Self-Confidence and Motivated Memory Loss: Evidence From Schools

Vivek Roy-Chowdhury[*]

June 2024

Please click here for the latest version.

Abstract

We use data on teenagers' memories of mathematics grades to provide novel evidence on the dynamic relationship between biased recall and beliefs. Recall of grades has a strong positive bias that increases when more time has passed since the relevant report card, supporting theories of motivated beliefs and memory loss. Students are also more likely to overconfidently predict their next grade in periods where they recall an incorrectly high grade. We estimate a dynamic discrete choice model of biased recall. Parameters are consistent with the two-way relationship between beliefs and recall proposed by Kőszegi et al. (2022). Simulations suggest informational interventions can harm welfare and possibly widen attainment gaps if not coupled with measures to protect self-esteem, especially for low achievers.

JEL codes: D91, I21, J83

[*] Faculty of Economics, University of Cambridge. I am grateful to Toke Aidt, Leonardo Bursztyn, Botond Kőszegi, Kai Liu, George Loewenstein, Christopher Rauh, Julia Shvets, Jack Willis, Weilong Zhang, Florian Zimmermann, and attendees of the European Winter Meeting of the Econometric Society, the CESifo Summer Institute, the Barcelona School of Economics Summer Forum, and the ECBE conference for their useful comments.

1 Introduction

Beliefs about one's own characteristics are fundamental to a broad range of choices. However, a notable failure of the canonical model of belief updating in this domain is that it cannot accommodate the common observation that individuals are persistently overconfident. Overconfidence has economically important implications; it has been demonstrated to affect decisions in financial markets (De Bondt and Thaler, 1995), CEOs' investment allocations (Malmendier and Tate, 2005), personal health (Oster et al., 2013), and political behaviour (Ortoleva and Snowberg, 2015). A leading explanation of how overconfidence can functionally be sustained against persistent contradictory information comes from biased recall. In the context of self-image, biased recall captures the notion that favourable past information about oneself tends to come to mind more easily than unfavourable information.

This chapter exploits longitudinal data on teenagers' recall of their recent mathematics grades at school to examine the dynamic relationship between biased recall and beliefs. It therefore entails the first attempt to test whether motivated beliefs are important during schooling, a particularly sensitive time for the formation of an individual's beliefs about their ability. We begin by replicating recent empirical tests of theories of motivated beliefs with longitudinal field data: we test whether individuals recall favourable information about themselves more readily than unfavourable information; and whether a given individual is more likely to have overconfident beliefs about future performance when they incorrectly recall high recent performance.[1]

Dynamic measures of both recall and beliefs then allow us to make several contributions to the existing literature on motivated beliefs. We provide the first field evidence of biased memory loss by demonstrating that recall of grades is substantially more positively biased at a longer time horizon. Then, in a dynamic structural model, we explore how biased recall shapes the impact of shocks to attainment and self-esteem, and how these dynamics vary across unobserved types of student. It also provides a natural environment in which to indirectly test the key conjecture of 'fragile self-esteem' theory: that positive or negative

[1] See Bénabou and Tirole (2002) for a notable early theory. Related theories of belief distortion include Brunnermeier and Parker (2005) and Köszegi (2006).

beliefs can be self-sustaining through associative recall (Kőszegi et al., 2022).

The dataset we use is the Beginning School Study (BSS). The BSS was a longitudinal study of Baltimore City public schools in the 1980s, combining administrative data on attainment from schools with subjective survey data. The BSS is distinct from other comparable datasets in that it contains repeated measures of recalled grades alongside their factual counterparts.

Our results show that students in the BSS make positive (flattering) errors in recall much more often than negative (unflattering) ones. Using within-individual variation, we find that they are also much more likely to make recall errors when receiving low grades. Additionally, we find that the same student is much more likely to make an overconfident prediction of their next grade in periods where they recall a higher grade than they received. These analyses corroborate existing experimental and field evidence on the asymmetric recall of ego-relevant information and its link to overconfidence (Eil and Rao, 2011; Huffman et al., 2022). We then exploit an idiosyncrasy of the BSS to test for biased memory loss in the field, replicating Zimmermann (2020)'s study in the laboratory: we compare recall in Spring survey periods, which were about 2 months after the relevant report cards, to Fall periods which were about 4–5 months after report cards. Recall of grades is substantially more positive in Fall periods, indicating that memory loss is tilted towards ego enhancement.

We then estimate a discrete choice model which represents biased recall as a tool to optimise psychological welfare. Our parameter estimates support models of ego utility with reality-constrained belief distortion. In order to test Kőszegi et al. (2022)'s distinctive assumption that beliefs influence recall, we allow preferences for positively biased recall to depend on prior academic self-esteem. We find evidence consistent with this mechanism, and that shocks to self-esteem can be very persistent: a poor report card in mathematics negatively affects the distribution of self-esteem for around 2–3 years.

In an extended version of the model, we permit unobserved heterogeneity in both preferences and state variables. We allow *mindset*, the student's belief in the malleability of their ability, to be a predictor of unobserved type. Mindset has been posited as a key factor underpinning the formation of self-confidence in developmental psychology (Dweck, 2002) and related concepts have received attention in the economics of education (Alan et al., 2019). The model reveals substantial heterogeneity across students and finds that having a growth mindset is a strong predictor of both unobserved type and attaiment. High achieving types with a growth mindset are estimated to have disproportionately strong preferences for recalling positive grades, meaning their self-esteem is substantially less responsive to actual attainment. Additionally, we empirically quantify the fragility of belief equilibria, and show that an equilibrium with unrealistically high self-esteem is more fragile for low achievers. Further analysis suggests that this may be because high achievers find it more threatening to admit to themselves that they have low ability; even though lower recalled grades are less likely to affect their self-esteem, their enjoyment of school is more sensitive to self-esteem.

Aside from testing prominent theories of ego utility and beliefs, our analysis also contributes to research on the economics of education. A key focus in recent years has been to understand gaps in educational attainment and why they widen over schooling years (Heckman et al., 2006). One strand of this literature explores the role of self-confidence as a non-cognitive skill (Cunha and Heckman, 2007; Cunha et al., 2010; Alan et al., 2019). Our results indicate that self-deception is inextricable from self-confidence at school.

One concern with unincentivised measures of recall is that reporting may not be truthful. However, two pieces of evidence provide reassurance that dishonesty is unlikely to play a major role in the results. The first is our finding that recall biases starkly increase when beliefs are elicited after a longer delay, which cannot be rationalised by dishonesty. The second is that recalled grades have a material impact on self-esteem, grades, and wellbeing at school, all of which are highly persistent through time.

Our results have several implications for research and policy focusing on the provision of ego-relevant information, both inside and outside of education. In some circumstances, implementing realistic beliefs about ability is desirable. Under this prerogative, our results on biased memory loss demonstrate that unfavourable grades may need to be reiterated over time. However, a simulation of our dynamic model suggests that an intervention to correct recall errors has economically significant, negative effects on both academic self-esteem and attainment, at least within the relatively short horizon measured in our sample. The effects are also heterogeneous across the population: our model suggests low achievers' future attainment is more sensitive to their awareness of recent attainment. This suggests a more complicated set of consequences from policies aiming to reduce informational frictions, for example between schools, students, and parents (Dizon-Ross, 2019). Our results therefore throw into sharp relief the importance of expanding the classical perspective on information — that more is better — to account for non-standard preferences for beliefs and information.

Outline. This chapter proceeds as follows. In Section 2, we describe the dataset. In Section 3, we outline results from reduced-form analysis testing the basic predictions of motivated beliefs theory. In Section 4, we outline and estimate our dynamic structural model. In Section 5, we conclude.

2 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).[2] The BSS began in 1982, tracking a representative cohort of 790 students attending 20 Baltimore City public schools in the USA.

The backdrop of the BSS was one of striking educational disadvantage. From the beginning of the study, Baltimore was in the midst of a protracted economic decline, with profound effects on sample students' households (Alexander et al., 2014). 61% of responding parents in the sample did not finish high school, 28% were sole earners without employment, and 25% were the sole resident parent (Table 1). In spite of these dispiriting circumstances, parents were relatively optimistic about their child's educational prospects: in the first run of the survey, 98% expected their child to finish high school; only 70% eventually did so without initially dropping out.

[2] Alexander and Entwisle (2003).

Table 1: 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.[3] 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.

BSS surveys asked students 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), students were 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), students 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 2). Notwithstanding this substantial seasonal variation, students' recall accuracy was largely stable over time, other than in fourth grade (which we omit from all ensuing analysis). Imperfect recall observed in Spring suggests that there may have been ample opportunities for adolescents to avoid absorbing information from report cards.[4] 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 the analysis to follow.

The BSS contains a large set of variables that we do not use in our analysis.

[3] The sample is substantially smaller in Fall '88 and Spring '89, when the BSS was unable to track many students who had moved from the initial set of elementary schools to a range of middle schools. The study made a concerted effort to locate most missing students by the end of the academic year, but not in time to collect viable measures of recall for these two survey sweeps.

[4] 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 2: Summary statistics

Sweep	N	School grade (modal)	Mean mathematics mark (recalled)	Mean mathematics 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 mathematics 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.

Almost invariably, our reason for ignoring them is infrequency or inconsistency in how the variable was recorded. For example, the BSS also surveyed parents and teachers on their beliefs about students’ ability and, in the former case, their recall of students’ grades. While students themselves were surveyed twice a year, parents and teachers were only sampled once a year or less. Combined with an unbalanced panel of students, this means longitudinal analyses using data from parents and teachers would be very poorly powered. Thus, our focus is on students’ own recall and beliefs, other than where we allow cross-sectional variation in parents’ ‘mindset’ to predict students’ unobserved type in our structural model. Similarly, the BSS contains administrative data on report card grades, as well as measures of recall, for a range of other subjects: Reading, Writing, English, and Science. None of these subjects appears consistently through the sample, largely due to changes in the curriculum as students progress through school. Thus, all of our analysis centres on mathematics, for which both grades and recall were collected in all key periods and retained a consistent definition throughout.

Table 3: Grades and scores

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

3 Reduced form analysis

We now present some reduced form evidence consistent with motivated beliefs theory. In what follows, t denotes the survey sweep, so the subscript it denotes the last available observation of the given variable for individual i when survey t was collected. Throughout the chapter, g_{it} denotes report card grades in mathematics. r_{it} denotes recalled report card grades in mathematics. Grade recall was elicited using the following survey 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 g_{it} and r_{it} is $\{E, G, S, U\}$; see Table 3 for a translation into numerical marks.[5] The other notable variable we use later in this section is each student's expected grade in their next report card, also collected on the $\{E, G, S, U\}$ scale.

3.1 Recall errors

While theories of motivated beliefs differ on the precise motivations for belief distortions, they all predict that unfavourable information should be more unattractive to an individual with some form of ego utility.[6] Before moving to our structural analysis in Section 4, which quantifies the trade-off in distorting beliefs, we

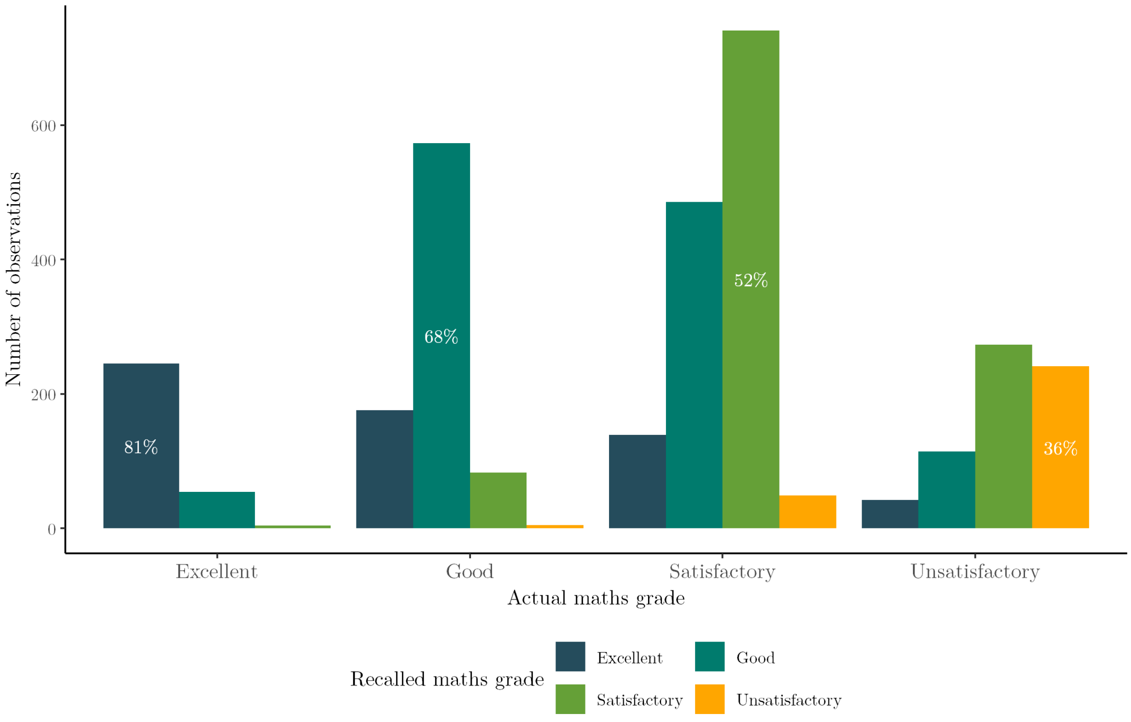
[5] Not all schools used the $\{E, G, S, U\}$ grading scale. We use a conversion table in the BSS documentation to map actual grades to that scale. Students were always asked to recall grades on the $\{E, G, S, U\}$ scale, but they were also provided with the percentage score mapping of each grade category in Table 3.

