta_{1,S}+\beta_{5,S}$	5.121^{**} (2.123)
$\beta_{1,U}+\beta_{5,U}$	3.598^{*} (2.103)
$\tilde{\theta}(.)$ to D_t	
ζ_0	-0.422 (2.449)
ζ_1	0.610^{***} (0.154)
Thresholds for D_t	
τ_1^{θ}	-4.393^{**} (2.187)
τ_2^{θ}	-1.314 (1.466)
τ_3^{θ}	2.393^{**} (0.936)
τ_4^{θ}	3.736^{**} (1.461)
Observations	1,682

Note: *p<0.1; **p<0.05; ***p<0.01. Parameter estimates generated by numerical maximisation. Standard errors in brackets, using the sandwich estimator (Huber et al. [1967]).

Table IX: Parameter estimates for discrete choice model: Sample heterogeneity

	Fall vs Spring	Subject preference	Low self-confidence
Self-confidence motive			
λ	1.226***	0.973***	0.464**
	(0.351)	(0.313)	(0.230)
λ_{fall^+}	0.178	0.213	0.120
	(0.165)	(0.172)	(0.150)
λ_{subm^+}		0.620**	
		(0.269)	
$\lambda_{\mathbb{1}(D_{t-1}<3)^+}$			0.983**
			(0.388)
Reality constraints			
γ_1	−1.925***	−1.915***	−1.710***
	(0.097)	(0.112)	(0.193)
γ_2	−4.032***	−4.117***	−4.172***
	(0.239)	(0.296)	(0.465)
γ_3	−5.506***	−5.548***	−4.806***
	(0.501)	(0.625)	(0.965)
$\gamma_{1,fall^+}$	0.897***	0.905***	0.922***
	(0.124)	(0.125)	(0.123)
$\gamma_{2,fall^+}$	1.577***	1.591***	1.615***
	(0.277)	(0.283)	(0.263)
$\gamma_{3,fall^+}$	1.515***	1.541**	1.566***
	(0.570)	(0.604)	(0.554)
$\gamma_{1,subm^+}$		−0.066	
		(0.140)	
$\gamma_{2,subm^+}$		0.037	
		(0.334)	
$\gamma_{3,subm^+}$		−0.092	
		(0.690)	
$\gamma_{1,\mathbb{1}(D_{t-1}<3)^+}$			−0.259
			(0.196)
$\gamma_{2,\mathbb{1}(D_{t-1}<3)^+}$			0.136
			(0.457)
$\gamma_{3,\mathbb{1}(D_{t-1}<3)^+}$			−0.861
			(0.895)
Observations	1,680	1,680	1,680

Note: *p<0.1; **p<0.05; ***p<0.01. Parameter estimates generated by numerical maximisation. Standard errors in brackets, using the sandwich estimator (Huber et al. [1967]). Only λ and γ_j are allowed to vary; belief updating coefficients omitted for brevity. Standard errors and significance stars for subsample x^+ coefficients reflect differences relative to the omitted category. *subm* = student's favourite subject was maths in previous survey. $\mathbb{1}(D_{t-1} < 3)$ = low self confidence in previous survey.

large part of positively biased recall is enabled by genuine memory loss as time passes after information transmission: $2/3$ of the positive bias in recall errors in Fall is related to the decay of information in roughly the second and third months after receipt. Interestingly, the structural estimates vindicate this finding, and go further in suggesting that much of the effect of time elapsing seems to be in eroding the relative cost of smaller memory distortions.[20] The second piece of supporting evidence is that reported grade memories appear to have a material impact on students' sequential updating of self-confidence, which is highly persistent through time.

Recall accuracy of report card grades in the BSS is generally poor, which is unsurprising given the relatively acute challenges faced by most sample members in their adolescent years.[21] Considering that almost 30% of the sample dropped out of high school, recalling exact quarterly grades received in a specific subject likely presented particular issues for many students, whose default option may even have been to not pay attention to those grades in the first instance. Indeed, the fact that evidence of motivated beliefs is still strong in the Spring edition of the survey, which was generally administered a few weeks after the relevant report cards, leaves open the possibility that some component of the results is explained by selective attention or information avoidance. This large degree of information loss provides an appropriate setting for a study of the behavioural motivations underlying it, as well as heterogeneity in those motivations across individuals. Even if measured recall of grades is more accurate in settings where the mechanisms for information transmission are stronger, my results likely contain insight on how adolescents with varying levels of self-confidence may weight positive and negative attainment information. This phenomenon could materially complicate the potential impact of informational interventions like those considered by Dizon-Ross (2019), in that signals may differently affect children with varying tendencies for cognitive self-flattery. Indeed, Dizon-Ross does find evidence that parents are more responsive to positive information than negative, and speculates that this tendency could have impacted the average treatment effect of her intervention on enrollment. Arguably more novel is the evidence of motivated memory loss, which suggests that feedback interventions in education would do well to focus as much on later reiteration of feedback as on how information is transmitted in the moment.

