# Self-Confidence and Motivated Memory Loss: Evidence From Schools

Vivek Roy-Chowdhury*

August 2022

**Abstract**

Using longitudinal data on teenagers' memories of recent mathematics report card grades and administrative data on actual grades, this paper finds evidence of motivated beliefs during schooling. Students i) make more errors in recalling lower grades and ii) have more flattering memories of grades when the survey was administered with a longer delay, suggesting a substantial role for biased memory loss in enabling belief distortion. Recalling a higher grade is associated with higher academic self-esteem, enjoyment of school, expected grades and future actual grades, consistent with theories which link biased recall to higher utility through enhancements of both beliefs and future outcomes.

**JEL codes:** D91, I21, J83

**Keywords:** motivated beliefs, education, ability, recall, self-confidence

*Faculty of Economics, University of Cambridge, Cambridge, CB3 9DD, United Kingdom; vr277@cam.ac.uk. I am grateful to Toke Aidt, Leonardo Bursztyn, Caterina Calsamiglia, Christopher Rauh, Julia Shvets, Jack Willis, Weilong Zhang, and attendees of the European Winter Meeting of the Econometric Society, the Barcelona School of Economics Summer Forum, the CESifo Summer Institute and the Early Career Behavioural Economics conference for their useful comments.

# 1 Introduction

There is now mounting evidence that self-relevant beliefs are systematically detached from reality in a number of settings, including financial markets (De Bondt and Thaler 1995) and personal ability (Svenson 1981). These beliefs matter for important decisions. Malmendier and Tate (2005) find that CEOs of Forbes 500 companies overestimate the returns to their own investment projects, explaining large distortions in investment allocations. Individuals may even distort beliefs when their own lives are at risk: Oster et al. (2013) find that individuals at very high risk of Huntington disease avoid diagnostic tests, and behave in general as if they do not have it.

A major research agenda on *motivated beliefs* has arisen to reconcile classical economic theories of belief updating, which suggest that information is always treated neutrally, with the notion that agents can ignore or forget unfavourable, self-relevant information to protect their egos. However, while laboratory studies on motivated beliefs are manifold, identification challenges mean there is currently scant empirical evidence on how this phenomenon may affect belief formation in the real world. Similarly, little is known empirically about whether motivated belief distortion enhances, or creates inefficiencies in, decision-making and outcomes.

This paper finds field evidence of motivated beliefs in the important formative context of schooling, suggests that biased memory loss may play a major role, and finds that self-enhancing memories are associated with better short-run outcomes. The analysis uses longitudinal data on adolescents' memories of report card grades in mathematics, comparing them to administrative data on actual grades. It begins by using within-individual variation to characterise biases in recall behaviour. This exercise generates two main findings. First, recall is positively biased and students make more errors in recalling low grades, even when using within-individual variation in the latter case. Second, memories of grades become significantly more positive when the survey was administered with a longer delay, indicating that biased memory loss enables a large part of belief distortion.

Exploiting broad-ranging longitudinal data, the analysis then turns to the possible motivations for biased recall by analysing associations between recalled grades and a range of short-run outcomes, both subjective to the student and material. Independent of actual grades, recalling a higher grade is positively associated with changes in academic self-esteem, enjoyment of school, teacher-rated enthusiasm, expected grades and future grades for a given student, but negatively

with time spent on homework. Taken together, these findings are consistent with theories in which belief distortion enhances utility through both beliefs themselves (Brunnermeier and Parker 2005; Köszegi 2006) and future outcomes (Bénabou and Tirole 2002; Compte and Postlewaite 2004).

The dataset we use is the Beginning School Study (BSS). The BSS is an educational study of students attending Baltimore City public schools during the 1980s, combining administrative data on attainment from schools with subjective data from surveys, primarily of children but also their parents and teachers. The BSS stands apart from other educational field datasets in its direct elicitation of subjects' perceptions of objective report card grades. In collecting students' recollections of recent signals, it permits the empirical analysis to avoid the problematic practice of inferring how signals are perceived from how beliefs update (Benoît and Dubra 2011).[1]

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

An acute concern for studies of self-belief updating in the laboratory is that real-life stakes may not be adequately replicated. For instance, the inconsistency across Eil and Rao (2011) and Mobius et al. (2014) relative to Ertac (2011) may be because Eil and Rao (2011) consider results on an IQ test, whereas Ertac (2011) examines a much more specialised assessment. For similar reasons, experiments like Coutts (2019) cannot alone discredit motivated beliefs theory: parity in belief updating across self-relevant and self-irrelevant signals can result from insufficient stakes in the self-relevant case. These discrepancies suggest room for field research to clarify what real-world stakes, if any, are sufficient to generate motivated beliefs in practice.

Oster et al. (2013) conduct a notable field study of information avoidance as a mechanism for belief protection. In that setting, individuals avoid testing for Huntington disease, seemingly

[1]The study's authors originally identified a positive recall bias in their own pooled cross-sectional analysis of grades (Alexander, Entwisle, and Bedinger 1994).

to avoid the psychic trauma of a positive diagnosis, consistent with Brunnermeier and Parker (2005)'s optimal expectations theory. However, since testing for Huntington disease is not costless (costs range from $200-$600), their analysis cannot definitively show that the underlying motive for low testing rates is belief distortion. Closer to the present research is Huffman et al. (2022), which presents field evidence associating biased recall with self-assessment, using cross-sectional data on a workplace tournament. Much as in this paper, the first stage of their analysis involves demonstrating that recall errors are more likely when performance was poorer. The second stage demonstrates that those managers who have flattering memories of their performance are the ones most likely to overpredict their future performance relative to a model-based prediction.

This study contributes to existing field research in several ways. The first is in demonstrating a substantial increase in positive recall biases over time, indicating that biased memory loss accounts for a large part of the bias observed in ex-post recall. The second is in the use of longitudinal data, which both permits us to examine links between recall biases and outcomes and, importantly, addresses potential concerns relating to unobserved cross-sectional heterogeneity. In the absence of within-individual variation, it would be difficult to disprove that individuals with lower ability and effort are the ones most likely to receive poor performance signals and fail to recall or pay attention to them. The results in Section 3.1 offer some reassurance to existing field research on motivated beliefs since, at least in this context, individual heterogeneity is largely inconsequential for the observed relationship between recall accuracy and signal content. Finally, in contrast to settings like Huffman et al. (2022)'s workplace tournament — in which participants are highly self-selected — this paper studies recall biases in the more universal context of schooling and academic ability.

Many theories of belief-based utility, including Bénabou and Tirole (2002) and Köszegi (2006), accommodate biases in both the acquisition and retention of information. The distinction between selective attention and selective memory is of critical importance in practice, especially in understanding how socially valuable but negative information affects beliefs over time. The notion that beliefs can change over time, notably in the absence of any new information, presents a challenge to broad swathes of economic research and policy which typically assumes that beliefs stay fixed. Until recently, empirical research was silent on whether motivated signal processing tends to occur at the time of signal transmission or afterwards, via memory loss. As previously highlighted, an idiosyncrasy of the survey design offers an opportunity to study this phenomenon in the field: the

Fall edition of the survey was administered a few months after report cards, while the Spring edition was typically delivered with just a month's delay. Zimmermann (2020) conducts a similar analysis in the laboratory and also supports the existence of motivated memory loss. However, it leaves open the question of whether earlier signals about ability in the field, such as those received during school, selectively decay over time in a similar fashion. The incentives to retain signals about ability may be very different when prior beliefs are weaker and signals contain more information, as is the case during schooling years.

The final contribution of this paper is its use of longitudinal variables to examine the links between recall distortion, beliefs and outcomes, an analysis which has not been feasible in previous empirical work. Owing to this lack of empirical study, the purpose of biased belief updating is not yet well understood, while theory offers competing explanations: in Brunnermeier and Parker (2005), beliefs increase felicity through anticipatory utility, while in Bénabou and Tirole (2002) and Compte and Postlewaite (2004) they serve to self-motivate or improve physical and psychological performance. Controlling for actual grades, individual fixed effects and lagged dependent variables, we find that recalled grades have a positive relationship with academic self-assessments, enjoyment of school, teacher-rated enthusiasm, expected grades and future actual grades, but negatively with time spent on homework. These conclusions suggest that individuals may be served well by belief distortion, consistent with the idea that it is a (subconscious) choice variable. We also present descriptive evidence that recalling higher grades is associated with subjects having more favourable beliefs in their mathematics ability around a decade later, after they have left school.

One concern with the use of survey data to measure recall is that responses could be dishonest. Incentivisation does not provide an easy solution given that the most basic feature of motivated beliefs models is the responsiveness of beliefs to incentives: in addition to mitigating dishonesty, incentives for accuracy create measurement error by attenuating any intrinsic incentive to suppress bad memories.[2] Dishonesty cannot convincingly explain the positive drift in recall of grades as the survey is administered after a longer delay. As for the basic result that recall is less accurate for worse grades, several findings indicate that dishonesty plays at most a small quantitative role: while a minority of students rated as particularly dishonest by their teachers do display unusual recall behaviour, the 73% of the sample rated as 'not at all' dishonest show almost identical asymmetries

[2]In fact, biased recall essentially disappears when incentives are very high (Zimmermann 2020).

in recall to those in the main results. Furthermore, recalling a higher grade is associated with improvements in a range of subjective and material outcomes; actual grades are overwhelmingly the most important determinant of within-student variation in recalled grades; and incorrect recall is common (23%) in the Fall survey even when students achieve the highest possible grade, for which only unflattering errors are possible.