[6] For example, in Bénabou and Tirole (2002), overconfidence helps to overcome present bias, so belief distortion has 'instrumental' benefits. In Brunnermeier and Parker (2005), belief distortion has psychological benefits through anticipatory utility (Caplin and Leahy, 2001) but leads to instrumental costs by distorting actions away from their optimum.

follow most of the existing empirical literature in simply examining whether the probability of making a recall error differs across favourable and unfavourable information.

In the context of the BSS, the opportunities for information avoidance and suppression began when report cards were 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). Afterwards, biased memory loss may play a role. Both of these mechanisms are likely to be present in measured recall errors. While we examine their combined effects in this section, we demonstrate that biased memory loss plays a distinct and significant role in Section 3.2.

Figure 1: Remembered grades vs. actual grades, split by survey edition



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

Figure 1 provides a clear indication of biased recall in the BSS data. It plots actual grades received in mathematics and splits them by the grade the student remembered. The first observation is that recall errors are positively skewed, other than for the highest grade (for which upward errors are impossible): conditional on making an error, students generally remembered getting better grades than they actually did. Secondly, errors become increasingly prevalent the lower the actual grade is. Nonetheless, in all cases other than for the worst grade, the modal outcome is to recall the correct grade.

In order to formalise this analysis, we estimate a regression whose dependent variable is the event that a student incorrectly recalls their mathematics grade.[7] 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 are cognitive ability and engagement with schooling, which one might expect to be correlated with both grades and the event that they are correctly recalled. Since the BSS observed students longitudinally, we can make major progress in addressing these concerns by including individual and year fixed effects in the analysis. Intuitively, that means we examine whether the same individuals are more likely to forget lower grades than higher ones. The empirical specification is a linear probability model with robust standard errors.

As in the raw data, the estimated coefficients under column (1) in Table 4 provide clear evidence that adolescents are more likely to fail to recall poor grades. This effect is very large: students are almost 50 percentage points (pp) more likely to make recall errors when they receive the lowest grade, *Unsatisfactory* (scores lower than 70%), than when they receive *Excellent* (scores above 90%). The effect of actual grades on recall accuracy is monotonic and large: each progressively worse grade is much less accurately recalled by students. As already indicated, the probability of a recall error increases substantially (by 23.7pp) in Fall periods, which was delivered much longer after report card grades than the Spring edition. The results therefore vindicate the pattern in the raw data (Figure 1) and closely match similar patterns in experimental and field research on adults (Eil and Rao, 2011; Huffman et al., 2022).

Column (2) in Table 4 omits individual fixed effects. The results undergo lit-

[7] This dependent variable also permits negative recall errors. We 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*. We also report the results of an analysis in which we examine whether probability of making a positive recall error depends on the actual grade received, excluding observations where *Excellent* was actually received (Table A1) The results are very similar, with striking increases in the probability of making a positive recall error when receiving *Satisfactory* and *Unsatisfactory* relative to *Good*.

Table 4: Reduced form model — recall errors

	Dependent variable:	
	Incorrect recall	
	(1)	(2)
Actual grade: *Good*	0.116*** (0.040)	0.128*** (0.030)
Actual grade: *Satisfactory*	0.276*** (0.042)	0.251*** (0.029)
Actual grade: *Unsatisfactory*	0.520*** (0.047)	0.461*** (0.032)
Fall	0.237*** (0.020)	0.238*** (0.020)
Individual FE	Yes	No
Academic year FE	Yes	Yes
Observations	2,694	2,694
R^2	0.371	0.129
Adjusted R^2	0.196	0.127

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Linear probability model, robust standard errors in parentheses. All grades are from quarterly report cards for mathematics, and recall of the most recent quarterly grade is elicited in a biannual survey. The omitted actual grade category is *Excellent*.

tle qualitative change other than some small differences in relative magnitudes. This conclusion could provide some reassurance to comparable analyses in field research without longitudinal variation (Huffman et al., 2022), since it indicates that fixed unobserved individual characteristics, such as cognitive ability or engagement with schooling, may not be an important joint determinant of signal valence and the accuracy of recall. However, we will soon see that unobserved factors appear to be a more important joint determinant of biased recall and overconfidence.

We also examine sex and racial differences in recall asymmetries (Table A2). Surprisingly, there are no differences between boys and girls, matching a related analysis in Huffman et al. (2022). On the other hand, Black students appear to display a greater propensity to recall information asymmetrically than others: incorrect recall of *Good* and *Unsatisfactory* is more likely relative to *Excellent* grades than for other students. However, this pattern is uneven: *Satisfactory* grades are no less accurately recalled by Black students.

3.2 Biased memory loss

Biases in recall could result from a combination of information avoidance and memory loss. However, the design of the BSS presents a compelling opportunity to isolate the role of the latter.[8] In both versions of the survey within the year, in 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. In the raw data, we have already seen sizeable differences in the overall accuracy of recall between Spring and Fall (Table 2). The analysis in this section exploits this naturally occurring variation in survey delivery to examine whether recall is more positively biased in Fall than in Spring.

Figure 2 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 noted above, 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 the deterioration in recall accuracy is in a flattering direction; the incidence of unflattering errors is barely higher in Fall relative to Spring. By comparing the difference between flattering and unflattering errors in Fall and Spring, we can determine that at least 2/3 of the positive bias in recall in Fall is explained solely by memory loss. Assuming truthful reporting, the remaining 1/3 could be explained by either memory loss during the first month or information avoidance.[9]

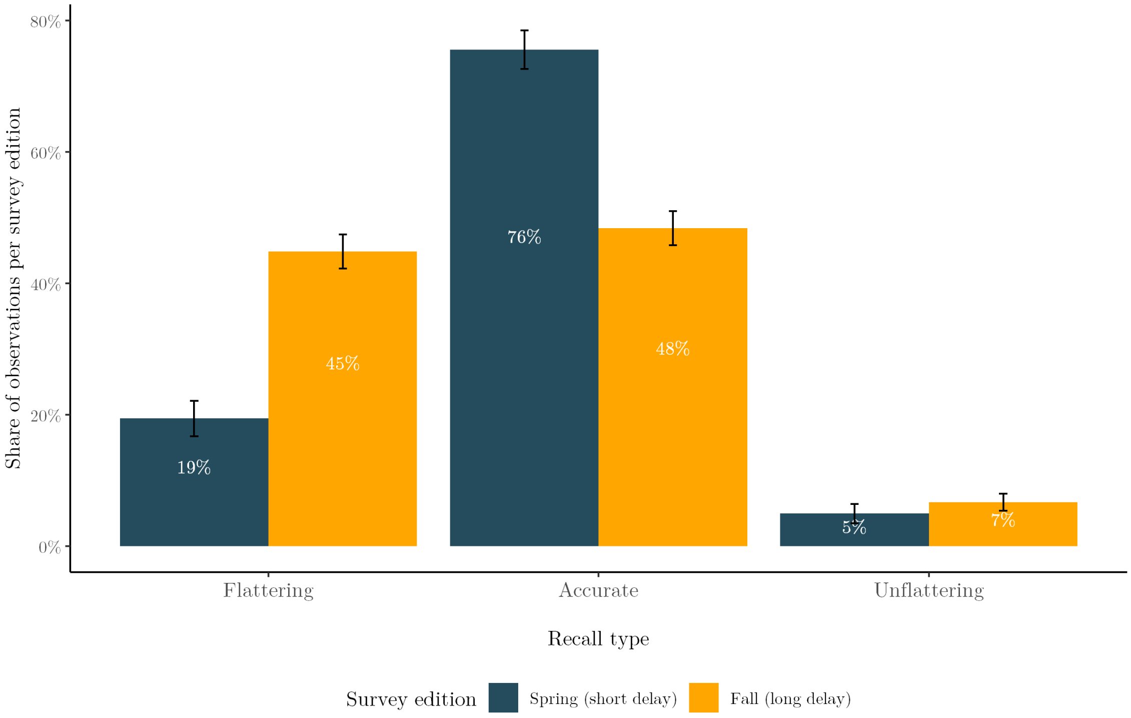
We also study biased memory loss in a simple regression, once again using within-individual variation. Our approach is to examine the event that the student recalls getting an *Excellent* grade — as the dependent variable, controlling for the actual grade they received, g_{it}.[10] The key coefficient of interest is on *Fall*:

[8] Our strategy is similar to the main analysis in Zimmermann (2020), which examines how recall of negative signals of IQ deteriorates over time relative to positive signals.

[9] Even in the Spring survey, recall is less accurate for the lower grades, *Satisfactory* and especially *Unsatisfactory* (Figure A13). 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.

[10] Another option would be to examine whether the probability of making a positive recall error increases in Fall, conditional on making an error, as in Figure 2. However, as in Section 3.1, this

Figure 2: 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.

we want to establish that the probability of recalling the top grade increases in *Fall* compared to *Spring*, for the same student.

Table 5: Reduced form model — memory loss

	Dependent variable:	
	Recalled grade: *Excellent*	
	(1)	(2)
Actual grade: *Good*	-0.598^{***} (0.035)	-0.600^{***} (0.030)
Actual grade: *Satisfactory*	-0.680^{***} (0.035)	-0.724^{***} (0.027)
Actual grade: *Unsatisfactory*	-0.691^{***} (0.037)	-0.741^{***} (0.028)
Fall	0.084^{***} (0.013)	0.084^{***} (0.013)
Individual FE	Yes	No
Academic year FE	Yes	Yes
Observations	2,694	2,694
R^2	0.371	0.129
Adjusted R^2	0.196	0.127

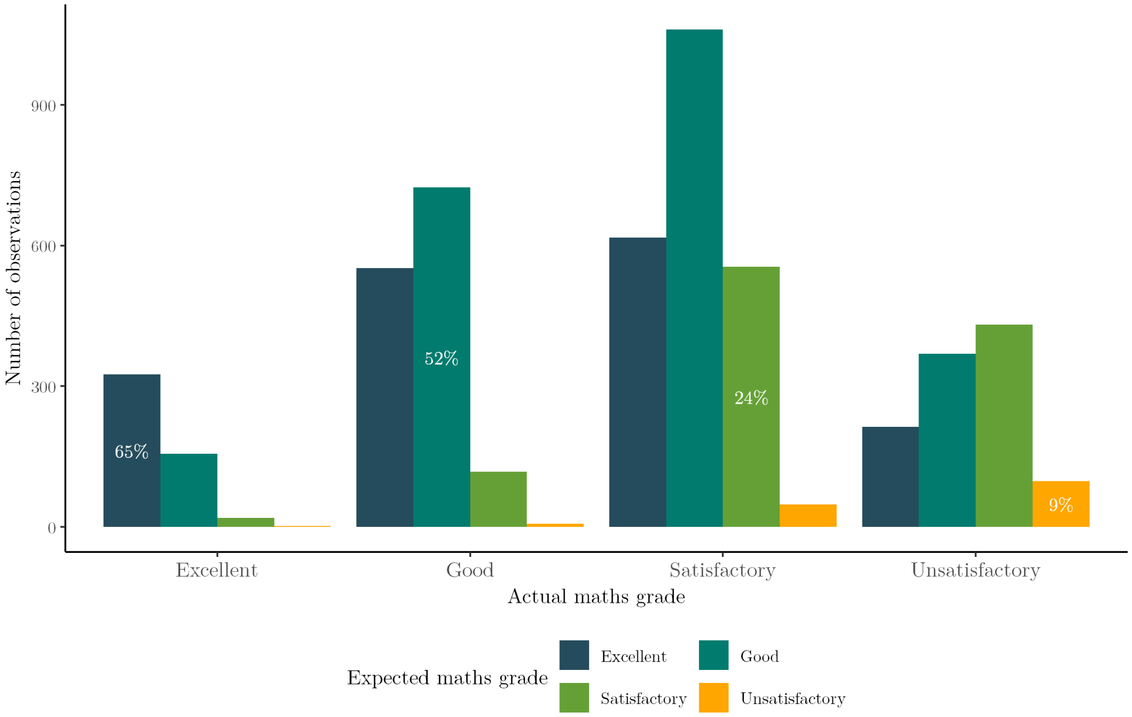
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Linear probability model, robust standard errors in parentheses. All grades are from quarterly report cards for mathematics, and recall of the most recent quarterly grade is elicited in a biannual survey. The omitted actual grade category is *Excellent*.

The results are reported in Table 5. The coefficient on Fall in column (1) implies that being surveyed in Fall has a large, positive impact on students' propensity to remember getting the highest grade. This formalises the positive drift in memory illustrated in the raw data, in Figures 2 and A13. That is, the same students are more likely to remember higher grades when surveyed in Fall than in Spring, controlling for actual grades. Thus, we obtain evidence matching the main exercise in Zimmermann (2020): the positive bias in recall becomes more accentuated as time passes, suggesting that memory loss serves the ego. Unsurprisingly, the coefficient on *Fall* does not differ across specifications with and without individual fixed effects in columns (1) and (2), since the sample should be balanced across editions of the survey.

Again, we examine heterogeneity along sex and race dimensions (Table A3). In this case, we find no evidence of differences in the effect of memory loss across

strategy only permits us to examine cases where the middle two grades, *Good* and *Satisfactory*, were received. It would involve discarding information on how memory loss affects students receiving the top and bottom grades.

Figure 3: Expected grades vs actual



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

either dimension. That being said, Black students are less likely to recall *Excellent* grades (across both Spring and Fall) having actually achieved either of the lowest two grades.