The BSS contains a much broader array of variables than are considered in this paper. For instance, the survey collected information on parents' and teachers' perceptions of sample members' academic abilities,

[20] Additional analysis also suggests that children's recall accuracy does not depend on interviewer sex or race.

[21] Nonetheless, recall appears to be much better than in Dizon-Ross (2019), for a low-income setting in Malawi.

and also elicited parents' memories of their children's report card grades. The breadth of the data may permit further research on the intergenerational transmission of beliefs and preferences for motivated memory loss, among other topics.

References

Alan, S., Boneva, T., & Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, *134*(3), 1121–1162.

Alexander, K., & Entwisle, D. (2003). *The Beginning School Study, 1982-2002*. https://doi.org/10.7910/DVN/NYYXIO

Alexander, K., Entwisle, D., & Olson, L. (2014). *The long shadow: Family background, disadvantaged urban youth, and the transition to adulthood*. Russell Sage Foundation. http://www.jstor.org/stable/10.7758/9781610448239

Aydogan, I., Baillon, A., Kemel, E., & Li, C. (2017). Signal perception and belief updating.

Bartolucci, F., & Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. *Econometrica*, *78*(2), 719–733.

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

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S. et al. (2002). Hybrid choice models: Progress and challenges. *Marketing Letters*, *13*(3), 163–175.

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

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal expectations. *American Economic Review*, *95*(4), 1092–1118.

Bursztyn, L., Egorov, G., & Jensen, R. (2019). Cool to be smart or smart to be cool? Understanding peer pressure in education. *The Review of Economic Studies*, *86*(4), 1487–1526.

Cheng, H., & Furnham, A. (2002). Personality, peer relations, and self-confidence as predictors of happiness and loneliness. *Journal of adolescence*, *25*(3), 327–339.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, *22*(2), 369–395.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47.

Cunha, F., Heckman, J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78*(3), 883–931.

De Bondt, W. F., & Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in Operations Research and Management Science*, *9*, 385–410.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review*, *109*(8), 2728–65.

Dweck, C. S. (2002). The development of ability conceptions. *Development of Achievement Motivation* (57–88). Elsevier.

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–38.

Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, *80*(3), 532–545.

Falk, A., Kosse, F., Schildberg-Hörisch, H., & Zimmermann, F. (2020). Self-assessment: The role of the social environment.

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, *96*(4), 1043–1068.

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, *55*(1), 96–135.

Heckman, J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, *24*(3), 411–482.

Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, *32*, 100170.

Huber, P. J. et al. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *1*(1), 221–233.

Huffman, D., Raymond, C., & Shvets, J. (2019). Persistent overconfidence and biased memory: Evidence from managers. *Pittsburgh: University of Pittsburgh*.

Kinsler, J., Pavan, R. et al. (2016). Parental beliefs and investment in children: The distortionary impact of schools. *University of Chicago, Human Capital and Economic Opportunity Working Group*.

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

Kwang, T., & Swann Jr, W. B. (2010). Do people embrace praise even when they feel unworthy? A review of critical tests of self-enhancement versus self-verification. *Personality and social psychology review*, *14*(3), 263–280.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*(2), 153–174.

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

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2014). Managing self-confidence.

Oster, E., Shoulson, I., & Dorsey, E. (2013). Optimal expectations and limited medical testing: Evidence from Huntington disease. *American Economic Review*, *103*(2), 804–30.