In addition to populating a scarce bank of empirical evidence on motivated beliefs, this paper also contributes to a more recent research agenda within the economics of education. Economists have exerted considerable effort to understand gaps in educational attainment and why they widen over schooling years (Heckman et al. 2006). This literature has begun to allude to the possibility of beliefs in ability detaching from reality, particularly in a recent focus on self-confidence as a non-cognitive skill (Cunha and Heckman 2007; Saltiel 2022). However, relatively little is understood, especially empirically, about how self-confidence develops and responds to positive and negative feedback.[3] The results in this paper indicate that motivated protection of self-confidence occurs in early life, when beliefs about general ability are starting to be formed (Dweck 2002). Most laboratory studies, and Huffman et al. (2022)'s in the field, concern narrower, domain-specific ability since credible signals about general ability are harder to provide later in life.

This paper also enriches an emerging literature focusing more closely on feedback in schools. Dizon-Ross (2019) examines how uncertainty about children's academic abilities distorts investments in human capital in the context of a field setting. She demonstrates that poorer parents face greater informational frictions in assessing their children's abilities. Kinsler and Pavan (2021) find evidence that parents' beliefs about ability are distorted by local grade averages. Other authors focus more closely on children's own beliefs: Alan et al. (2019) conduct a field intervention on "grit", demonstrating the malleability of children's beliefs on the relationship between effort and skill; Bursztyn et al. (2019) and Falk et al. (2020) illustrate the importance of the social environment in shaping children's attitudes towards self-assessment. The results in this paper are consequential for how schooling feedback interventions, such as in Dizon-Ross (2019), should be modelled and evaluated. They could have highly heterogeneous, dynamic impacts if some students are more able to inure themselves to negative feedback than others through biased memory loss.

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

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

The backdrop of the BSS was one of striking educational disadvantage. When the study began, Baltimore was in the midst of a major economic decline which profoundly affected most of the students sampled (Alexander, Entwisle, and Olson 2014). These challenges are starkly visible in the demographic statistics in Table 1: 61% of responding parents in the sample did not finish high school, 28% were non-employed sole earners, and 25% were the sole resident parent. In spite of the dispiriting context, parents were relatively optimistic about their children's educational prospects: in the first run of the survey, 98% expected their children to finish high school. Only around 65% eventually did so without initially dropping out.[5]

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. It then made attempts to track respondents after they left school and began adult life, once in 1998 (with most respondents aged around 22) and again in 2006. It also surveyed parents and teachers, collecting subjective variables relating to students and other educational matters, but less frequently than students themselves. We make use of the follow-up and teacher surveys in an analysis of the effects of recalled grades in Section 4, but the bulk of the

[4] Alexander and Entwisle (2003).

[5] Based on a follow-up survey in 1998–2002.

Table 2: Grade interpretation

| Grade | Mark |
|---|---|
| *Excellent* | 90–100% |
| *Good* | 80–89% |
| *Satisfactory* | 70–79% |
| *Unsatisfactory* | <70% |

analysis focuses purely on students' own responses to surveys taken during schooling years.[6]

BSS survey sweeps repeatedly asked children what grade they remembered getting in their last quarterly report card. It collected information on grades in reading, science and mathematics. Since reading and science grades were less consistently recorded, we focus solely on mathematics grades in this paper. Survey waves were implemented twice a year, in Fall and Spring. In Fall (around November), children are asked to try to remember the grade they received at the end of the last academic year in June. In Spring (mostly in April or May), children were asked to remember the grade they got a few weeks earlier, in the third quarter.

Grade recall was elicited using the following question: *"Remember the last report card you got when school ended for the summer? You could have gotten marks like E (Excellent), G (Good), S (Satisfactory), or U (Unsatisfactory). What mark did you get in Mathematics?"* As such, the domain for recalled and actual grades is $\{E, G, S, U\}$; see Table 2 for a translation into numerical marks.[7]

Recall of grades is generally imperfect throughout the data (Table 3), even in the Spring edition of the survey collected roughly one month after the relevant grades. The fraught socioeconomic context of Baltimore City public schools is likely to have played some role in impeding information transmission (Alexander, Entwisle, and Bedinger 1994).[8] Recall accuracy is around 20 percentage points lower in Fall editions of the survey, which were collected after a much longer delay relative to the relevant grades: 3–5 months, compared to 1–2 months for the Spring edition. This presents

[6]It would also be of interest to examine parents' attitudes, but less frequent sampling means sample sizes can become rather small when the econometric design involves a dynamic component.

[7]Not all schools used the $\{E, G, S, U\}$ scale for actual grades. We use a conversion table in the BSS documentation to map actual grades to that scale. Students were asked to recall grades on the $\{E, G, S, U\}$ scale, but they were also provided with the percentage score mapping of each grade category in Table 2.

[8]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. It should also be noted that a delay of one month was sufficient for subjects in Zimmermann (2020)'s experiment to fail to recall past signals.

a valuable opportunity to explore the effects of the passage of time on recall bias, to be revisited in the analysis to follow. Notwithstanding the substantial seasonal variation, children's recall accuracy was largely stable over time, other than in fourth grade when they may have been too young to interpret report cards correctly. We thus omit Fall '85 from all of the ensuing analysis.

Table 3: Summary statistics

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

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

# 3 Analysis: Recall of Grades

We now proceed to examine patterns of grade recall demonstrated by BSS students. The first objective is to test the main prediction of motivated beliefs theory: that students' recall of grades is positively biased, in that recalled grades are generally higher than actual grades and recall errors are more likely for lower grades. The second is to investigate how much of a role is played by memory loss by testing whether recall of grades is more positive when the survey is delivered with a greater delay.

## 3.1 Recall By Grade

It is clear from the summary statistics in Table 3 that recall of grades was generally imperfect in the BSS. Theories of motivated beliefs go further in implying that recall distortion is used as a tool

to gain, or maintain, a desired belief. In this context, recalled grades are a tool in a subconscious optimisation process.

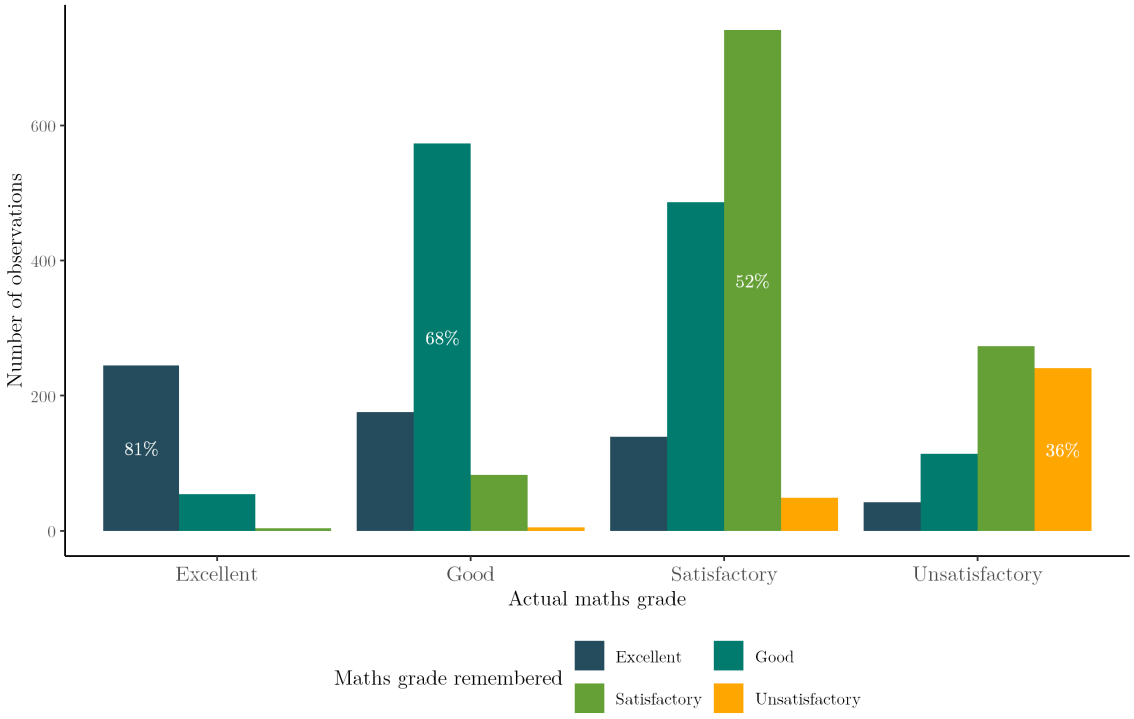
On one side of the trade-off is the incentive to preserve positive self-image by distorting or ignoring incoming bad news. Empirically, academic self-confidence has been associated with greater happiness in adolescents (Cheng and Furnham 2002): it embodies a hedonic benefit. More recent work in economics also outlines an instrumental benefit of higher self-confidence (Cunha and Heckman 2007; Cunha, Heckman, and Schennach 2010; Alan et al. 2019): it motivates students to work harder or even enhances performance directly. Later (in Section 4), we provide evidence of both benefits in the short run.

On the other side of the trade-off are the costs of distortion: first, information suppression or avoidance may incur a psychological and practical burden, and second, students face material incentives to hold correct beliefs (Brunnermeier and Parker 2005). In the context of schooling, excessively distorted beliefs could lead to students avoid making costly mistakes in choices over tasks and effort, as well as more significant decisions on whether to remain in education. This trade-off means recall should be selectively distorted, when it has a more negative impact on the individual. In this section, after verifying that recall errors are overwhelmingly flattering when they do occur, we follow most of the empirical literature by simply examining whether the probability of a given student making a recall error differs across good and bad grades.

Concretely, the course of events leading to recall of a grade ensues can be envisaged as follows. The opportunities for information avoidance and suppression begin when report cards are transmitted to students, who could ignore their report card if anticipating its poor content via other informal signals, or even avoid fully digesting information while seeing it (Gabaix et al. 2006). After that initial assimilation of information comes the subconscious decision to retain it, and it is at this point that the decay of information via memory loss becomes more important. Both of these mechanisms, contemporaneous and retrospective, culminate in measured recall errors. While 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 provides a clear indication of biased recall in the BSS data. It plots actual grades received in maths and splits them by the grade the child remembered. The first observation is that recall errors are positively skewed, other than for the highest grade (for which upward errors are

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




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

impossible): children generally remembered getting better grades than they actually did. Indeed, for the middle two grades, *Good* and *Satisfactory*, the probability of making a positive recall error conditional on making an error is 85%. The second observation, and the one that we take forward into the longitudinal analysis, is that errors become increasingly prevalent the lower the actual grade is.