3.3 Overconfidence

In each survey period, the BSS also collected students' expectations of the mathematics grade they would receive in the following quarterly report card. By comparing this variable to actual grades, we can directly measure overconfidence. Our aim is to establish whether biased recall is associated with overconfident expectations of future grades, replicating the main exercise in Huffman et al. (2022) but with the significant benefit of within-individual variation. In doing so, we provide insight on the important matter of whether biased recall allows overconfident beliefs at school to persist despite repeated signals to the contrary.

First, we report on the raw data (Figure 3). Students frequently make errors in predicting their next grade. Prediction errors are more common than recall errors. This disparity is particularly stark for the low grades: only 9% of those receiving *Unsatisfactory* expected to do so, compared with the 36% accurately recalling *Unsatisfactory* when it is received. The differences remain stark for *Satisfactory* and *Good* grades: accurate expectations occur at rates of 24% and 52% respectively, compared with accurate recall rates of 52% and 68% respectively. Notably, expectation errors are also more common than recall errors for the upper grade, *Excellent*, suggesting that part of the difference is driven by genuine forecast uncertainty. Nonetheless, students' expectations are strikingly overconfident in general. Other than for students eventually receiving *Good* and *Excellent*, the modal outcome is to expect a higher grade than is actually received. Students also commonly expect a higher grade than they recall receiving in their last report card (Figure A12).

To establish a link between biased recall and overconfidence, we estimate a regression which uses longitudinal variation to establish whether a *given student* is more likely to have overconfident expectations of their future grade in a period where they have recalled a higher grade than was actually received.[11] Notably, a within-individual strategy naturally precludes several of the concerns that are addressed individually in Huffman et al. (2022), perhaps most importantly that overconfidence could reflect stable differences in prior beliefs (Benoît and Dubra, 2011) or cognitive ability.

The results are reported in Table 6. First, in column (1), we find that a given student is indeed more likely to be overconfident in periods where they they recall an incorrectly high grade. The effect on the probability of overconfidence is very large, at 8.2 percentage points compared to a baseline probability of 28.1% for students who recall the correct grade or one lower than they received. Notably, actual attainment has no impact on the likelihood of overconfidence, suggesting the interpretation that recall is all that matters for beliefs.

In column (2), we add a control variable for Fall survey periods. This specification is of particular interest because, as established in the last section, recall

[11] This strategy requires us to drop all observations where students receive *Excellent* in the prior or next report card. An alternative specification, which uses the whole sample, considers whether prediction errors are more likely when recall errors have been made. The results are qualitatively very similar (Table A4).

Table 6: Reduced form models — overconfidence

	Dependent variable:		
	Overconfident prediction		
	(1)	(2)	(3)
Positively biased recall	0.082*** (0.023)	0.069*** (0.024)	0.129*** (0.020)
Actual grade: *Satisfactory*	−0.022 (0.029)	−0.018 (0.029)	0.040* (0.023)
Actual grade: *Unsatisfactory*	−0.039 (0.037)	−0.030 (0.037)	0.076*** (0.028)
Fall		0.051** (0.024)	0.036 (0.024)
Individual FE	Yes	Yes	No
Academic year FE	Yes	Yes	Yes
Observations	2,473	2,473	2,473
R^2	0.341	0.342	0.041
Adjusted R^2	0.130	0.131	0.038

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Linear probability model, robust standard errors in parentheses. All grades are from quarterly report cards for mathematics, and recalled and expected quarterly grades are elicited in a biannual survey. The sample excludes all observations where the previous or next grade is *Excellent*. The omitted actual grade category is *Good*.

errors are much more likely in Fall than Spring. Interestingly, students are more likely to be overconfident in Fall even though we control for recall errors. This may reflect that the fact that in Fall, expectations become more optimistic in a way that is not captured by recall errors. One example could be that the student could perceive a weaker link between past performance and future performance when more time has passed since the last grade.[12]

In column (3), we omit individual fixed effects from the specification. The most notable effect of this change is to almost double the size of the coefficient on recall errors, relative to the comparable specification in column (2). This suggests that unobserved cross-sectional heterogeneity is an important joint determinant of biased recall and overconfidence. This comparison reveals that students persistently receiving lower actual grades are more likely to make persistent errors in predicting their future grades: those receiving *Satisfactory* grades are 4pp more likely to make errors in predicting their next grade than those receiving *Good*,

[12] Since the length of time to the next grade is approximately the same in both the Spring and Fall editions of the survey, the coefficient on Fall cannot be explained by a difference in forecast horizon.

with an even larger difference of 7.6pp for those receiving *Unsatisfactory*. Thus, controlling for cross-sectional unobserved heterogeneity plays an important role in convincingly identifying the link between biased recall and overconfidence.

4 Structural analysis

We now outline a structural model that positions recall distortions at the centre of a psychological framework to optimise welfare. It assumes students (subconsciously) maximise lifetime utility by trading off the benefits and costs of distorting their recall of grades in each period. Since the model embodies a dynamic interaction of recall and self-esteem, with links possibly running in both directions, it provides a natural environment in which to test the key mechanisms underpinning fragile self-esteem theory (Kőszegi et al., 2022).

A further attraction of the structural approach is that, in an extended version of the model, we can separate students into unobserved types with potentially differing preferences for inflating their recall of grades. In particular, we model unobserved heterogeneity as a finite mixture distribution where key parameters for both preferences and state variables differ across discrete classes of student. This approach means the model can determine whether some students have stronger preferences for positively biased recall than others, also allowing these differences to be correlated with unobserved heterogeneity in attainment and self-esteem.

4.1 Basic model

For ease of computation and interpretation, we redefine our measures of recalled and actual grades as binary rather than categorical. In this updated definition, $r_t = 1$ if the student recalls *Good* or *Excellent* in period t. g_t has the same interpretation as r_t but refers to the relevant grade actually achieved. Also key to the model is y_t, an indicator of self-esteem based on the survey measure in Table 7. As is visible from the summary statistics, the majority of students select the middle category. We therefore define our binary indicator such that $y_t = 1$ for values greater than 3; this threshold extracts the maximal amount of information from the survey indicator. This definition implies $y_t = 1$ if a student believes they are above average ability for their school and year group.

Table 7: Self-esteem measure

How smart do you think you are compared to other kids in your school this year?	Value	Share of observations
One of the smartest	5	14.7%
Smarter than most kids	4	21.0%
About as smart as everybody else	3	57%
Not as smart as most kids	2	6.5%
Not very smart at all	1	0.1%

In each period, students choose r_t to maximise expected lifetime utility,[13]

$$E \sum_{t=1}^{\infty} \beta^{t-1} u(r_t, y_{t-1}, g_t, f_t, \epsilon_t; \theta, \lambda). \tag{1}$$

In (1), f_t is an indicator for the Fall survey, and $\epsilon_t = (\epsilon_{0t}, \epsilon_{1t})$ are choice-specific shocks to utility observed by the student in period t but unobserved in the data. We make the standard assumption that they are drawn from a Type 1 Extreme Value distribution. Periodic utility has the following functional form, where $\Theta = (\theta, \theta_y)$ and $\Lambda = (\lambda, \lambda_f)$,

$$u(r_t, y_t, g_t, \epsilon_t; \Theta, \Lambda) = \begin{cases} \theta + \theta_y y_{t-1} - (\lambda \mathbb{1}(g_t \neq r_t) + \lambda_f \mathbb{1}(g_t \neq r_t) f_t) + \epsilon_{1t} & \text{for } r_t = 1 \\ -(\lambda \mathbb{1}(g_t \neq r_t) + \lambda_f \mathbb{1}(g_t \neq r_t) f_t) + \epsilon_{0t} & \text{for } r_t = 0. \end{cases} \tag{2}$$

θ and θ_y capture the hedonic benefit of recalling a higher grade, while λ and λ_f capture the cost of recall distortions. The cost of recall distortions is symmetric, in that recalling an incorrectly high or an incorrectly low grade incurs the same cost.

To understand identification of the key parameters, consider a student and survey period where g_t, y_{t-1} and f_t are all equal to 0. For such a student, the

[13] Since we only observe key variables for an intermediate part of students' school careers, and since the process of interpreting signals about ability likely continues past schooling years, we set up the problem with an infinite horizon.

marginal utility of distorting recall is $\theta - \lambda$. Thus, if positive recall errors are more common *relative to* negative recall errors, we will obtain a larger estimate of θ. If instead $g_t = 1$, the marginal utility of a negative recall error $r_t = 0$ is $-\lambda$. Thus, λ is only smaller if both negative and positive recall errors are less common. Identification therefore exploits information from all types of recall: flattering, accurate, and unflattering.

The parameter λ_f also appears directly in utility, capturing the fact that the costs of memory distortion are lower in Fall. We would expect that $\lambda_f < 0$. This specification naturally accommodates the finding that recall becomes more positively biased in Fall: for a student with low prior self-esteem $y_{t-1} = 0$ achieving a low grade $g_t = 0$, the marginal utility of $r_t = 1$ is $\theta - \lambda$ in Spring and $\theta - (\lambda + \lambda_f)$ in Fall. The marginal utility of $r_t = 1$ conditional on $g_t = 1$ is θ in both Spring and Fall. If $\lambda_f < 0$, our model can embody biased memory loss: the prospect of recalling a high grade is equally attractive in both Spring and Fall, but it should be harder to do so in Spring when relatively little time has passed since the report card in question. Finally, θ_y allows preferences for $r_t = 1$ to vary with prior self-esteem. In particular, if $\theta_y > 0$, students with high prior self-esteem have a stronger preference for self-enhancing recall distortions, capturing the key mechanism underpinning fragile self-esteem (Kőszegi et al., 2022).

Our specification of hedonic utility imposes no explicit structure on the motivations for biasing recall. An alternative approach would be to do so, perhaps by specifying that the benefits of biased recall accrue through its expected effects on self-esteem and grades. There are two main justifications for the more parsimonious approach we use. Instead of attempting to explain why students might bias their recall of grades, our model focuses on estimating the strength of preferences for positively biased recall. This has significant benefits: identification is transparent; and our model folds all possible motivations for distorting grades, including unobserved ones, into the essentially reduced form parameters θ and θ_y. Thus, we do not make any assumptions about *what* motivates students to bias their recall, freeing up the model to focus on our main question of interest: how the strength of preference for biased recall relates to the dynamics of beliefs.

The second justification is that it would be very difficult to achieve plausible identification of any indirect marginal benefits of biased recall. For instance, in a model in which periodic utility were a direct function of $\hat{y}_t = E(y_t | r_t, y_{t-1}, g_t, f_t)$

or $\hat{g}_t = E(g_t|r_t, y_{t-1}, g_{t-1}, f_t)$, identification of the marginal utility of both variables would need to exploit heterogeneity in choice probabilities across observations with different marginal effects of r_t on $\hat{y}_t$ and $\hat{g}_t$. We would not have a source of variation in the marginal effect of r_t on $\hat{y}_t$ and $\hat{g}_t$ other than any non-linearity contrived from functional form assumptions, which we do not think would be behaviourally credible.

The two state variables g_t and y_{t-1} are specified to follow first-order Markov processes,

$$P(g_t = 1|g_{t-1}, y_{t-1}, r_{t-1}) = \text{logit}(\mu_1^g + \mu_y^g y_{t-1} + \mu_g^g g_{t-1} + \mu_r^g r_{t-1}), \tag{3}$$

and

$$P(y_t = 1|g_{t-1}, y_{t-1}, r_{t-1}) = \text{logit}(\mu_0^y + \mu_y^y y_{t-1} + \mu_y^y g_t + \mu_r^y r_t). \tag{4}$$

Let $x_t = (y_{t-1}, g_t)$. We make the standard assumption that (x_t, ϵ_t) is a stationary controlled first-order Markov process, with transition

$$P(x_{t+1}, \epsilon_{t+1}|x_t, \epsilon_t, r_t) = P(\epsilon_{t+1}|x_{t+1}, r_t)P(x_{t+1}|x_t, r_t), \tag{5}$$

so that ϵ_t is serially independent conditional on x_t, and x_{t+1} is independent of ϵ_t conditional on x_t and r_t. The student observes ϵ_t in period t but not before. Under these assumptions, the model is stationary and can be estimated by full information maximum likelihood after computing the expected value function via iteration of the Bellman equation (Rust, 1987). To elaborate on this procedure in our context, let $\bar{u} = u - \epsilon$ and define the expected value function $\bar{V}(x_t, r_t)$. The value function is

$$V(x_{t+1}, \epsilon_{t+1}) = \max_{r_{t+1}\in\{0,1\}} \{\bar{u}(x_{t+1}, r_{t+1}; \Theta, \Lambda) + \epsilon_{t+1} + \beta\bar{V}(x_{t+1}, r_{t+1})\}. \tag{6}$$

By taking expectations over $(x_{t+1}, \epsilon_{t+1})$ conditional on (x_t, r_t), we obtain the Bellman equation for $\bar{V}(x_t, r_t)$, where X is the state space for x_t,

$$\bar{V}(x_t, \epsilon_t) = \sum_{x_{t+1}\in X} \log\left[\sum_{r_{t+1}\in\{0,1\}} \exp\left(\bar{u}(x_{t+1}, r_{t+1}; \Theta, \Lambda) + \beta\bar{V}(x_{t+1}, r_{t+1})\right)\right] P(x_{t+1}|x_t, r_t). \tag{7}$$

The sample log-likelihood for the basic model can be written,

$$l(\Theta, \Lambda, \mu, \omega, \gamma) = \sum_{i=1}^{N} \log \left\{ \prod_{t=1}^{T} P(r_{it} | x_{it}; \Theta, \Lambda, \mu) P(x_{it} | x_{i,t-1}; \mu) \right\}. \tag{8}$$

In all of what follows, we assume $\beta = 0.9$ since it is not identified (Rust, 1987). However, we report the results of estimation with a range of values of β in the Appendix. Computation and maximisation of the sample log-likelihood is then straightforward because choice probabilities $P(r_{it}|x_{it}; \Theta, \Lambda, \mu)$ have the well-known logistic closed form, and state transition probabilities stem from the dynamic logistic regressions defined in (4) and (3).