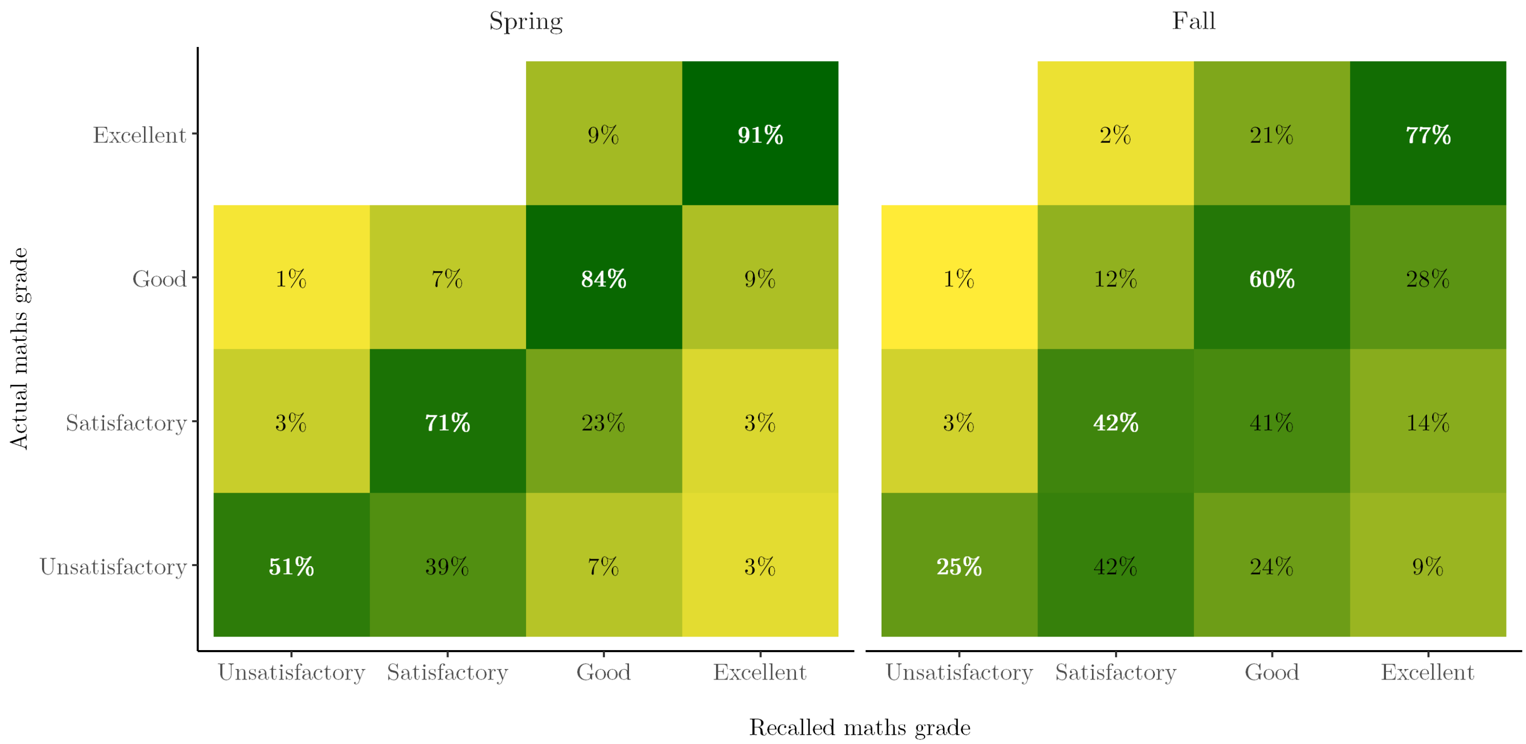
Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, *114*(1), 37–82.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta psychologica*, *47*(2), 143–148.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, *110*(2), 337–61.

A Additional Figures and Tables

Figure A4: Actual and remembered grades, split by survey edition



Note: Aggregated over survey sweeps. Percentages indicate shares of recalled maths grades by actual grade (row-wise).

Table A1: Parameter estimates for discrete choice model: demographic heterogeneity

	Estimate
Marginal value of self-confidence	
λ	1.019*** (0.368)
λ_{fall^+}	0.163 (0.166)
λ_{male^+}	0.069 (0.178)
λ_{black^+}	0.325 (0.224)
λ_{pdrop^+}	−0.099 (0.192)
Reality constraints	
γ_1	−1.856*** (0.171)
γ_2	−3.887*** (0.384)
γ_3	−4.583*** (0.754)
$\gamma_{1,fall^+}$	0.905*** (0.124)
$\gamma_{2,fall^+}$	1.592*** (0.279)
$\gamma_{3,fall^+}$	1.560*** (0.573)
$\gamma_{1,male^+}$	0.020 (0.136)
$\gamma_{2,male^+}$	0.071 (0.309)
$\gamma_{3,male^+}$	−0.376 (0.639)
$\gamma_{1,black^+}$	−0.057 (0.146)
$\gamma_{2,black^+}$	−0.067 (0.330)
$\gamma_{3,black^+}$	−0.819 (0.650)
$\gamma_{1,pdrop^+}$	−0.084 (0.143)
$\gamma_{2,pdrop^+}$	−0.310 (0.311)
$\gamma_{3,pdrop^+}$	−0.674 (0.657)
Observations	1,682

Note: *p<0.1; **p<0.05; ***p<0.01. Parameter estimates generated by numerical maximisation. Standard errors in brackets, using the sandwich estimator (Huber et al. [1967]). Only λ and γ_j are allowed to vary; belief updating coefficients omitted for brevity. Standard errors and significance stars for subsample x^+ coefficients reflect differences relative to the omitted category. *pdrop* = student's responding parent did not finish high school.