Based on the fact that recall errors are overwhelmingly positive when they do occur, we can conduct a longitudinal analysis using a regression whose dependent variable is simply the event that a student incorrectly recalls their grade.[9] If we write actual grades as $m_{it}$ and recalled grades as $\tilde{m}_{it}$, then incorrect recall is the event $\mathbb{1}(\tilde{m}_{it} \neq m_{it})$. The principal aim is to check whether the relationship between recall accuracy and the qualitative content of grades, visible in the raw data in Figure 2, is robust to confounding variables. Arguably the most problematic of those here are cognitive ability and engagement with schooling, which one might expect to be correlated

[9]This dependent variable also permits negative recall errors. We include 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*, but the recall accuracy of those achieving both the highest and lowest grades are still worth studying.

with both grades, $m_{it}$, and the event that they are incorrectly recalled, $\mathbb{1}(\tilde{m}_{it} \neq m_{it})$. Since the BSS observed students longitudinally, we can make major progress in addressing these concerns by including individual fixed effects in the analysis. Intuitively, that means we examine whether the same individuals are more likely to forget lower grades than higher ones. We exclude Fall '85, when most students were in $4^{th}$ grade and recall was very poor (Table 3), from the sample. The regression sample therefore largely follows the cohort while in their early adolescence (mostly aged 11–15), passing through school grades 6–9 and 12. We use a linear probability model with robust standard errors,

$$\mathbb{1}(\tilde{m}_{it} \neq m_{it}) = \beta_0 + \sum_{z \in \{G,S,U\}} \beta_{1z} \mathbb{1}(m_{it} = z) + \beta_2 Fall_t + \beta_3 Year_t + \alpha_i + \epsilon_{it}. \quad (1)$$

As in most of the existing empirical literature, $\sum_{z \in \{G,S,U\}} \beta_{1z} \mathbb{1}(m_{it} = z)$ in Equation (1) establish a good news-bad news effect: they measure the relationship between the probability of incorrectly recalling a grade and the quality of the grade itself. $Fall_t$ is a dummy variable for the Fall edition of the survey, in which recall is regularly much less accurate. $\alpha_i$ is an individual fixed effect, and $Year_t$ is a survey-year effect.

Like 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. Noting that *Excellent* is the omitted category, all grades are less likely to be remembered than *Excellent*, and each grade down from *Excellent* has a progressively lower probability of being correctly recalled. Students were 10 percentage points (pp) less likely to recall *Good* than *Excellent* grades, but almost 50 percentage points (pp) less likely to recall *Unsatisfactory* than *Excellent*. As already indicated, incorrect recall is substantially more likely in the Fall edition of the survey, which was delivered much later relative to report card grades than the Spring edition.

In addition to the basic analysis, it is also of interest to test whether students with apparently higher stakes — either via higher prior beliefs in their ability, meaning that they could have more to lose from bad news (Bénabou and Tirole 2002), or simply a greater preference for mathematics — demonstrate larger asymmetries in recall. This hypothesis finds support from columns (3) and (5) in Table 4, which include an interaction of actual grades with, respectively, an indicator of high prior self-confidence and an indicator that mathematics was the student's favourite subject prior to

Table 4: Recall of mathematics grades

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

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. High SC indicates the student believes they are at least smarter than the average child in their year and school. The omitted grade category is *Excellent*.

the survey in question.[10] For both sets of students, the gap between recall of *Excellent* grades and all the others is wider: these students were more likely to correctly recall getting *Excellent* grades and less likely to correctly recall the others. In the absence of higher stakes (when a preference for maths was not expressed, or when initial beliefs in ability were lower) students were equally likely to remember *Good* and *Excellent* grades, but still demonstrated asymmetric recall patterns for high and low grades.

The results in column (5) also help to differentiate between models of motivated beliefs, in which information is selected in order to maintain self-confidence, and alternatives. The most prominent of those alternatives is confirmation bias (Rabin and Schrag 1999), in which individuals are more likely to misinterpret signals which do not accord with their existing belief, regardless of whether those signals convey good or bad news about the ego.

Under a model of confirmation bias, the expectation would be for students with poor initial beliefs about their ability to prefer further negative signals. The results in column (3) indicate that this is unlikely to be the case. The coefficients on the actual grades $\{Satisfactory, Unsatisfactory\}$ demonstrate that even students who believe they are of average or, very rarely, below average ability are much more prone to forgetting bad grades than good ones.[11]

Each of the even-numbered columns in Table 4 omits individual fixed effects from the specification to its left. In all cases, the results undergo little qualitative change other than some small differences in relative magnitudes. This conclusion provides some reassurance to existing field research lacking in longitudinal variation, since it indicates that fixed unobserved individual characteristics, such as cognitive ability or engagement with schooling, are not an important joint determinant of the quality of signals and the ability or willingness to recall them correctly.

We also examine gender and racial differences in recall asymmetries (Table A1). Surprisingly, there are no differences between boys and girls. On the other hand, Black students display a greater propensity to recall information asymmetrically than others: relative to *Excellent* grades, incorrect recall of *Good* and *Unsatisfactory* is more likely than for other students.

[10] Both variables are measured in the last *available* period prior to $t$ for each student to minimise data loss.
[11] Only 7% of students believe they have below average ability for their school and age.

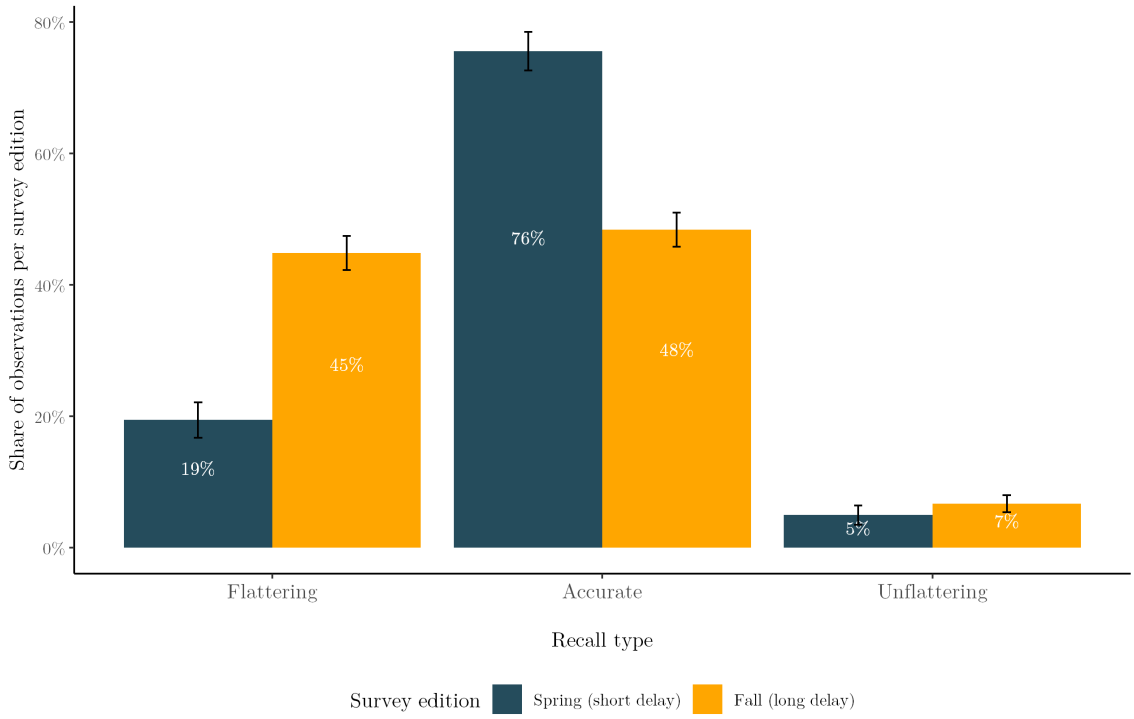
### 3.2 The Role of Memory Loss

It is well known that observed biases in recall alone cannot be used to infer biases in memory loss. The findings in Table 4, much like those in existing empirical work measuring ex-post recall, could reflect selective signal avoidance: since students receive a stream of signals on their performance through the academic quarter, they could avoid reading a report card grade if anticipating its unfavourable contents. And even if report cards were unavoidable, students could likely leverage mechanisms like directed cognition (Gabaix et al. 2006) to avoid absorbing unfavourable information in the moment. However, the design of the BSS presents a compelling opportunity, closely mirroring the main exercise in Zimmermann (2020)'s laboratory experiment, to investigate whether memory loss serves the cause of motivated beliefs.

In both editions of the survey within each year, Spring and Fall, adolescents were asked to remember their most recent grades. In Spring those grades were likely from just over a month earlier (in the second quarterly report card), but in Fall the most recent grades were likely received more than three months ago, prior to the summer holidays. The substantial differences in recall between Fall and Spring (Table 3) can thus be taken to reflect memory loss. The analysis in this section establishes that, since patterns of recall are markedly more motivated (positively biased) in Fall than in Spring, a large part of the information selection underlying motivated beliefs in this context appears to be enabled by biased memory loss.

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

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.

avoidance.[12] Nonetheless, Figure A3 indicates that recall accuracy deteriorates (from 91% to 77%) even for students receiving the top grade, *Excellent*, for which only unflattering recall errors are possible; self-deprecating memory loss does seem to occur, although it is far less frequent.