4.2 Unobserved heterogeneity

We also estimate a version of the model with unobserved heterogeneity modelled as a finite mixture distribution (Heckman and Singer, 1984). We use this approach to allow for heterogeneous preferences which may be correlated with unobserved determinants of state variables. We assume students can be one of two unobserved types, $c \in \{1, 2\}$.[14] The crucial feature we would like to capture is heterogeneity in the key parameters driving preferences for self-enhancing recall. As such, θ^c and θ^c_y are now permitted to vary across types, while we impose that the cost parameters are fixed across types — our interest *a priori* is not in heterogeneity in the accuracy of recall across students. Thus, periodic utility can now be written as:

$$u(r_t, y_t, g_t, c, \epsilon_t; \Theta^c, \Lambda) = \begin{cases} \theta^c + \theta^c_y y_{t-1} - (\lambda \mathbb{1}(g_t \neq r_t) + \lambda_f \mathbb{1}(g_t \neq r_t) f_t) + \epsilon_{1t} & \text{for } r_t = 1 \\ -(\lambda \mathbb{1}(g_t \neq r_t) + \lambda_f \mathbb{1}(g_t \neq r_t) f_t) + \epsilon_{0t} & \text{for } r_t = 0. \end{cases} \tag{9}$$

Since heterogeneity in preferences may be correlated with state variables, we control for type-specific fixed effects acting on both actual grades and self-esteem. We also want to be able to say if heterogeneity in preferences for recall is correlated with heterogeneity in the marginal effect of recall on state variables, so we also

[14] The model is not identified with more than two types, as is the case in other similar applications (Arcidiacono, 2005).

include the parameters ω_r^g and ω_r^y in the transition equations:

$$P(g_t = 1|g_{t-1}, y_{t-1}, r_{t-1}, c = 2; \mu_g, \omega^y, \omega_r^g) = P(g_t = 1|g_{t-1}, y_{t-1}, r_{t-1}; \mu_g) + \omega^g + \omega_r^g r_t, \tag{10}$$

and

$$P(y_t = 1|g_t, y_t, r_t, c = 2; \mu_y, \omega^y, \omega_r^y) = P(y_t = 1|g_{t-1}, y_{t-1}, r_{t-1}; \mu_y) + \omega^y + \omega_r^y r_t. \tag{11}$$

Although unobserved heterogeneity should be well identified by within-individual correlation in choices and state variables over time, we provide further information to the model and aid interpretation by permitting type membership to be predicted by two key covariates from Fall '85, about three years before the first period in which the rest of the variables are observed. A growth mindset is understood to be crucial in developing resilience to negative feedback and improving attainment (Dweck, 2002; Alan et al., 2019; Yeager et al., 2019), and Kőszegi et al. (2022) also speculate that mindset could introduce heterogeneity in self-esteem equilibria. We thus specify the following auxiliary model for unobserved type probabilities:

$$P(c = 2|X; \gamma) = \pi(X; \gamma) = \text{logit}(X'\gamma), \tag{12}$$

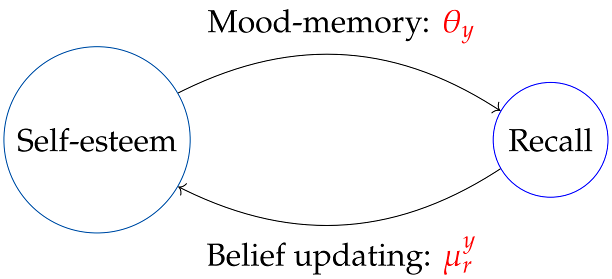
where X contains a constant term and the following variables:

1. $pfixed$: Parents' responses to the question *'For students to do well in school which is most important?'*. 'Effort/Personality' = 0; 'Ability' = 1.

2. $fixed$: Students' own responses to the question *'What do you think matters most in school, how hard you try, or how smart you are?'*. 'How hard you try' = 0; 'How smart you are' = 1.

Choice probabilities have the same logistic closed form as in the basic model. The only difference is that the sample likelihood is a weighted sum across the two types. Define $\Theta = (\theta^1, \theta^2, \theta_y^1, \theta_y^2)$, $\Lambda = (\lambda, \lambda_f)$, $\mu = (\mu_g, \mu_y)$, and $\Omega = (\omega^g, \omega_r^g, \omega^y, \omega_r^y)$. We can then write the sample log-likelihood $l(\Theta, \Lambda, \mu, \omega, \gamma)$,

Figure 4: Interaction of self-esteem and recall in FSE



where we reintroduce the individual index $i \in \{1, 2, ..., N\}$,

$$l(\Theta, \Lambda, \mu, \omega, \gamma) = \sum_{i=1}^{N} \log \left\{ \sum_{c=1}^{2} \pi(X_i; \gamma) \left[\prod_{t=1}^{T} P(r_{it} | x_{it}; \Theta, \Lambda, \mu, c) P(x_{it} | x_{i,t-1}; \mu, \Omega, c) \right] \right\}. \tag{13}$$

Consistent parameter estimates can be obtained by maximising $l(\Theta, \Lambda, \mu, \omega, \gamma)$, following computation of the value function via iteration of the Bellman equation as before.

4.3 Connection to Fragile Self-Esteem

Kőszegi et al. (2022)'s theory of 'fragile self-esteem' (henceforth FSE) hinges on a mutual interaction between beliefs and recall, outlined in Figure 4. In their framework, much as in other models of biased recall, 'mood' (utility) is determined by an agent's recall of favourable information. The novelty of FSE is in a 'mood-memory' relation that implies a relationship running in the other direction: an individual's recall bias depends directly on their mood. In particular, past outcomes are more likely to be recalled if they are concomitant with one's mood. If an individual feels positively about herself, associative memory allows her to recall memories in support of that belief. Since mood determines memory, and memory determines mood, 'self-esteem personal equilibrium' (SEP) in FSE is characterised by a fixed-point condition. Importantly, multiple equilibria may be possible, depending on the functional form of the mood-memory relationship. The authors' favoured specification is a functional form which gives rise to two SEP. The 'fragility' of each equilibrium is defined as the size of shock required to dislodge self-esteem to the other SEP.

Our data offer a novel opportunity to indirectly test the mechanisms underpinning FSE, since they offer dynamic measures of both self-esteem and recall. Recalling (4), μ_r^y captures the well-known impact of recall on self-esteem. The parameter introducing the novel mood-memory link proposed by FSE is θ_y: if mood influences memory, recalling a positive grade should be more likely when self-esteem is already high, meaning $\theta_y > 0$. Separately, the utility parameter θ, capturing students' preference for recalling favourable grades conditional on $y_{t-1} = 0$, has a close analogue in the FSE parameter k. In that model, k governs the strength of the individual's ego concern and functionally determines the extent of positive bias in the distribution of recall conditional on self-esteem.

Some judgement is required in translating FSE to an empirical setting. One particularly important point is how to characterise SEP empirically. Kőszegi et al. (2022) do not specify the speed of adjustment to equilibrium, but in our data, where each time period spans approximately 6 months, movements between self-esteem states ($y_t = 0, 1$) occur within single periods. Neither self-esteem state is absorbing in our data. State and choice variables thus form an ergodic Markov chain and the long-run distribution of self-esteem is invariant to initial conditions. In other words, starting at high or low self-esteem cannot influence the probability of having high self-esteem in the long run; in each consecutive period, it is possible to experience a shock large enough to shift from high self-esteem to low or vice-versa.

Kőszegi et al. (2022) define fragility of a self-esteem equilibrium as the smallest shock required to dislodge an individual from one equilibrium to another. On the assumption that $y_t = 0$ and $y_t = 1$ both capture equilibria, we can infer fragility from transition probabilities for y_t. To illustrate the intuition, consider the latent variable representation of y_t, where $y_t = \mathbb{1}(y_t^* > 0)$ and

$$y_t^* = \mu_0^y + \mu_y^y y_{t-1} + \mu_g^y g_t + \mu_r^y r_t + \epsilon_t^y, \tag{14}$$

where ϵ_t^y is a random variable whose distribution is logistic with location 0 and scale 1. y_t^* is thus a continuous measure of self-esteem like the one used in FSE. Now suppose that the high self-esteem equilibrium is located at $y_t^* = \bar{y}_1 > 0$. If we make the assumption that students are in equilibrium in each period, a shift to the low self-esteem equilibrium (say, at some $\bar{y}_0 \leq 0$) occurs for any shock large enough to imply $y_t^* \leq 0$. Suppose for simplicity that $g_t = 1$ and $r_t = 1$. Then, the critical value for ϵ_t^y,

$$\bar{\epsilon}^y = -\mu_0^y - \mu_y^y - \mu_g^y - \mu_r^y, \tag{15}$$

is a monotonic transformation of $1 - P(y_t = 1 | y_{t-1} = 1, g_t = 1, r_t = 1)$. Thus, a natural quantification of fragility in our setting arises from the transition probabilities $P(y_t = k | y_{t-1} = 1 - k)$ for $k \in \{0, 1\}$: they are monotonically related to the size of shock required to move students from one level of self-esteem to another.

In FSE, self-esteem is a function of an individual's recall of their entire history of ego-relevant information, which is unobserved in our setting. We only have access to recall of the most recent grades. This substantively implies that θ_y only partially identifies the mood-memory relation. If mood-memory acts through recall of any information other than the most recent grade, such a channel would be captured by μ_y^y: y_{t-1} affects y_t independently of its effect on r_t.

4.4 Results

Basic model. Parameter estimates for the basic model are reported in Table 9. State variables are strongly serially dependent, indicated by μ_g^g and μ_y^y respectively. Most important for our test of fragile self-esteem theory are the results that $\mu_y^y > 0$ and $\mu_r^y > 0$: we have that self-esteem is both serially dependent and affected by recall of ego-relevant information. However, the former significantly outweighs the latter. Notably, self-esteem does not depend on actual grades: $\mu_3^y = 0$.

Actual grades do depend on self-esteem ($\mu_y^g > 0$) and biases in recall ($\mu_r^g > 0$). This finding is particularly notable because it is consistent with belief distortion having instrumental benefits (Bénabou and Tirole, 2002; Compte and Postlewaite, 2004) rather than costs (Brunnermeier and Parker, 2005), at least within the horizon we examine. This has some relevance to a later discussion on the welfare effects of interventions to correct beliefs.

In interpreting the parameters of the utility function, notice that $\theta + \theta_y < \lambda + \lambda_f$, so even in *Fall* and when prior self-esteem is high, recalling the correct grade delivers higher periodic expected utility. This is driven by the fact that empirically, the modal behaviour is to recall the correct grade. However, when recall errors do occur, $\theta > 0$ indicates a (strong) preference for recalling positive grades over negative ones.

Notably, the magnitude of θ_y tells us that the marginal benefit of recalling a high grade increases by over a third if prior self-esteem is high. This result is highly consistent with the mood-memory mechanism distinctive to FSE. However, it should be noted that the impact of θ_y on the dynamics of y_t is quantitatively small. Both θ_y and μ_r^y are small enough that any impact of y_{t-1} on the distribution of r_t filters through to little impact on the distribution of y_t. By far the most important source of persistence in y_t is through direct serial correlation, captured by μ_{yy}. However, as previously noted, μ_{yy} is likely to capture additional mood-memory effects: if prior self-esteem is high, recall of signals other than r_t (such as for other subjects, or for mathematics in periods prior to t, could be more favourable).

Estimates of λ and λ_f indicate that the cost of making a recall error is almost halved in Fall relative to Spring, indicating a rapid time decay in recall accuracy. As previously illustrated, the main effect of this change is that positively biased recall becomes much more likely in Fall: if $g_t = 0$, the marginal utility of recalling $r_t = 1$ is substantially higher when $f_t = 1$.

Recalling a higher grade also affects continuation values, in that both g_t and y_t are more likely to equal 1 if a higher grade is recalled. However, these feedback effects on state variables are only of second-order importance for decision making.[15]

Unobserved heterogeneity. In Table 9, we report estimates for the model which permits students to take one of two unobserved types, $c \in \{1, 2\}$. Recall that in this model, we permit c to influence both preferences for ego-enhancing recall and the evolution of state variables.

Parameter estimates for this specification reveal that the basic model masks substantial heterogeneity in recall preferences across students. $\theta^1 = 2.3$, so type 1 is characterised by very strong preferences for self-enhancing recall of grades. Type 1's preferences for recall are also characteristic of fragile self-esteem: $\theta_y^1 > 0$, meaning positive memories of grades are more attractive when prior self-esteem

[15] As previously noted, we take the standard practice of fixing the discount factor β since it is not identified in stationary dynamic discrete choice models. However, in Table A6 we report coefficient estimates for a wide set of values of β. Parameter estimates vary very little across specifications. The only parameter significantly affected by changes to β is θ: if β is smaller, the present benefit of biased recall has more work to do in explaining choice probabilities.

Table 8: Parameter estimates for basic structural model

Utility function	
θ	0.65*** (0.01)
θ_y	0.23*** (0.02)
λ	2.23*** (0.02)
λ_f	−1.00*** (0.02)
Markov process: actual grades	
μ_1^g	−1.96*** (0.02)
μ_y^g	0.42*** (0.02)
μ_g^g	1.62*** (0.03)
μ_f^g	−0.08*** (0.02)
μ_r^g	0.90*** (0.03)
Markov process: self-esteem	
μ_1^y	−1.78*** (0.02)
μ_2^y	1.86*** (0.02)
μ_3^y	0.01 (0.03)
μ_4^y	0.07*** (0.02)
μ_5^y	0.67*** (0.03)
Observations	1,176

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors in parentheses. Model estimated by full-information maximum likelihood. Discount factor set to $\beta = 0.9$.

is higher. On the other hand, type 2 has substantially weaker preferences for self-enhancing recall: $\theta^2 = 0.19$. Type 2 also exhibits fragile self-esteem, although less so than for type 1. Notably, our auxiliary model reveals that both indicators of a growth mindset strongly predict membership of type $c = 1$.[16]

Another substantial difference between the two types is in the unobserved factors affecting attainment. Type 2 achieves much more poorly than type 1: $\omega^g = -0.25$, indicating a 25pp lower conditional probability of receiving $g_t = 1$. They also have lower self-esteem: $\omega^y = -0.04$, indicating the conditional probability of having high self-esteem is about 4pp lower for type 2. Thus, our approach of permitting heterogeneity in recall preferences to be correlated with heterogeneity in state variables is justified by the data.

There are also differences across types in the effects of recalled grades on state variables. Since $\omega_r^y > 0$, the self-esteem of type 1 is *less* vulnerable to recalling low grades. This suggests that in addition to being less likely to inflate recall of poor grades, type 2s are also less likely to defend their self-esteem when they do recall low grades. One interpretation is that there are two layers of self-deception when receiving a poor grade. First, if possible, students recall a higher grade than actually received. However, if it is infeasible or psychologically too costly to recall an incorrect grade, type 1s further inure themselves to low attainment by perceiving both a weaker mapping from grades to ability and a higher unconditional mean for ability. Another possibility is of course that type 1 is right to be less threatened by low attainment because it is objectively less informative. There is some support for this interpretation from the fact that $\omega_r^g > 0$, implying attainment is genuinely less responsive to recall for type 1 than type 2.

While the self-esteem of type 1 is more resilient to low grades than that for type 2, a natural question is whether this conclusion extends to wellbeing. As previously noted, θ in our model is related to the parameter k in Kőszegi et al. (2022)'s model, governing the strength of the individual's concern for their ego. In FSE, higher values of k increase the probability of recalling positive outcomes and reduces the probability of recalling negative ones, holding mood constant. However, a larger value of k also crucially implies that mood is more sensitive to self-esteem. For this purpose we re-estimate the model with an auxiliary state

[16] Excluding the auxiliary model for type probabilities leads to similar results.

variable capturing school-centric welfare, measured by the question 'How much do you like school in general?', and allow y_t to have a heterogeneous marginal effect on welfare by type.[17] Instructively, the marginal effect of self-esteem on welfare is *greater* for type 1. This suggests an interpretation of the difference in θ^c across types; recalling favourable grades is more important to type 1s because self-esteem is more important for their welfare.

One concern with this interpretation may be that θ^c differs across types simply because, when grades are forgotten, recall converges on historical attainment and type 1 does substantially better than type 2 on average. We can directly examine the plausibility of this claim by inspecting the steady-state distributions of r_t and g_t for both types. For type 1, $P(r_t = 1|f_t = 1, c = 1) = 0.90$ in the long run, whereas $P(g_t = 1|f_t = 1, c = 1) = 0.57$. Likewise for Spring periods, $P(r_t = 1|f_t = 0, c = 1) = 0.80$, but $P(g_t = 1|f_t = 0, c = 1) = 0.55$. So, type 1 recalls high grades 90% of the time in Fall, and 80% of the time in Spring, but only achieve them 55–57% of the time in the long run. The difference in long-run recall *bias* relative to type 2 is very stark. In Spring, the long-run probability of type 2 recalling a high grade (30%) is only 9pp higher than the probability of achieving one (21%), much smaller than the 25pp bias for type 1. In Fall, the positive bias in recall is 16pp for type 2s, relative to a 33pp bias for type 1. Thus, differences in recall preferences across types cannot be rationalised by average differences in attainment.

4.5 The dynamics of self-esteem

The model outlined in Sections 4.1 and 4.2 creates a rich dynamic interaction of recall choices, self-esteem and actual attainment. In this section, we explore how the dynamic distributions of these variables depend on initial conditions and across the two unobserved types of student.

As previously noted, shocks to self-esteem are possible in every period. Neither high self-esteem ($y_t = 1$) or low self-esteem ($y_t = 0$) is an absorbing state, so the long-run distribution of self-esteem is only a function of model parameters. Nonetheless, our model can tell us for how long various shocks to initial conditions affect the distribution of self-esteem. As previously outlined, this analysis is

[17] More detail on this alternative version of the model can be found in Appendix B.

Table 9: Parameter estimates for structural model with unobserved heterogeneity

Utility function	
θ^1	2.27^{***} (0.25)
θ^2	0.19^{***} (0.03)
$\theta^2 - \theta^1$	-2.08^{***} (0.23)
θ_y^1	0.54^{*} (0.28)
θ_y^2	0.14^{**} (0.06)
$\theta_y^2 - \theta_y^1$	-0.40 (0.42)
λ	2.37^{***} (0.03)
λ_f	-1.20^{***} (0.03)
Markov process: actual grades	
μ_1^g	-0.80^{***} (0.13)
μ_y^g	0.25^{***} (0.02)
μ_g^g	1.55^{***} (0.03)
μ_f^g	0.12^{***} (0.01)
μ_r^g	0.06 (0.15)
ω^g	-0.25^{***} (0.01)
ω_r^g	0.12^{***} (0.01)
Markov process: self-esteem	
μ_0^y	-1.55^{***} (0.13)
μ_y^y	1.82^{***} (0.02)
μ_g^y	0.02 (0.03)
μ_f^y	0.06^{***} (0.02)
μ_r^y	0.42^{***} (0.14)
ω^y	-0.04^{***} (0.00)
ω_r^y	0.06^{***} (0.01)
Auxiliary model for $c = 2$	
γ_1	0.41^{***} (0.13)
γ_2	1.09^{***} (0.34)
γ_3	0.78^{*} (0.44)
Observations	1,176

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors in parentheses. Model estimated by full-information maximum likelihood. Discount factor set to $\beta = 0.9$.

related to the notion of fragility used in Kőszegi et al. (2022): if a self-esteem equilibrium is more fragile, it may be more easily perturbed by shocks in the short run.

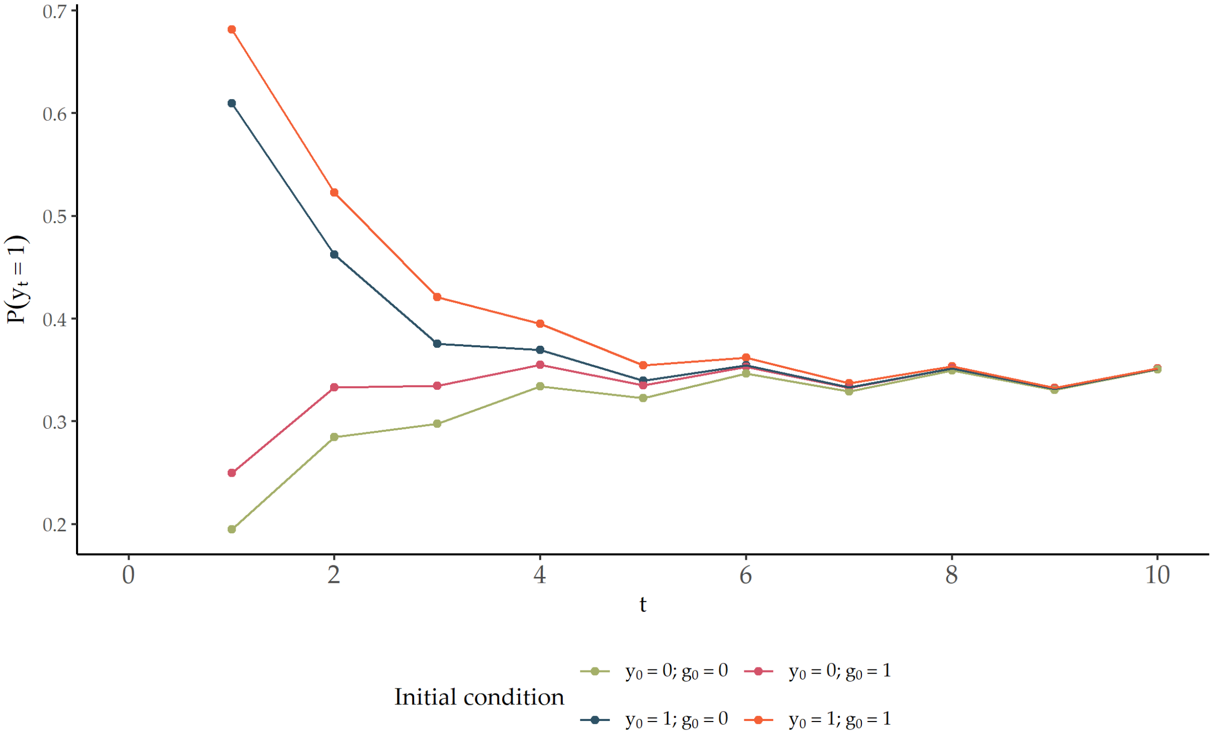
Basic model. Figure 5 illustrates the dynamics of self-esteem for the basic model, without unobserved heterogeneity. Since t corresponds to survey periods, each time period is about half a year. For this reason, self-esteem is also higher in even-numbered periods (corresponding to *Fall*, when higher grades are recalled and self-esteem is also higher conditional on recall).

Unsurprisingly, the largest dynamic impact on the distribution of self-esteem is from a shock to initial conditions for self-esteem. An initial condition with $y_0 = 0$ rather than $y_0 = 1$ implies just over a 40-percentage-point decrease in the likelihood of having high self-esteem in the following period. Notably, even 2 years later (in $t = 4$), $P(y_t = 1)$ is about 5 percentage points lower in scenarios with $y_0 = 0$. This persistence is largely driven by serial dependence in y_t, captured by μ_y^y, while the mood-memory utility parameter θ_y is quantitatively unimportant. Thus, if mood-memory plays an important role in the dynamics of self-esteem, it must be through other means than just recall of the most recent mathematics grade. As previously noted, this is plausible if y_{t-1} affects recall of unobserved signals in t.

Although actual grades g_t have no direct impact on y_t ($\mu_g^y = 0$), they significantly impact what grade is recalled and therefore indirectly impact self-esteem because $\mu_r^y > 0$. Figure 5 indicates a highly persistent impact on self-esteem of receiving a single low grade in a quarterly report card. $P(y_t = 1)$ is around 8 percentage points lower in starting conditions with $g_0 = 0$, and the effect barely dissipates until about 1.5–2 years after the initial shock.

Model with unobserved heterogeneity. Figure 6 demonstrates the dynamics of $P(y_t = 1)$ by type in the extended model with unobserved heterogeneity. Types 1 and 2 have quite different long-run equilibria for self-esteem. Around half of the difference in the long-run equilibrium level of $P(y_t = 1)$ is driven by a fixed effect ω^y, and the other half by differences in recall preferences (θ and θ_y). That is, type 1 has higher self-esteem, and around half of this can be explained by the fact that in each period, they have a stronger preference to recall higher grades than type 2.

Figure 5: Dynamics of $P(y_t = 1)$ — basic model



Note: $y_t = 1$ implies the student believes they have above average academic ability for their school and year group. t corresponds to survey periods, so each period is approximately half a year.