The next step is to verify that this positive drift in memory is identified in an analysis where unobserved cross-sectional heterogeneity can be neutralised. In this case, within-student variation can be used to verify that when the same student is sampled in Fall, after a longer delay relative to grades, their memory tends to be more flattering. Since the objective is now to examine whether memory becomes more optimistic over a longer horizon, given the actual signal received, we use $\mathbb{1}(\tilde{m}_{it} = \textit{Excellent})$ — the event that the student recalls getting an *Excellent* grade — as the dependent variable, controlling for the actual grade they received, $m_{it}$,[13]

$$\mathbb{1}(\tilde{m}_{it} = E) = \delta_0 + \sum_{z \in \{G,S,U\}} \delta_{1z} \mathbb{1}(m_{it} = z) + \delta_2 Fall_t + \delta_3 Year_t + \alpha_i + \epsilon_{it}. \tag{2}$$

The statistically significant and positive estimate for $\delta_2$, the coefficient on $Fall_t$ in column (1) of Table 5, implies that being surveyed in Fall has a positive impact on students' propensity to remember getting the highest grade. This formalises the positive drift in memory illustrated in raw data in Figures 2 and A3, and demonstrates its robustness to a within-individual identification strategy. That is, the same students are more likely to remember the top grade when surveyed in Fall than in Spring, accounting for the grade they actually achieved in each case.

As in Section 3.1, we can check whether the generally positive drift in memory over time is different for children who express a greater preference for mathematics or have higher prior self-confidence. This is achieved by interacting $Fall_t$ with those two variables, as in specifications (2) and (3) in Table 5. Once again, we find strong evidence that both groups of students exhibited stronger motivated information processing: in this case, they were more prone to an upward drift in their memory of grades as time elapses. Combined with the heterogeneity in general recall in Table 4, these results suggest that those with higher initial confidence and enthusiasm were the

[12] It should be noted that in the raw data (see Figure A3 in Appendix A) is that even in the Spring edition, recall is worse for the lower grades, *Satisfactory* and especially *Unsatisfactory*. Given the relatively short delay for the Spring survey, of just over a month, this could support some role for selective attention or information avoidance. That being said, Zimmermann (2020) successfully uses a delay of one month to measure the effects of memory loss in his laboratory test of the same phenomenon.

[13] Another option would be to examine whether the probability of making a positive recall error increases in Fall, as in Figure 2. However, as in Section 3.1, this strategy would require us to exclude all cases in which the top and bottom grades were received.

Table 5: Spring/Fall recall comparison

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Recall: *Excellent* | | | | | |
| | Base | Base | High SC | High SC | Subject preference | Subject preference |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Actual: *Good* | −0.598*** | −0.600*** | −0.573*** | −0.631*** | −0.598*** | −0.604*** |
| | (0.035) | (0.030) | (0.044) | (0.035) | (0.035) | (0.030) |
| Actual: *Satisfactory* | −0.680*** | −0.724*** | −0.640*** | −0.735*** | −0.676*** | −0.715*** |
| | (0.035) | (0.027) | (0.045) | (0.033) | (0.035) | (0.027) |
| Actual: *Unsatisfactory* | −0.691*** | −0.741*** | −0.655*** | −0.759*** | −0.685*** | −0.725*** |
| | (0.037) | (0.028) | (0.046) | (0.033) | (0.037) | (0.028) |
| Fall | 0.084*** | 0.084*** | 0.052*** | 0.057*** | 0.053*** | 0.057*** |
| | (0.013) | (0.013) | (0.018) | (0.019) | (0.015) | (0.014) |
| High SE $(t-1)$ | | | 0.002 | 0.022 | | |
| | | | (0.020) | (0.014) | | |
| Fall*High SE $(t-1)$ | | | 0.074*** | 0.062** | | |
| | | | (0.028) | (0.029) | | |
| Maths pref $(t-1)$ | | | | | −0.002 | 0.042*** |
| | | | | | (0.019) | (0.013) |
| Fall*Maths pref $(t-1)$ | | | | | 0.066*** | 0.059*** |
| | | | | | (0.023) | (0.023) |
| Individual FE | Yes | No | Yes | No | Yes | No |
| Academic year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,694 | 2,694 | 2,148 | 2,148 | 2,639 | 2,639 |
| $R^2$ | 0.541 | 0.330 | 0.559 | 0.337 | 0.535 | 0.345 |
| Adjusted $R^2$ | 0.413 | 0.328 | 0.416 | 0.333 | 0.411 | 0.342 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. High SC indicates the student believes they are at least smarter than the average child in their year and school. The omitted grade category is *Excellent*.

ones most likely to inure themselves to incoming negative feedback, and that selective memory may have been an important tool in facilitating this defence mechanism. Notably, though, even students with lower prior beliefs in their ability experienced a significant positive drift in memory over time, demonstrated by the positivity of the coefficient on $Fall_t$ in column (3). This corroborates a similar finding in Zimmermann (2020), and, matching his interpretation, further strengthens the suggestion that the results reflect motivated memory loss rather than, for example, a drift of memories towards prior beliefs.

Once again, we examine heterogeneity along sex and race dimensions (Table A2). In this case, we find no evidence of differences in the effect of memory loss across 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 Dishonesty

A perennial issue with survey data is that it is generally impossible to establish that reporting of subjective variables is truthful. In this case, a concern is that students could have exaggerated grades to avoid embarrassment. On the premise that biases in recall are *motivated*, incentives for accuracy are not a panacea: while incentives may mitigate dishonesty, they would also disrupt the natural (subconscious) trade-off between ego enhancement and the costs of belief distortion. Indeed, Zimmermann (2020) finds evidence that recall biases are essentially eliminated by sufficiently large incentives for accurate recall, suggesting that bad memories are suppressed rather than erased.

It is unlikely that dishonesty plays a major role in the results on memory loss in Section 3.2: we would not expect students to become more dishonest when recalling grades further back in time. Indeed, errors become more likely even when the top grade, *Excellent*, is received (Figure A3). And, since biased memory loss plays a dominant role in the *level* of bias over time, this also provides some reassurance to the interpretation of the results in Section 3.1. Later on, in Section 4, we will also see that recalling higher grades is associated with positive changes in a range of subjective and objective variables, and that actual grades are overwhelmingly the most important determinant of within-student variation in recalled grades.

We can conduct closer analysis of the role of deceit using a direct indicator of students' honesty: teachers' ratings in response to the prompt 'For each statement, check under the appropriate

column, how much like that the child is' — 'Cheats; tells lies; is deceitful'.[14] The small proportion of students (approximately 7% of available observations) assessed by their teacher to be 'pretty much', 'very much' or 'exactly' like this do demonstrate much larger asymmetries between recall of *Excellent* grades and all others than other students.[15] These students are particularly distinct in almost always recalling *Excellent* grades regardless of their actual grade. However, we observe asymmetries in recall very similar to those seen in the basic results (Table 4) among the remaining students. Notably, the results are not qualitatively affected by restricting the sample to just those students rated as 'not at all' dishonest, covering around 73% of observations (Table A5). Conducting the same exercise for the results on the Fall/Spring comparison, Table A6 supports our intuition in identifying no difference in the patterns of biased memory loss across honesty categories, even the small share explicitly rated as dishonest by their teachers.

Overall, the data suggest that dishonesty plays at most a minor role in the magnitude of the results on biased recall in Section 3.1. It also seems to be irrelevant for the results on memory loss in Section 3.2, in line with our intuition.

# 4 Analysis: Possible Motives For Recall Distortion

Having established patterns of biases in recall, we now take advantage of the breadth of the BSS to analyse how recalled grades are associated with the evolution of a range of outcomes, both subjective to the student and material.

## 4.1 Short-Run Outcomes

We begin by analysing the connection between students' recall of grades and a set of short-run outcomes.[16] In doing so, we can establish suggestive evidence on the possible motives for recall distortions. For instance, some models suggest that self-enhancing recall enhances current utility through either a direct channel (Köszegi 2006) or anticipated outcomes (Brunnermeier and Parker

[14]It should be noted that, in principle, it is unclear that teacher-rated dishonesty is only relevant to misreporting: students rated as more dishonest could also face lower cognitive costs of memory distortion. For this reason, we classify this analysis as a robustness check and do not exclude these students from the main results.

[15]The interaction terms are no longer significant if this category is expanded to also include the larger proportion of students who are rated as only being 'somewhat' like this.

[16]While the BSS contains a very broad range of possibly relevant variables, relatively few — those analysed in this section — met the criterion of being elicited in consecutive rounds with consistent definitions.

2005). Others feature self-deception as a means of overcoming motivation problems resulting from time inconsistency (Bénabou and Tirole 2002), or performance issues from low confidence (Compte and Postlewaite 2004). The analysis presented here finds some support for both channels: recall of higher grades is associated with more favourable self-beliefs, utility and future grades, and the measured effect on utility appears to act through anticipated impacts on future grades.

In analysing the effects of recalled grades on a range of outcomes, it is important to understand where within-individual variation in recall (conditional on actual grades) comes from. Theoretical work on belief-based utility suggests a complex array of factors in the choice and belief environment could play a role, including the consumption value of self-esteem, the marginal value and cost of effort, social networks, the costs of overconfidence, or exogenous shocks to confidence from life events. As a result, it is likely that recall is endogenous to most of the subjective variables analysed in this section. However, we propose that endogeneity presents less of an obvious concern here than is typical for two reasons. First, it is difficult to envisage a source of exogenous variation in recall which would not also directly affect a wide array of other psychological variables. It is therefore informative simply to know how these variables are associated with recall, rather than attempting to answer the more amorphous question of how recall itself 'causes' outcomes. Second, the analysis is able to leverage the longitudinal content of the data by conditioning on both lagged dependent variables and individual fixed effects. In order to do so, we estimate Bartolucci and Nigro (2010)'s Quadratic Exponential (QE) model on all variables of interest, with recalled grades as the most important covariate throughout. Owing to this strategy, the analysis addresses the more mundane statistical concern that variation in recall biases is correlated with existing dynamics of the outcome variable or with cross-sectional heterogeneity.