Figure 6: Dynamics of $P(y_t = 1)$ — model with unobserved heterogeneity

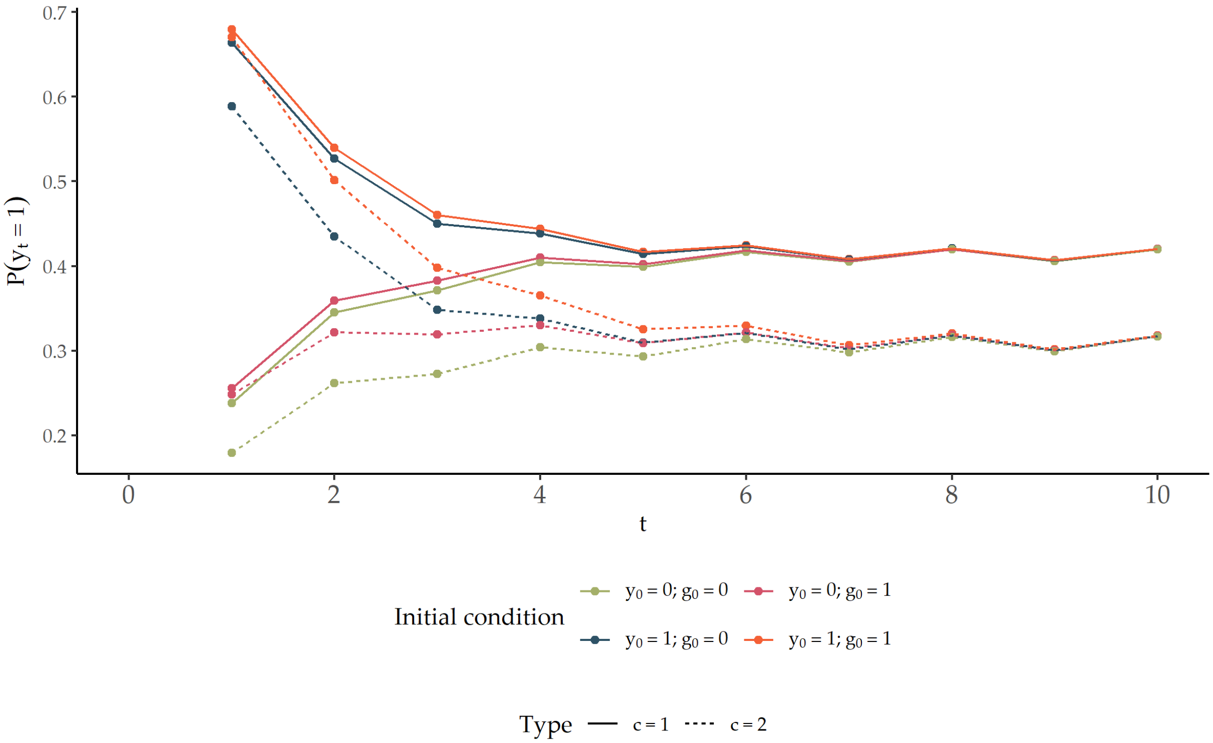
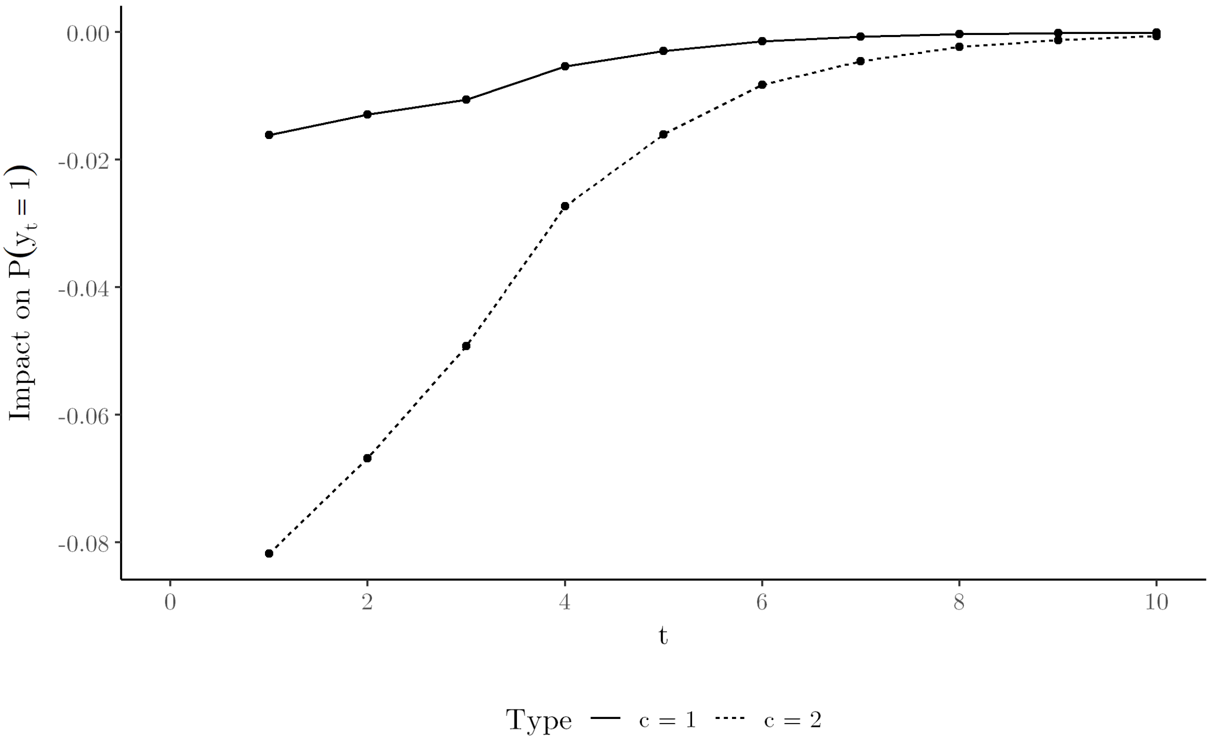


Figure 7: Impact of initial condition $g_0 = 0$ vs. $g_0 = 1$



Note: $y_t = 1$ implies the student believes they have above average academic ability for their school and year group. t corresponds to survey periods, so each period is approximately half a year. Initial condition in both cases has $y_0 = 1$.

Recalling our earlier discussion, one measure of the fragility of an SEP at $y_t = k$ is $Pr(y_t = k|y_{t-1} = k)$. This measure is monotonically related to the size of shock required to displace a student from $y_t = k$ to the other equilibrium in the following period. Using this definition, we can see that the high SEP at $y_t = 1$ is equally stable for type 1 and type 2, conditional on receiving a high grade $g_t = 1$. However, conditional on receiving a low grade, the high SEP is more fragile for type 2. This type has only a 60% chance of remaining at $y_t = 1$ in the following period, compared to 70% for type 1. In other words, an equilibrium with unrealistically high self-esteem is more stable for type 1, who places a much higher value on ego.

While both types are affected similarly by direct shocks to self-esteem, they differ significantly in their responses to shocks to actual grades. In particular, type 1 is much more resilient to attainment shocks than type 2, largely because they are more likely to recall a high grade even when they receive a low one. This is illustrated more clearly in Figure 7: the impact of an initial condition ($y_0 = 1; g_0 = 1$); relative to ($y_0 = 1; g_0 = 1$) is sharply negative for type 2 but negligible for type 1. Just as in FSE, those who place a higher value on ego tend to have a more positively biased recall process, making high self-esteem less vulnerable to shocks.

Information intervention. One category of policy intervention that has attracted particular interest in the literature on education is the provision and reinforcement of feedback information (see e.g. Dizon-Ross, 2019). In our setting, we can model this intervention by setting $r_t = g_t$ and examining how the dynamics of our two key state variables, y_t and g_t, are affected.

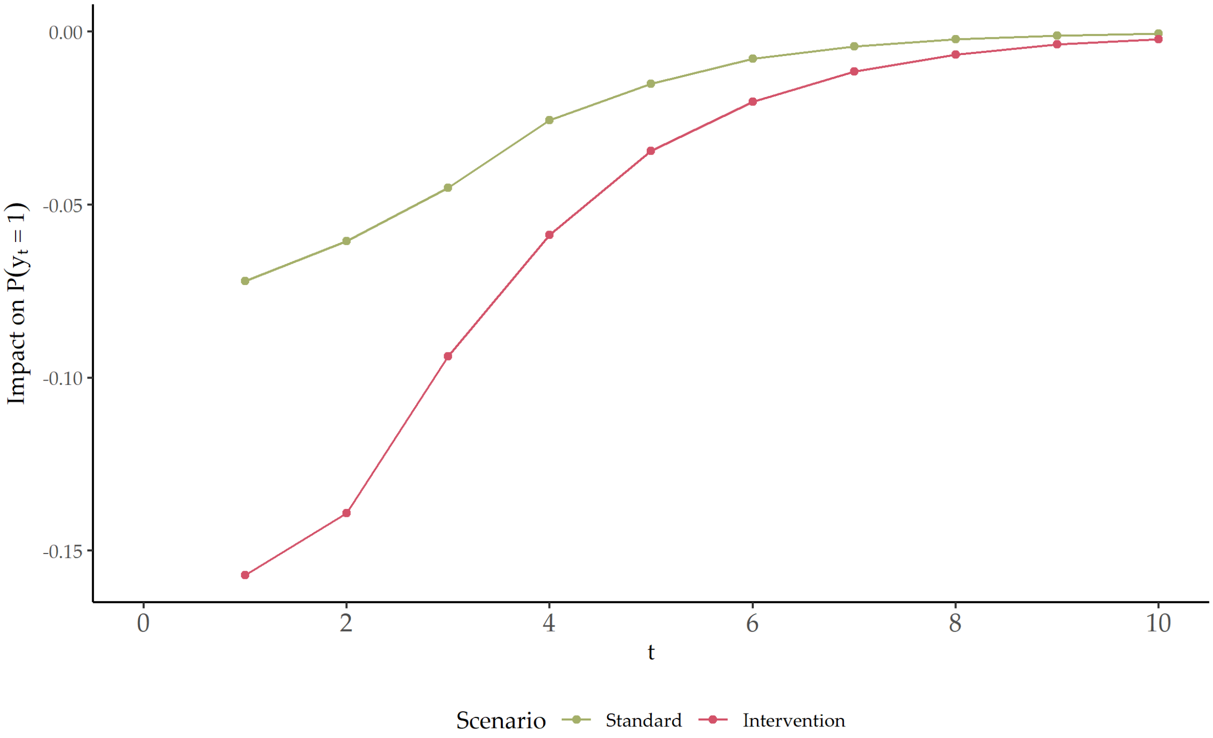
The effects of this intervention are most obvious when a poor grade is received. To explore this case, we plot the impact on $P(y_t = 1)$ of starting in an initial condition ($y_0 = 1; g_0 = 0$) relative to an initial condition ($y_0 = 1; g_0 = 1$), and explore how that impact differs with and without the intervention. To begin with, we do so using the basic model without unobserved heterogeneity (Figure 8).

The initial impact of receiving a low grade on self-esteem is almost three times larger under the informational intervention. Two years after the intervention, the expected impact is almost as large as would have been the initial impact without the intervention. Thus, self-deception is crucial for the protection of self-esteem, especially for low achievers. That protective mechanism would precluded by better information transmission by limiting opportunities for biased beliefs.

Clearly, if self-esteem is founded on self-deception, it is likely to be unrealistic. A natural question is whether correcting unrealistically high self-esteem could be beneficial for students. For two reasons, we can be fairly sure that such an effect is implausible in our context. First, we know that higher self-esteem is associated with higher psychological wellbeing while at school, controlling for grades (Appendix B). Second, our basic parameter estimates indicate that high self-esteem (and recalling a higher grade) is associated with higher future attainment; both μ_y^g and μ_r^g are greater than 0.

To illustrate this point, we also plot the impact of the informational intervention on how g_t responds after a shock to its own initial condition (Figure 9). Under

Figure 8: Impact of poor initial grade on $P(y_t = 1)$ — with and without information intervention



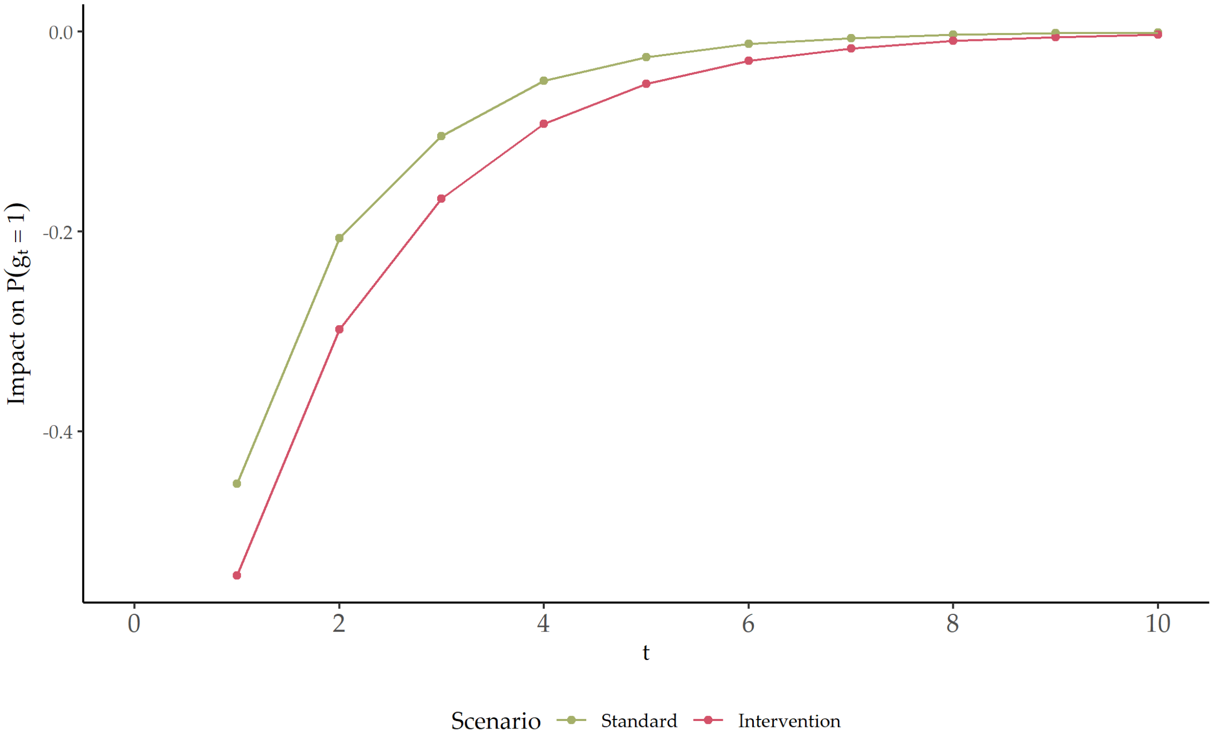
Note: $y_t = 1$ implies the student believes they have above average academic ability for their school and year group. t corresponds to survey periods, so each period is approximately half a year. Initial condition in both cases has $y_0 = 1$.

the informational intervention, we see that the impact on g_t is more pronounced: recalling a low grade directly affects attainment in the following period, and this negative impact is inherited in future periods. Strikingly, the impact of the informational intervention in this setting is that students are approximately 10pp (25%) more likely to achieve another poor grade in the period after a low one. While one should not use this result to claim that unrealistically high self-esteem is beneficial in general, our results do highlight a possible negative side-effect of informational interventions: for students who achieve poorly, constraints on self-deception can result in psychological costs and a negative impact on attainment. These findings is consistent with theories connecting self-confidence to performance (Bénabou and Tirole, 2002; Compte and Postlewaite, 2004), but has received little attention in applied work examining the removal of informational frictions in education.