In Equation (3), $y_{it}$ is a binary outcome variable, and $x_{it}$ is a vector of covariates (generally including both recalled and actual grades, as well as year effects). The QE model then implies a conditional likelihood of the following form for $y_{it}$:

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

where, for $t < T$,

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

and

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

It is easier to understand the QE model using the structural equation describing the latent variable $\theta_{it}$, satisfying $y_{it} = \mathbb{1}(\theta_{it} > 0)$:

$$\theta_{it} = \alpha_i + x'_{it}\beta + y_{i,t-1}\gamma + e_{t+1}^*(\alpha_i, X_i) + \epsilon_{it}, \tag{6}$$

where $\epsilon_{it}$ are independent with a logistic distribution and $\alpha_i$ are individual fixed effects. Relative to standard dynamic models, the most unconventional aspect of Equation (6) is the error correction term $e_{t+1}^*(\alpha_i, X_i)$. However, theories of belief-based utility provide a natural fit for this element of the specification, assuming some forward-looking sophistication on the part of students. The presence of $e_{t+1}^*(\alpha_i, X_i)$ implies current outcomes are enhanced by the anticipated impact of $y_{it} = 1$ on the likelihood of $y_{i,t+s} = 1$ for $s > 0$, as detailed by Equation (4). This formulation is reminiscent of models of anticipatory utility, such as that forming the basis of Brunnermeier and Parker (2005), in which current utility is directly affected by expectations of future outcomes.

#### 4.1.1 Outcome Variables

We consider the following categories of outcome variable, in each case outlining its theoretical significance:

**Future recalled grades**

This analysis serves a dual purpose. First, we investigate whether changes in recall (given actual grades) persist over time within individuals, suggesting that belief distortion behaviours are retained through time. One indication that this could be the case is from Sections 3.1 and 3.2, which indicate that students with higher self-confidence are subsequently more likely to recall higher grades. Second, by establishing that actual grades have an important influence on recalled grades,

(conditional on past recall and individual fixed effects), we address one potential concern with the preceding analysis: that students employ simple response strategies such as simply recalling the same grade in every period.

**Future actual grades**

This outcome variable provides an important test of a subset of models of belief-based utility. Notably, there need not be a systematic positive link between current beliefs and future outcomes in models such as Brunnermeier and Parker (2005), in which belief distortions serve the purpose of increasing current felicity at the cost of *distorting* actions, which can induce worse future outcomes. For example, overconfidence could result in lower future grades by encouraging students to exert lower effort than they would otherwise (in the absence of an anticipatory utility motive to distort beliefs). In contrast, mechanisms such as those in Bénabou and Tirole (2002) and Compte and Postlewaite (2004) suggest that current belief distortions are essentially strategic and should result in more favourable future outcomes by inducing higher effort or confidence-enhanced performance respectively. Students may be aware of positive links between self-confidence and performance, either through their own experience or by learning from others, and subconsciously internalise those impacts by enhancing their self-confidence using biased recall.

**Beliefs**

- *How smart do you think you are compared to other kids in your school?*
- *Expected mathematics grades (in the next report card)*

The first variable captures general academic self-esteem, whereas the second captures expectations of performance in mathematics. In both cases, if biased recall is in use and if mathematics is relevant to the belief in question, a basic requirement of models of belief-based utility is that recalling a higher grade is associated with more favourable self-beliefs and expectations.

**Enjoyment of school**

- *Now, I'd like to find out how you feel about school in general. Do you like school a lot, think it's just O.K., or not like it much at all?*

In models of belief-based utility, recalling a higher grade generally increases contemporaneous utility. However, this is not always the case. Notably, in Bénabou and Tirole (2002)'s core model of motivated beliefs, nothing strictly prevents utility from being lower in the current period: beliefs have no intrinsic consumption value, but in the presence of time inconsistency, higher self-confidence induces higher *current* effort costs with the aim of maximising lifetime utility. Thus, if recalled grades are associated with higher enjoyment of school, it provides an indication that either self-beliefs have some intrinsic value to students, or that wellbeing is augmented by more optimistic beliefs about the future through anticipatory utility (as in Brunnermeier and Parker 2005). We check for the latter using supplementary analysis adding expected grades as another explanatory variable.

**Schooling effort**

- *How much time do you usually spend on homework each night?*

Analysis of this variable serves predominantly as a tentative test of the motivation channel in Bénabou and Tirole (2002). That is, given that recall is asymmetric, it should be the case that effort and recalled grades are positively associated. In their core model, Bénabou and Tirole (2002) propose that a primary driver of recall distortions is for the subconscious to coerce a present-biased self into overcoming its motivation problem. The essential basis of that link is a complementarity between ability and the return to effort. Of course, homework time investment is a highly incomplete measure of effort. We address this point by also analysing teacher ratings of classroom enthusiasm, as below.

**Teacher evaluations**

- *Very enthusiastic, interested in a lot of things, likes to express his/her ideas*
- *Can't concentrate, can't pay attention for long* (analysis uses reversed scale: 1 if better concentration, 0 otherwise)
- *Is awfully restless, fidgets all the time, can't sit still* (analysis uses reversed scale, so 1 if better concentration, 0 otherwise)

In measuring classroom enthusiasm and focus, these variables provide an additional dimension of measured effort relative to students' reports of time spent on homework. As is the case for reported homework effort, they provide some insight into the motivation and performance mechanisms proposed by Bénabou and Tirole (2002) and Compte and Postlewaite (2004): both demotivation and performance-limiting anxiety may be overcome by distorting recall and enhancing self-confidence. However, it should be noted that teacher evaluations were only measured in the Spring edition of the survey, when recall distortions were less significant. Most importantly, because sampling frequency is halved, statistical power is lower and we must use simple linear fixed effects regressions (dropping lagged dependent variables) in place of the QE model.

#### 4.1.2 Results

We begin by analysing dynamics in recalled and actual grades.[17] The regression on dynamics in recalled grades asks whether shocks to recall are persistent, and whether recalled grades anchor to actual grades in each period (as they should). For the second regression, the question is whether recalled grades are associated with *future* actual grades, and thus whether self-deception appears to serve students in terms of material as well as subjective outcomes. In both cases, we use the QE specification so that the coefficient estimates are conditional on individual fixed effects and a lagged dependent variable. As in a logistic regression, the coefficients have a log-odds interpretation.

Column (1) in Table 6 contains the results relating to the dynamics of recalled grades. The much larger coefficient on *current* actual grades than even prior recalled grades indicates that they provide a substantial anchor for recall in each period, and in turn provides reassurance that the basic results for asymmetric recall in Table 4 cannot be explained by students consistently recalling the same (flattering) grade. The second result is that recalled grades exhibit some first-order persistence. Finally, consistent with the results pertaining to memory loss in Table 5, recalled grades are higher in Fall.

Column (2) in Table 6 captures the relationship between *prior* recalled grades and actual grades. Actual grades exhibit even more significant autocorrelation than do recalled grades, suggesting

[17] In this set of regressions, all variables relating to maths grades are measured as binary rather than categorical: we specify $y_{it} = \mathbb{1}(\text{Actual} > \textit{Satisfactory})$. The reason for doing so is that the lagged counterpart of the dependent variable $y_{it}$ must be binary, so including a more granular covariate would arbitrarily penalise the lagged dependent variable for measurement error. Dummy variables are defined as 1 if $\in \{\textit{Excellent}, \textit{Good}\}$ because only 7% of students on average actually achieve the top grade, *Excellent*.

ability is more stable over time than self-deception. More notably, prior recalled grades have a positive association with actual grades, independent of prior actual grades and individual fixed effects. The size of the coefficient on prior recalled grades is economically significant, at around a third of the effect of prior actual grades. This result is compatible with theoretical accounts in which self-deception enhances performance or motivation, rather than having a purely ephemeral effect on subjective wellbeing. Tentatively, it thus provides more definitive support for models like Bénabou and Tirole (2002) and Compte and Postlewaite (2004), in which belief distortion helps to correct inefficiencies, than those like Brunnermeier and Parker (2005) and Köszegi (2006) in which beliefs are traded off against optimal actions because the former affect hedonic utility. In this context, belief distortion appears to be outcome-enhancing.

The next set of results pertains to variables capturing beliefs, subjective wellbeing and preferences (Table 7). Half of the variables are highly autocorrelated even when using within-individual variation and controlling for year effects, justifying the use of the QE model. Before proceeding to analyse the effects of recalled and actual grades, we pause to note that expected mathematics grades are systematically higher in Fall relative to Spring. Notably, this effect is estimated independently of recalled grades, which we have already demonstrated to increase in Fall relative to Spring. This likely reflects that the summer months prior to the Fall survey provides opportunities for self-enhancing memory loss over a much deeper set of signals than just the most recent mathematics grades, with the outcome being a generalised increase in optimism. Nonetheless, the previously observed positive drift in mathematics grade recall in Fall relative to Spring is robust to controlling for expected mathematics grades and current academic self-esteem, indicating memory loss in grade recall is largely independent of this separate process of enhanced optimism (Table A3).

Recalled grades have a positive and monotonic relationship with perceived ability and grade expectations, in columns (1) and (2) of Table 7, though with only the comparison between *Excellent* and *Unsatisfactory* being statistically significant in the latter case.[18] This first result provides some reassurance that students are using recalled grades to update their beliefs in some sense. Second, we have evidence from column (3) that students experience greater wellbeing at school

[18]There is also weakly significant evidence that actually receiving an *Excellent* grade is associated with *lower* perceived ability when holding recalled grades constant. This may reflect heterogeneity in the effect of recalled grades across actual grades. The group of students actually receiving *Excellent* grades is relatively small, at around 10% on average.

when recalling higher grades. This result is particularly evocative of models featuring either a consumption value of anticipatory utility or current self-confidence, as proposed by Brunnermeier and Parker (2005) and Köszegi (2006) respectively and empirically observed in Oster et al. (2013). Instructively, additional analysis indicates that all of the positive effect of recalled mathematics grades on current enjoyment of school passes through their effect on expected mathematics grades (Table A4), suggesting particular support for the anticipatory utility mechanism in Brunnermeier and Parker (2005): recalled grades are no longer related to enjoyment of school when expected grades are also controlled for. Finally, in column (4), recalling *Unsatisfactory*, the lowest possible grade, is significantly associated with more reported time spent on mathematics homework. This provides an suggestive contradiction of Bénabou and Tirole (2002)'s model, in which students should exhibit 'defensive pessimism' — a preference for negative signals — if effort and ability are substitutes. Instead, some dimensions of effort and perceived ability appear to be substitutes in this context, but students still exhibit self-enhancing biases in recall.