We also examine how the same intervention affects different types in our extended model with unobserved heterogeneity. As before, the main effect of the intervention is to increase students' sensitivity to achieving poor grades. Starting with the impact on self-esteem, we see that both types are strongly negatively affected by the intervention (Figure 10). If the intervention were only implemented on type 1, it would have the effect of completely closing the gap in the impact of receiving a poor grade: type 1 relies on biased recall to maintain their self-esteem in the face of poor attainment. Even though type 2 is much less likely to bias their recall of poor grades without the intervention, the information intervention has almost as large an impact on self-esteem: since $\omega_r^y = 0.06$, their self-esteem is more sensitive to what grade is recalled.

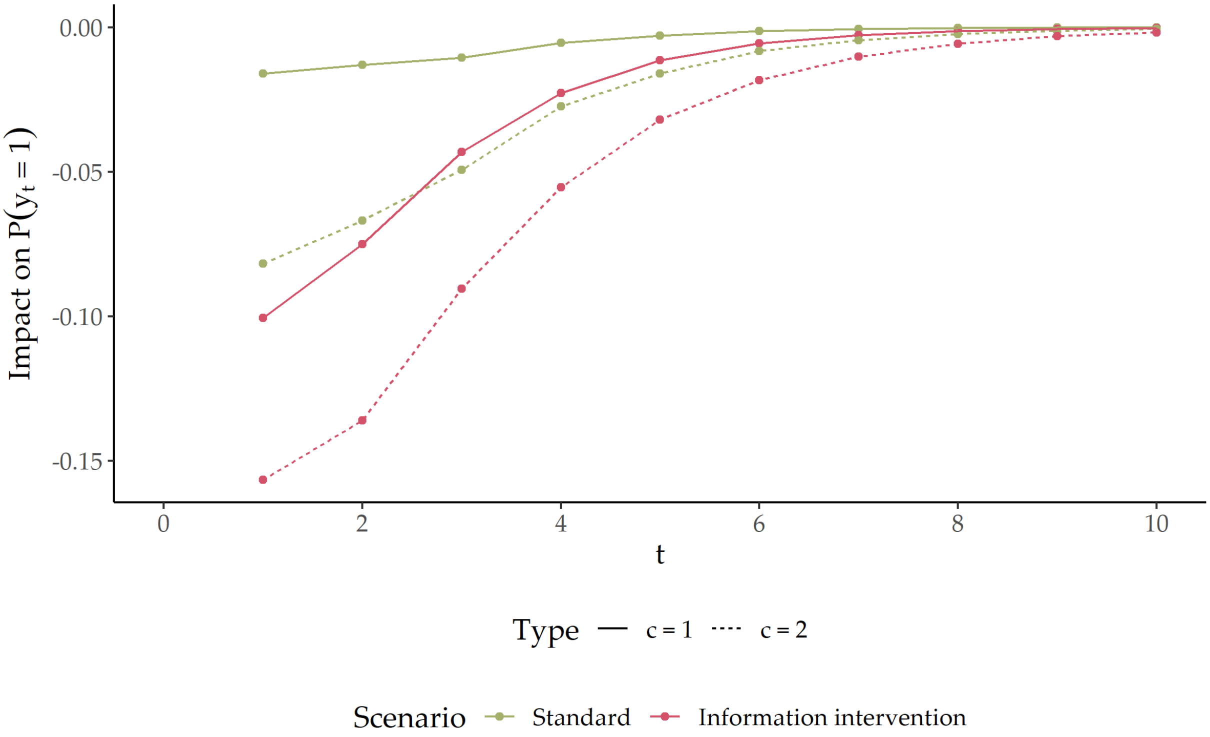
We can also examine how the intervention would impact dynamics for g_t (Figure 10). Here, the difference between types is significant: type 1 is essentially unaffected by the intervention, since r_{t-1} has little impact on g_t. However, recall is more strongly associated with future attainment for type 2, who is generally low achieving. Thus, an informational intervention could worsen attainment gaps because shocks to self-esteem are more consequential for low achievers.

Figure 9: Impact of poor initial grade on $P(g_t = 1)$ — with and without information intervention



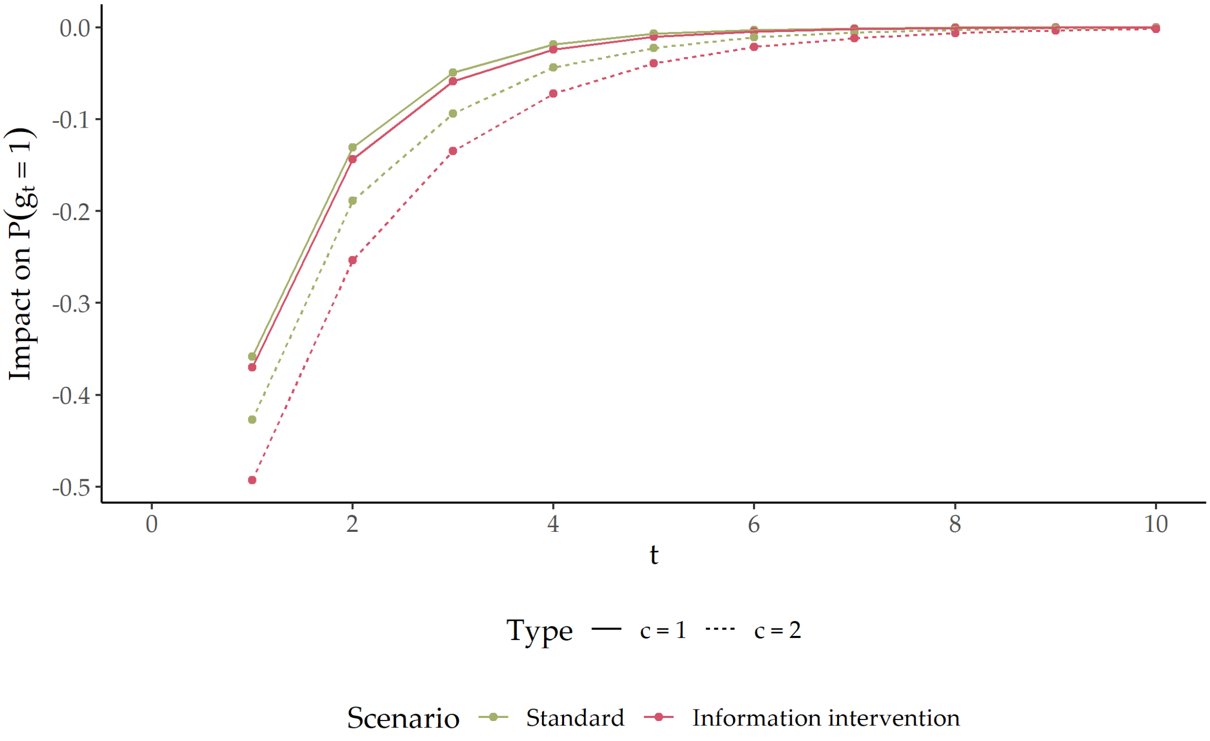
Note: $g_t = 1$ indicates an actual grade of *Good* or *Excellent*. t corresponds to survey periods, so each period is approximately half a year. Initial condition in both cases has $y_0 = 1$.

Figure 10: Impact of poor initial grade on $P(y_t = 1)$ — with and without information intervention, 2-type model



Note: $y_t = 1$ implies the student believes they have above average academic ability for their school and year group. t corresponds to survey periods, so each period is approximately half a year. Initial condition in all cases with $y_0 = 1$.

Figure 11: Impact of poor initial grade on $P(g_t = 1)$ — with and without information intervention, 2-type model



Note: $g_t = 1$ indicates an actual grade of *Good* or *Excellent*. t corresponds to survey periods, so each period is approximately half a year. Initial condition in all cases with $y_0 = 1$.

5 Conclusion

This chapter finds support for theories predicting that individuals distort their recall of ego-relevant information to protect their self-esteem. It also demonstrates that these distortions are enabled by biased memory loss. Principally, these findings 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 we find for biased memory loss, matching Zimmermann (2020) in the laboratory, is particularly notable, since it highlights that beliefs may rapidly diverge from objective reality over time. It also suggests that greater effort may be required to bolster the retention of negative signals as time passes.

Our use of rich longitudinal data facilitates a test of a recently proposed two-way relationship between beliefs and recall (Kőszegi et al., 2022). This theory is particularly powerful because it suggests a possible path-dependence for beliefs through biased, associative recall. It also provides a more compelling explanation for biased recall than existing models: unfavourable signals may produce a disproportionate impact on self-esteem, relative to a standard Bayesian model, by stimulating recall of associated negative signals. Our model finds support for the key mechanism, that biases in recall depend on prior self-esteem. Like the theory suggests, the effect of sole signals on self-esteem can be both large and long-lasting.

The welfare implications of biased recall are generally not well studied empirically. In our model, we take it as given that recall distortions serve to optimise psychological welfare, taking cues from existing work on ego utility. This approach seems justified through auxiliary evidence that recall and self-esteem impact students' wellbeing at school. Additionally, it does not appear to be the case that inflated recall negatively affects academic performance; recalling a higher grade (and having higher self-esteem) is in fact associated with improved attainment in our sample, especially for low achievers. This is consistent with theories of belief- or confidence-based motivation (Bénabou and Tirole, 2002; Compte and Postlewaite, 2004) rather than theories in which belief distortion comes at the cost of distorting choices (Brunnermeier and Parker, 2005). Thus, our results highlight that policies aiming to correct recall errors and overconfidence should be combined with measures to protect the psychological welfare of low achievers.

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

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, *73*(5), 1477–1524.

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

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.

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

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

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. In *Development of achievement motivation* (pp. 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.

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.

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, 271–320.

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.

Huffman, D., Raymond, C., & Shvets, J. (2022). Persistent overconfidence and biased memory: Evidence from managers. *American Economic Review*, *112*(10), 3141–3175.

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

Kőszegi, B., Loewenstein, G., & Murooka, T. (2022). Fragile self-esteem. *The Review of Economic Studies*, *89*(4), 2026–2060.

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

Ortoleva, P., & Snowberg, E. (2015). Overconfidence in political behavior. *American Economic Review*, *105*(2), 504–535.

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

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369.

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

A Additional reduced form analyses

Table A1: Reduced form models — biased recall

	Dependent variable:	
	Flattering recall error	
	(1)	(2)
Actual grade: *Satisfactory*	0.257*** (0.025)	0.193*** (0.020)
Actual grade: *Unsatisfactory*	0.543*** (0.032)	0.434*** (0.024)
Fall	0.234*** (0.020)	0.240*** (0.020)
Individual FE	Yes	No
Academic year FE	Yes	Yes
Observations	2,448	2,448
R^2	0.431	0.158
Adjusted R^2	0.258	0.156

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Linear probability model, robust standard errors in parentheses. All grades are from quarterly report cards for mathematics, and recall of the most recent quarterly grade is elicited in a biannual survey. Observations with an actual grade of *Excellent* are omitted since flattering recall errors are impossible; the omitted actual grade category in this case is *Good*.

Table A2: Reduced form models — recall errors, heterogeneity by sex and race

	Dependent variable:	
	Incorrect recall	
	(1)	(2)
Actual grade: *Good*	0.071 (0.081)	0.113* (0.062)
Actual grade: *Satisfactory*	0.232*** (0.086)	0.245*** (0.059)
Actual grade: *Unsatisfactory*	0.403*** (0.095)	0.391*** (0.063)
Female	—	0.025 (0.051)
Black	—	−0.037 (0.051)
Fall	0.237*** (0.020)	0.238*** (0.020)
Actual grade: *Good**Female	−0.093 (0.080)	−0.066 (0.061)
Actual grade: *Satisfactory**Female	−0.008 (0.082)	0.015 (0.058)
Actual grade: *Unsatisfactory**Female	−0.067 (0.093)	−0.045 (0.064)
Actual grade: *Good**Black	0.153* (0.083)	0.085 (0.062)
Actual grade: *Satisfactory**Black	0.087 (0.088)	0.011 (0.060)
Actual grade: *Unsatisfactory**Black	0.227** (0.098)	0.144** (0.066)
Individual FE	Yes	No
Academic year FE	Yes	Yes
Observations	2,694	2,694
R^2	0.375	0.134
Adjusted R^2	0.199	0.128

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Linear probability model, robust standard errors in parentheses. All grades are from quarterly report cards for mathematics, and recall of the most recent quarterly grade is elicited in a biannual survey. The omitted actual grade category is *Excellent*.