Table 8 reports the results for teacher ratings of students on a range of behavioural and attitudinal measures. Teachers were only surveyed in Spring, reducing the sample size to the extent that we can no longer use the QE model, and statistical power is vitiated. Nonetheless, the measures provide some additional value on the basis that they provide independent evaluations of students' behaviour and somewhat richer measures of effort. These measures are analysed using linear regressions with fixed effects in Table 8. Low statistical power encumbers the analysis, especially given its use of individual fixed effects, which may explain the result that recalled grades do not have a statistically significant relationship with any outcome other than in column (1). The results in column (1) provide weak evidence that students become less enthusiastic in the classroom when they recall grades lower than *Excellent* — only the comparison with *Good* is weakly significant, but this may reflect larger standard errors on the other coefficients. Nonetheless, it suggests that the evidence of a negative correlation between effort (as measured by homework effort) and recalled grades is not definitive. Combined with the previous finding on homework effort, it appears that some dimensions of effort could in fact increase in recalled grades while others decrease, so our overall sense is that the data are inconclusive on the motivation channel proposed by Bénabou and Tirole (2002).

Combined, the results in Tables 6 to 8 sharpen our understanding of the dynamic effects of

students' self-enhancing recall distortions. Most notably, recalling more flattering mathematics grades is associated with higher actual grades in the following period: the likelihood of *obtaining* a grade higher than *Satisfactory* is positively associated with recalling a grade higher than *Satisfactory* in the previous period, conditional on the grade actually received at that point (Table 6). Using the subjective outcome variables in Tables 7 and 8, we also gain some insight into the short-term welfare implications of recall distortions. Recalling higher grades is associated with higher perceived ability and future expectations of mathematics grades, as well as higher overall enjoyment of school. While the first two results reflect that recalled information is likely used to update beliefs, the latter suggests that there may be an impact on instantaneous subjective wellbeing. This is suggestive of either the 'ego utility' of beliefs in models like Köszegi (2006), in which self-esteem features directly in utility, or a model of anticipatory utility such as Brunnermeier and Parker (2005), in which more optimistic expectations of the future enhance current well-being. Tentative evidence of the latter is provided by adding expected mathematics grades as a covariate in this analysis: the effect of recalled grades on current enjoyment of school appears to be mediated by their effect on expected grades (Table A4).

While a more definitive conclusion on the mechanisms underlying performance improvements cannot be drawn without richer measures of effort intensity, we observe mixed results on how recalled grades are associated with effort: time spent on homework weakly decreases in recalled grades, while teacher-rated enthusiasm weakly increases. As such, the results leave open the possibility of both the performance-enhancing effect of self-confidence in Compte and Postlewaite (2004) and the self-motivation channel proposed by Bénabou and Tirole (2002) as drivers of the positive relationship between recalled and future grades. A fruitful direction for future work could be to make a sharper distinction between these channels in a more controlled setting.

Table 6: Effect of recall — recalled and actual grades

| | $y_{it}$ | |
|---|---|---|
| | Recall > *Satisfactory* | Actual > *Satisfactory* |
| | (1) | (2) |
| Actual > *Satisfactory* | 2.982*** | |
| | (0.393) | |
| Actual > *Satisfactory* $(t-1)$ | | 1.518*** |
| | | (0.162) |
| Recall > *Satisfactory* $(t-1)$ | 0.716*** | 0.476*** |
| | (0.270) | (0.157) |
| Fall | 1.572*** | −0.103 |
| | (0.277) | (0.147) |
| Year effects | Yes | Yes |
| Observations | 1,698 | 2,324 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Quadratic exponential model; coefficients are log-odds. Standard errors in brackets. Actual grades are measured at the time of the survey, so we skip intermediate quarters between Fall (Q4) and Spring (Q2).

Table 7: Effect of recall — self-esteem and preferences for school

| | $y_{it}$ | | | |
|---|---|---|---|---|
| | Academic self-esteem | Expected maths grade | Likes school a lot | >10 mins on maths hw |
| | (1) | (2) | (3) | (4) |
| Recall: *Good* | −0.338 | 0.354 | −0.321 | 0.129 |
| | (0.349) | (0.338) | (0.372) | (0.315) |
| Recall: *Satisfactory* | −0.742* | −0.392 | −0.506 | 0.439 |
| | (0.439) | (0.361) | (0.438) | (0.372) |
| Recall: *Unsatisfactory* | −1.501** | −0.801* | −1.338** | 1.006** |
| | (0.615) | (0.470) | (0.653) | (0.510) |
| Actual: *Good* | 0.875 | 0.577 | 0.384 | −0.246 |
| | (0.541) | (0.570) | (0.576) | (0.466) |
| Actual: *Satisfactory* | 0.923 | 0.346 | 0.107 | −0.076 |
| | (0.580) | (0.571) | (0.637) | (0.502) |
| Actual: *Unsatisfactory* | 1.080* | 0.480 | 0.156 | −0.411 |
| | (0.656) | (0.612) | (0.720) | (0.576) |
| Fall | 0.301 | 0.564*** | 0.351 | −0.063 |
| | (0.244) | (0.184) | (0.250) | (0.207) |
| $y_{i,t-1}$ | 0.663*** | 0.491** | 0.157 | 0.336 |
| | (0.257) | (0.206) | (0.275) | (0.214) |
| Year effects | Yes | Yes | Yes | Yes |
| Observations | 1,699 | 1,887 | 1,703 | 1,695 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Quadratic exponential model; coefficients are log-odds. Standard errors in brackets. The omitted grade category is *Excellent*. All dependent variables are self-reported and measured in period $t$ unless otherwise stated. Perceived ability measured as the student stating that they believe they are at least above average relative to other children in their school and year. Homework time investment is reported for the typical day.

Table 8: Effect of recall — teacher evaluations

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Enthusiastic | Concentrates | Not restless | Expected grade |
| | (1) | (2) | (3) | (4) |
| Recall: *Good* | −0.253* | −0.029 | −0.013 | −0.009 |
| | (0.134) | (0.103) | (0.072) | (0.142) |
| Recall: *Satisfactory* | −0.182 | −0.072 | −0.063 | −0.053 |
| | (0.152) | (0.115) | (0.092) | (0.149) |
| Recall: *Unsatisfactory* | −0.277 | 0.004 | 0.069 | −0.082 |
| | (0.178) | (0.151) | (0.123) | (0.163) |
| Actual: *Good* | 0.121 | 0.082 | 0.096 | −0.117 |
| | (0.149) | (0.113) | (0.079) | (0.160) |
| Actual: *Satisfactory* | −0.130 | 0.016 | 0.087 | −0.381** |
| | (0.165) | (0.118) | (0.099) | (0.171) |
| Actual: *Unsatisfactory* | −0.116 | −0.149 | −0.088 | −0.508*** |
| | (0.189) | (0.147) | (0.127) | (0.181) |
| Student FE | Yes | Yes | Yes | Yes |
| Academic year FE | No | No | No | No |
| Observations | 828 | 815 | 819 | 704 |
| $R^2$ | 0.661 | 0.672 | 0.680 | 0.801 |
| Adjusted $R^2$ | 0.219 | 0.231 | 0.253 | 0.489 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. The omitted grade category is *Excellent*. Variable (1) is measured as the teacher rating the student as being 'pretty much', 'very much' or 'exactly like' the statement 'Very enthusiastic, interested in a lot of things, likes to express his/her ideas'. Variable (2) is measured as a rating of 'somewhat', 'a little' or 'not much' like the statement 'Can't concentrate, can't pay attention for long'. Variable (3) uses the same coding as (2) but for the statement 'Is awfully restless, fidgets all the time, can't sit still'. In all cases, mathematics teacher ratings were used unless only the verbal teacher's ratings were available.

### 4.2 Long-Run Outcomes

The BSS also conducted a follow-up survey with subjects in a single round over 20 months from 1998 to the beginning of 2002, after normal schooling years were over and they were mostly aged around 22–25. They were asked about their current circumstances and beliefs, with a similar thematic focus to the school-time surveys. In this section, we briefly examine correlations between recalled grades, actual grades and long-term outcomes. Since longitudinal variation is unavailable, the analysis is vulnerable to bias from unobserved variables and is intended to be seen as descriptive.

The simple strategy taken here is to take recalled and actual grades from one sweep of the school survey and examine how they correlate with a range of long-run outcomes. We select Fall '87, the earliest period used in the preceding analysis, because most students had not dropped out of school by that point. Material outcomes in 1998–2002 are not well predicted by either recalled or actual grades in Fall '87, with the exception of the probability of finishing high school being strongly predicted by actual mathematics grades (Table 9). However, recalled grades during school have more predictive power for subjective variables (Table 10): recalling *Satisfactory* or *Unsatisfactory* in Fall '87 predicts a 20 percentage point lower probability of a subject believing they are 'good' or 'very good' at mathematics around a decade later, controlling for actual mathematics grades in the same period as in column (1). Notably, actual grades are similarly predictive of perceived mathematics ability, likely because they capture actual underlying mathematics ability. Also notable is the fact that recalling an *Unsatisfactory* grade in Fall '87 is associated with a lower long-run self-efficacy score, measured as the likelihood of disagreeing with the prompt 'People like me don't have much of a chance to be successful in life'. Actual grades in Fall '87 are positively associated with life satisfaction and expected professional success, in columns (3) and (4). Recalled grades have no predictive power for either outcome, except for a positive coefficient on recalling *Good* for expectations of professional success. It is unclear why this positive effect would not exist for recalling *Excellent*, likely reflecting the challenges of estimating long-run relationships between the variables in question using only cross-sectional data.

Table 9: Long-run analysis (Fall '87 grades and recall) — outcomes

| | *Dependent variable (1998–2002):* | | | | |
|---|---|---|---|---|---|
| | Graduated from HS | Had > 1 FT job | Never unemp | HH income | Never arrested |
| | (1) | (2) | (3) | (4) | (5) |
| Recall: *Good* (F87) | 0.047 | 0.008 | −0.015 | 0.011 | −0.064 |
| | (0.066) | (0.060) | (0.069) | (0.077) | (0.080) |
| Recall: *Satisfactory* (F87) | 0.021 | 0.004 | 0.045 | −0.023 | 0.002 |
| | (0.077) | (0.070) | (0.081) | (0.089) | (0.094) |
| Recall: *Unsatisfactory* (F87) | 0.058 | 0.059 | −0.004 | −0.151 | −0.093 |
| | (0.129) | (0.118) | (0.136) | (0.153) | (0.157) |
| Actual: *Good* (F87) | −0.284*** | −0.014 | 0.150* | −0.118 | 0.030 |
| | (0.083) | (0.075) | (0.087) | (0.096) | (0.103) |
| Actual: *Satisfactory* (F87) | −0.377*** | −0.035 | 0.042 | −0.132 | −0.026 |
| | (0.084) | (0.077) | (0.089) | (0.094) | (0.106) |
| Actual: *Unsatisfactory* (F87) | −0.507*** | −0.067 | 0.141 | −0.010 | −0.017 |
| | (0.096) | (0.088) | (0.101) | (0.108) | (0.123) |
| Female | 0.116** | 0.015 | −0.025 | 0.017 | −0.013 |
| | (0.046) | (0.042) | (0.049) | (0.055) | (0.057) |
| Black | 0.106** | −0.069 | 0.011 | 0.094 | 0.018 |
| | (0.051) | (0.047) | (0.054) | (0.058) | (0.060) |
| Constant | 0.736*** | 0.304*** | 0.479*** | 0.515*** | 0.512*** |
| | (0.069) | (0.063) | (0.073) | (0.084) | (0.084) |
| Observations | 421 | 422 | 422 | 283 | 286 |
| $R^2$ | 0.110 | 0.009 | 0.022 | 0.031 | 0.025 |
| Adjusted $R^2$ | 0.091 | −0.013 | 0.001 | −0.001 | −0.006 |
| Residual Std. Error | 0.471 (df = 411) | 0.431 (df = 412) | 0.497 (df = 412) | 0.501 (df = 273) | 0.502 (df = 276) |
| F Statistic | 5.671*** (df = 9; 411) | 0.416 (df = 9; 412) | 1.042 (df = 9; 412) | 0.970 (df = 9; 273) | 0.801 (df = 9; 276) |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. The omitted grade category is *Excellent*. All outcome variables are self-reported. Variable (4) is measured as the subject's household income being over $600 a week ($31,200 a year).

Table 10: Long-run analysis (Fall '87 grades and recall) — beliefs and expectations

| | *Dependent variable (1998-2002):* | | | |
|---|---|---|---|---|
| | Maths SE | Self-efficacy | Life satisf. | Expect prof. success |
| | (1) | (2) | (3) | (4) |
| Recall: *Good* (F87) | −0.046 | −0.016 | 0.074 | 0.130** |
| | (0.064) | (0.027) | (0.068) | (0.064) |
| Recall: *Satisfactory* (F87) | −0.200*** | 0.019 | −0.017 | 0.017 |
| | (0.074) | (0.031) | (0.080) | (0.075) |
| Recall: *Unsatisfactory* (F87) | −0.226* | −0.185*** | −0.035 | 0.133 |
| | (0.125) | (0.053) | (0.134) | (0.127) |
| Actual: *Good* (F87) | −0.126 | −0.046 | −0.190** | −0.194** |
| | (0.080) | (0.034) | (0.086) | (0.081) |
| Actual: *Satisfactory* (F87) | −0.223*** | −0.048 | −0.143 | −0.214*** |
| | (0.081) | (0.035) | (0.087) | (0.082) |
| Actual: *Unsatisfactory* (F87) | −0.256*** | −0.030 | −0.173* | −0.346*** |
| | (0.093) | (0.039) | (0.100) | (0.094) |
| Female | −0.118*** | 0.050*** | 0.006 | −0.055 |
| | (0.045) | (0.019) | (0.048) | (0.045) |
| Black | 0.149*** | 0.028 | 0.027 | 0.053 |
| | (0.049) | (0.021) | (0.053) | (0.050) |
| Constant | 0.852*** | 0.961*** | 0.702*** | 0.796*** |
| | (0.067) | (0.028) | (0.072) | (0.068) |
| Observations | 422 | 421 | 420 | 422 |
| $R^2$ | 0.112 | 0.066 | 0.019 | 0.050 |
| Adjusted $R^2$ | 0.095 | 0.048 | 0.0002 | 0.032 |
| Residual Std. Error | 0.457 (df = 413) | 0.192 (df = 412) | 0.488 (df = 411) | 0.463 (df = 413) |
| F Statistic | 6.499*** (df = 8; 413) | 3.644*** (df = 8; 412) | 1.010 (df = 8; 411) | 2.732*** (df = 8; 413) |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. All outcome variables are self-reported. The omitted grade category is *Excellent*. Variable (1) is measured as the subject stating they are 'good' or 'very good' at 'doing math'. Variable (2) is measured as the likelihood of disagreeing with the prompt 'People like me don't have much of a chance to be successful in life'. Variable (3) is measured as the subject being 'mostly' or 'very' satisfied with 'The way your life has gone since high school'. Variable (4) is measured as the subject responding with 'chances are high' of 'the chance that by age 35 you will...be successful in your line of work'.

# 5 Conclusion

This paper finds that self-assessments of ability may be *motivated* in early life. This is likely to have consequences for ego-relevant beliefs in adulthood: beliefs about ability are at their most malleable prior to adulthood (Dweck 2002); and children may learn the benefits of belief distortion, such as improvements in subjective and objective outcomes, and take them forwards into adulthood. While belief distortions appear to be welfare enhancing in this context, a fruitful direction for further research would be to examine whether they can be self-limiting in others. The answer likely hinges on whether any improvements in felicity, performance and motivation outweigh possible distortions in actions. This appears not to be the case in some contexts (Malmendier and Tate 2005), but the welfare implications are much less obvious in others (Oster et al. 2013).

The relationships we observe between recall and a range of short-run outcomes are consistent with prominent models which suggest that belief distortion can enhance both utility and future outcomes: recalling higher grades is associated with both higher subjective outcomes and higher attainment in the short term, and we observe some evidence of higher classroom enthusiasm. However, particularly because it lacks a rich measure of effort or any exogenous variation in incentives, the analysis also highlights an opportunity for future work to more closely examine the motives underlying belief distortion. A valuable contribution could be made in providing a more precise decomposition of how much belief distortion is driven by a myopic drive to feel better about oneself, and how much by a more strategic impulse to self-motivate or perform more effectively.

The results also make the case that motivated beliefs should be considered more carefully by the growing research agenda on the provision of attainment information to students and parents. Improving the transmission of information may mitigate distortions in important decisions in some contexts (Dizon-Ross 2019). But it may also have heterogeneous effects on beliefs, potentially widening existing gaps between children: students with lower self-confidence may be the ones least likely to suppress further negative feedback. Moreover, while negative feedback may lead to beliefs being better calibrated to reality, it could also be performance-limiting for students who are less willing or able to suppress it.

## 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., & Bedinger, S. (1994). When expectations work: Race and socioeconomic differences in school performance. *Social psychology quarterly*, 283–299.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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. *Forthcoming, American Economic Review*.

Kinsler, J., & Pavan, R. (2021). Local distortions in parental beliefs over child skill. *Journal of Political Economy*, *129*(1), 81–100.