Table A3: Reduced form model — memory loss, heterogeneity by sex and race

	Dependent variable:	
	Recalled grade: *Excellent*	
	(1)	(2)
Actual grade: *Good*	0.074 (0.082)	0.115* (0.062)
Actual grade: *Satisfactory*	0.235*** (0.086)	0.247*** (0.059)
Actual grade: *Unsatisfactory*	0.408*** (0.095)	0.393*** (0.063)
Female	—	0.017 (0.054)
Black	—	−0.021 (0.056)
Fall	0.252*** (0.039)	0.248*** (0.038)
Actual grade: *Good**Female	−0.094 (0.081)	−0.064 (0.061)
Actual grade: *Satisfactory**Female	−0.008 (0.083)	0.017 (0.058)
Actual grade: *Unsatisfactory**Female	−0.069 (0.092)	−0.043 (0.063)
Actual grade: *Good**Black	0.150* (0.084)	0.081 (0.063)
Actual grade: *Satisfactory**Black	0.084 (0.088)	0.007 (0.060)
Actual grade: *Unsatisfactory**Black	0.223** (0.099)	0.140** (0.065)
Female*Fall	−0.010 (0.035)	0.011 (0.035)
Black*Fall	−0.013 (0.039)	−0.022 (0.038)
Individual FE	Yes	No
Academic year FE	Yes	Yes
Observations	2,694	2,694
R^2	0.375	0.134
Adjusted R^2	0.198	0.128

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Linear probability model, robust standard errors in parentheses. All grades are from quarterly report cards for mathematics, and recall of the most recent quarterly grade is elicited in a biannual survey. The omitted actual grade category is *Excellent*.

Table A4: Reduced form models — grade prediction errors

	Dependent variable:		
	Incorrect prediction		
	(1)	(2)	(3)
Recall error	0.062*** (0.021)	0.042** (0.022)	0.089*** (0.019)
Actual grade: *Good*	−0.008 (0.044)	0.010 (0.044)	0.085** (0.035)
Actual grade: *Satisfactory*	0.016 (0.045)	0.037 (0.045)	0.164*** (0.033)
Actual grade: *Unsatisfactory*	0.015 (0.050)	0.043 (0.050)	0.201*** (0.036)
Fall		0.081*** (0.022)	0.078*** (0.022)
Individual FE	Yes	Yes	No
Academic year FE	Yes	Yes	Yes
Observations	2,883	2,883	2,883
R^2	0.306	0.310	0.045
Adjusted R^2	0.110	0.115	0.042

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Linear probability model, robust standard errors in parentheses. All grades are from quarterly report cards for mathematics, and recalled and expected quarterly grades are elicited in a biannual survey. The omitted actual grade category is *Excellent*.

B Structural model: Extensions and robustness checks

Here, we briefly outline an extended structural model with an additional state variable capturing school-centric welfare. The purpose of this model is to investigate whether different types of student place different weights on academic self-esteem in utility. As mentioned in the main body, the measure of welfare we use is based on the question 'How much do you like school in general', and we code responses $z_t = 1$ if the student's response was 'A lot' and $z_t = 0$ if they responded with 'A little' or 'Not at all'. We then specify the same model as in Section 4.2, but with z_t as an additional state variable which appears nowhere else. As with the other state variables, z_t follows a conditional Markov process,

$$P(z_t = 1|z_{t-1}, g_t, y_t, c = 1) = \text{logit}(\mu_1^z + \mu_z^z z_{t-1} + \mu_g^z g_t + \mu_y^z y_t), \tag{16}$$

for $c = 1$, and

$$P(z_t = 1|z_{t-1}, g_{t-1}, y_{t-1}, c = 2) = P(z_t = 1|z_{t-1}, g_t, y_t, c = 1) + \omega^z + \omega_y^z(y_t = 1), \tag{17}$$

for $c = 2$. Crucial is the coefficient ω_y^z, which lets the marginal effect of self-esteem vary by type. The coefficient estimates are reported in Table A5. The coefficients are qualitatively the same as in the baseline 2-type model, although the fragile self-esteem parameter θ_y^1 is no longer statistically significant because its magnitude is slightly reduced and standard errors are inflated. The main coefficient of interest in this case is ω_y^z, capturing heterogeneity in the impact of self-esteem on welfare. Self-esteem has a substantially smaller impact on the welfare of type 2: the marginal effect of $y_t = 1$ is 9pp smaller for type 2 than for type 1. This evidence is consistent with the interpretation that the much larger θ^c for type 1 reflects a larger relevance of self-esteem for welfare.

Table A5: Parameter estimates for structural model with unobserved heterogeneity and auxiliary state variable for welfare

Utility function	
θ^1	2.44^{***} (0.31)
θ^2	0.18^{***} (0.03)
θ^1_y	0.39 (0.36)
θ^2_y	0.17^{***} (0.06)
λ	2.35^{***} (0.03)
λ_f	-1.24^{***} (0.03)
Markov process: actual grades	
μ^g_1	-0.82^{***} (0.18)
μ^g_y	0.25^{***} (0.02)
μ^g_g	1.54^{***} (0.03)
μ^g_f	0.15^{***} (0.01)
μ^g_r	0.23 (0.17)
ω^g	-0.24^{***} (0.01)
ω^g_r	0.06^{***} (0.01)
Markov process: self-esteem	
μ^y_1	-1.61^{***} (0.15)
μ^y_y	1.84^{***} (0.03)
μ^y_g	−0.00 (0.03)
μ^y_f	0.08^{***} (0.02)
μ^y_r	0.49^{***} (0.16)
ω^y	-0.03^{***} (0.00)
ω^y_r	0.04^{***} (0.01)
Markov process: welfare	
μ^z_1	-1.84^{***} (0.09)
μ^z_z	1.69^{***} (0.03)
μ^z_g	-0.18^{***} (0.03)
μ^z_f	0.07^{***} (0.02)
μ^z_y	0.84^{***} (0.14)
ω^z	-0.01^{***} (0.00)
ω^z_y	-0.09^{***} (0.01)
Auxiliary model for $c = 2$	
γ_1	0.52^{***} (0.10)
γ_2	1.25^{***} (0.41)
γ_g	0.83^{**} (0.40)
Observations	1,158

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors in parentheses. Model estimated by full-information maximum likelihood. Discount factor set to $\beta = 0.9$.

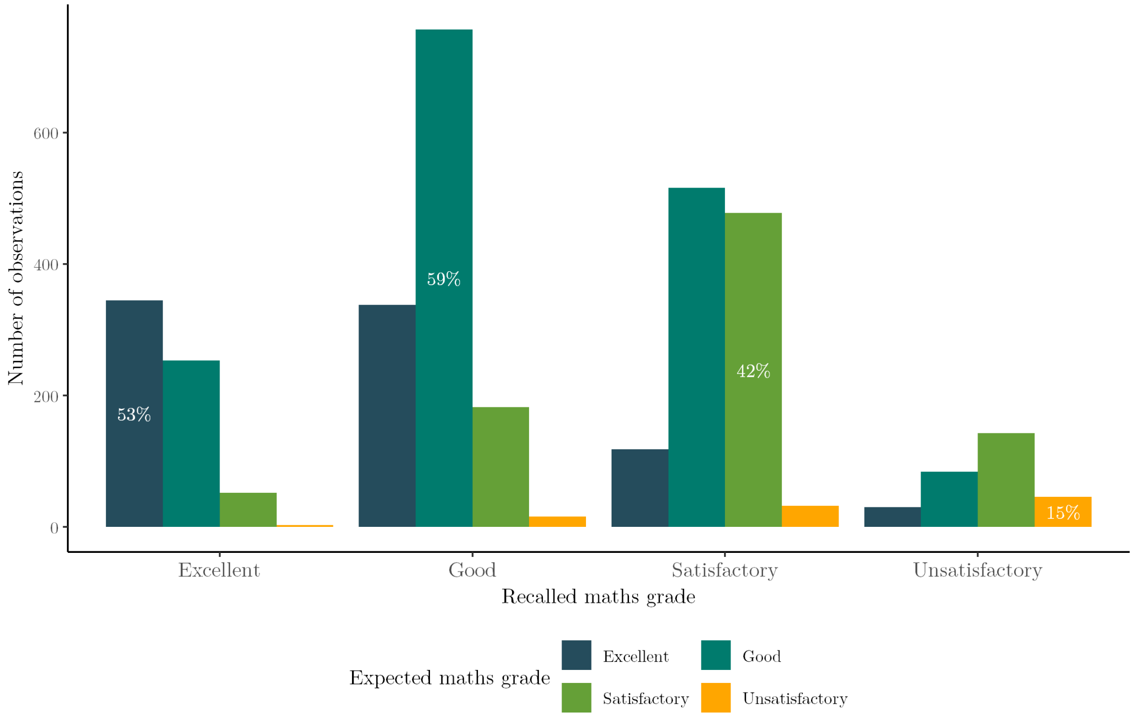
Table A6: Structural model: variations on β

			β		
	0.1	0.25	0.5	0.75	0.9
Utility function					
θ	0.79***	0.77***	0.73***	0.68***	0.65***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
θ_y	0.27***	0.26***	0.25***	0.24***	0.23***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
λ	−2.26***	−2.25***	−2.24***	−2.23***	−2.23***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
λ_f	1.00***	1.00***	1.00***	1.00***	1.00***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Markov process: actual grades					
μ_1^g	−1.78***	−1.78***	−1.78***	−1.78***	−1.78***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
μ_y^g	1.86***	1.86***	1.86***	1.86***	1.86***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
μ_g^g	0.00	0.00	0.01	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
μ_f^g	0.07***	0.07***	0.07***	0.07***	0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
μ_r^g	0.67***	0.67***	0.67***	0.67***	0.67***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Markov process: self-esteem					
μ_1^y	−1.96***	−1.96***	−1.96***	−1.96***	−1.96***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
μ_y^y	0.42***	0.42***	0.42***	0.42***	0.42***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
μ_g^y	1.62***	1.62***	1.62***	1.62***	1.62***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
μ_f^y	−0.08***	−0.08***	−0.08***	−0.08***	−0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
μ_r^y	0.89***	0.89***	0.89***	0.89***	0.90***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Observations			1,176		

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors in parentheses. Model estimated by full-information maximum likelihood.

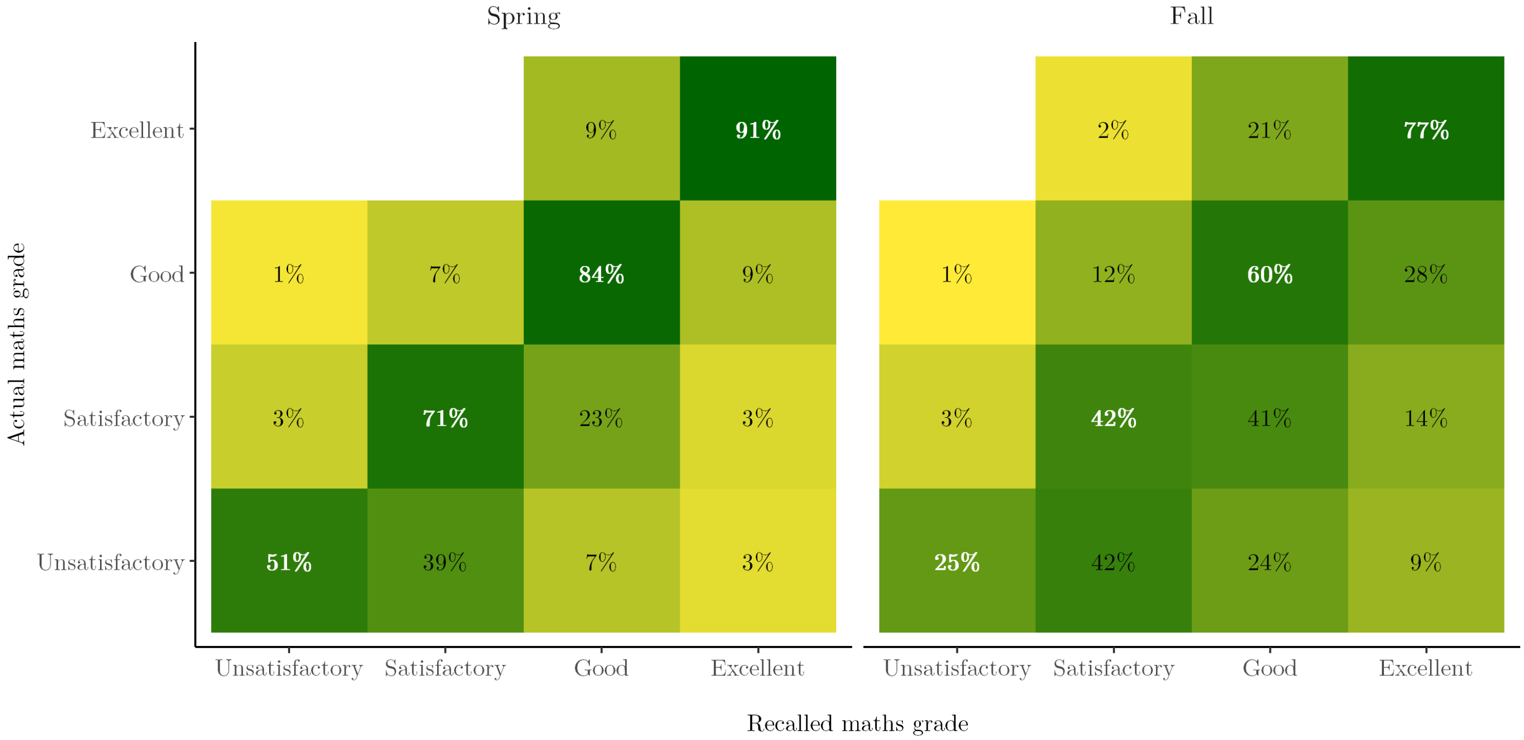
C Additional figures

Figure A12: Expected grades vs recalled



Note: Aggregated over survey sweeps. White percentages are the share of observations with expected grades equal to recalled grades, by recalled grade.

Figure A13: Actual and remembered grades, split by survey edition



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