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

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

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

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

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

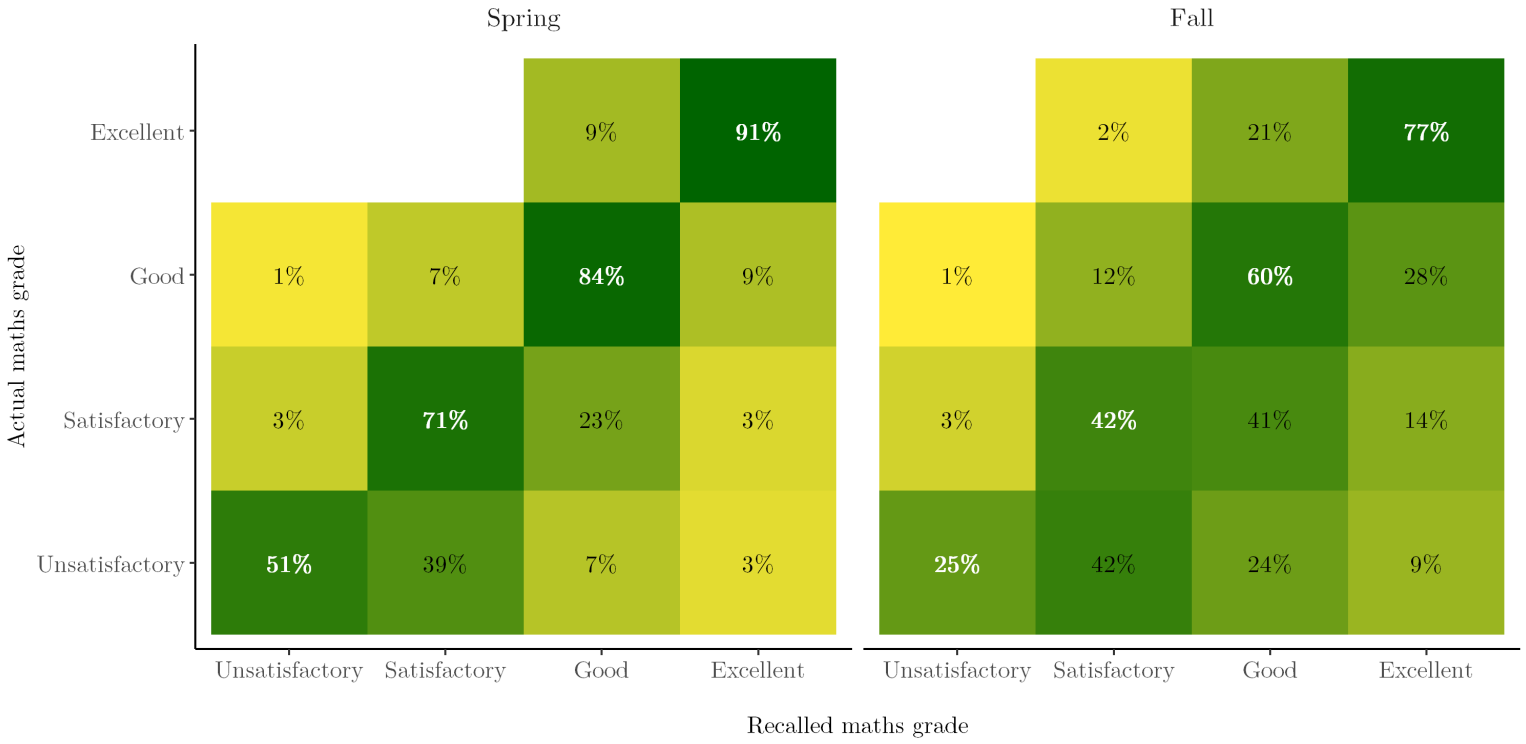
Saltiel, F. (2022). What's math got to do with it? Multidimensional ability and the gender gap in STEM.

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

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

# A Additional Figures and Tables

Figure A3: Actual and remembered grades, split by survey edition




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

Table A1: Recall of maths grades — heterogeneity by sex and race

| | *Dependent variable:* | |
|---|---|---|
| | Incorrect recall | |
| | (1) | (2) |
| Actual: *Good* | 0.071 | 0.113* |
| | (0.081) | (0.062) |
| Actual: *Satisfactory* | 0.232*** | 0.245*** |
| | (0.086) | (0.059) |
| Actual: *Unsatisfactory* | 0.403*** | 0.391*** |
| | (0.095) | (0.063) |
| Female | | 0.025 |
| | | (0.051) |
| Black | | −0.037 |
| | | (0.051) |
| Fall | 0.237*** | 0.238*** |
| | (0.020) | (0.020) |
| Actual: *Good**Female | −0.093 | −0.066 |
| | (0.080) | (0.061) |
| Actual: *Satisfactory**Female | −0.008 | 0.015 |
| | (0.082) | (0.058) |
| Actual: *Unsatisfactory**Female | −0.067 | −0.045 |
| | (0.093) | (0.064) |
| Actual: *Good**Black | 0.153* | 0.085 |
| | (0.083) | (0.062) |
| Actual: *Satisfactory**Black | 0.087 | 0.011 |
| | (0.088) | (0.060) |
| Actual: *Unsatisfactory**Black | 0.227** | 0.144** |
| | (0.098) | (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 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. The omitted grade category is *Excellent*. High SC indicates the student believes they are at least smarter than the average child in their year and school.

Table A2: Fall/Spring comparison: heterogeneity by sex and race

| | *Dependent variable:* | |
|---|---|---|
| | Recall: *Excellent* | |
| | (1) | (2) |
| Actual: *Good* | −0.575*** | −0.622*** |
| | (0.068) | (0.059) |
| Actual: *Satisfactory* | −0.577*** | −0.700*** |
| | (0.070) | (0.053) |
| Actual: *Unsatisfactory* | −0.568*** | −0.683*** |
| | (0.073) | (0.055) |
| Female | | −0.031 |
| | | (0.052) |
| Fall | 0.079*** | 0.078*** |
| | (0.024) | (0.024) |
| Black | | 0.064 |
| | | (0.054) |
| Female*Fall | −0.004 | 0.006 |
| | (0.022) | (0.022) |
| Fall*Black | 0.012 | 0.005 |
| | (0.025) | (0.025) |
| Actual: *Good**Female | 0.004 | 0.008 |
| | (0.068) | (0.060) |
| Actual: *Satisfactory**Female | 0.009 | 0.036 |
| | (0.068) | (0.054) |
| Actual: *Unsatisfactory**Female | −0.016 | −0.001 |
| | (0.071) | (0.055) |
| Actual: *Good**Black | −0.047 | 0.011 |
| | (0.073) | (0.061) |
| Actual: *Satisfactory**Black | −0.160** | −0.077 |
| | (0.074) | (0.056) |
| Actual: *Unsatisfactory**Black | −0.170** | −0.102* |
| | (0.078) | (0.057) |
| Individual FE | Yes | No |
| Academic year FE | Yes | Yes |
| Observations | 2,694 | 2,694 |
| $R^2$ | 0.544 | 0.334 |
| Adjusted $R^2$ | 0.415 | 0.330 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. The omitted grade category is *Excellent*.

Table A3: Spring/Fall recall comparison — robustness check

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Recall: *Excellent* | | | | | |
| | Base | Base | Subject preference | Subject preference | Prior belief | Prior belief |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Actual: *Good* | −0.567*** | −0.543*** | −0.523*** | −0.556*** | −0.569*** | −0.549*** |
| | (0.037) | (0.032) | (0.049) | (0.038) | (0.037) | (0.032) |
| Actual: *Satisfactory* | −0.644*** | −0.639*** | −0.579*** | −0.631*** | −0.639*** | −0.635*** |
| | (0.038) | (0.030) | (0.050) | (0.037) | (0.037) | (0.030) |
| Actual: *Unsatisfactory* | −0.651*** | −0.641*** | −0.582*** | −0.636*** | −0.645*** | −0.631*** |
| | (0.039) | (0.032) | (0.052) | (0.039) | (0.039) | (0.032) |
| Fall | 0.071*** | 0.079*** | 0.040** | 0.057*** | 0.037** | 0.046*** |
| | (0.012) | (0.012) | (0.018) | (0.017) | (0.014) | (0.014) |
| Maths pref $(t-1)$ | | | 0.001 | 0.009 | | |
| | | | (0.022) | (0.016) | | |
| High SC $(t-1)$ | | | | | −0.005 | 0.027* |
| | | | | | (0.020) | (0.014) |
| High SC | 0.049*** | 0.037*** | 0.033 | 0.016 | 0.051*** | 0.040*** |
| | (0.019) | (0.014) | (0.021) | (0.016) | (0.019) | (0.014) |
| Expected: *Good* $(t)$ | −0.100*** | −0.150*** | −0.089*** | −0.161*** | −0.102*** | −0.145*** |
| | (0.024) | (0.021) | (0.027) | (0.025) | (0.024) | (0.021) |
| Expected: *Satisfactory* $(t)$ | −0.110*** | −0.181*** | −0.091*** | −0.183*** | −0.114*** | −0.173*** |
| | (0.025) | (0.022) | (0.029) | (0.025) | (0.025) | (0.022) |
| Expected: *Unsatisfactory* $(t)$ | −0.148*** | −0.209*** | −0.150*** | −0.212*** | −0.153*** | −0.195*** |
| | (0.036) | (0.025) | (0.042) | (0.028) | (0.036) | (0.025) |
| Fall*Maths pref $(t-1)$ | | | 0.071** | 0.049* | | |
| | | | (0.029) | (0.029) | | |
| Fall*High SC $(t-1)$ | | | | | 0.073*** | 0.066*** |
| | | | | | (0.024) | (0.023) |
| Individual FE | Yes | No | Yes | No | Yes | No |
| Academic year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,494 | 2,494 | 1,985 | 1,985 | 2,453 | 2,453 |
| $R^2$ | 0.564 | 0.359 | 0.576 | 0.358 | 0.563 | 0.372 |
| Adjusted $R^2$ | 0.433 | 0.355 | 0.425 | 0.353 | 0.434 | 0.368 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Linear probability model, robust standard errors in brackets. The omitted grade category is *Excellent*. High SC indicates the student believes they are at least smarter than the average child in their year and school.

Table A4: Effect of grade recall on enjoyment of school: robustness to expectations

| | $y_{it}$ |
|---|---|
| | Likes school a lot |
| Recall: *Good* | −0.417 |
| | (0.389) |
| Recall: *Satisfactory* | −0.375 |
| | (0.460) |
| Recall: *Unsatisfactory* | −0.822 |
| | (0.705) |
| Actual: *Good* | 0.078 |
| | (0.607) |
| Actual: *Satisfactory* | −0.206 |
| | (0.660) |
| Actual: *Unsatisfactory* | −0.135 |
| | (0.744) |
| Fall | 0.385 |
| | (0.258) |
| Expected: *Good* $(t+1)$ | 0.344 |
| | (0.358) |
| Expected: *Satisfactory* $(t+1)$ | −0.415 |
| | (0.434) |
| Expected: *Unsatisfactory* $(t+1)$ | −1.664** |
| | (0.816) |
| $y_{i,t-1}$ | 0.203 |
| | (0.286) |
| Observations | 1,708 |

*Note:* *p<0.1; **p<0.05; ***p<0.01. Quadratic exponential model; coefficients are log-odds. Standard errors in brackets.

Table A5: Recall of Maths grades: robustness check for student honesty

| | *Dependent variable:* | | |
|---|---|---|---|
| | Incorrect recall | | |
| | Base | Honesty split | Highest honesty only |
| | (1) | (2) | (3) |
| Actual: *Good* | 0.057 | 0.049 | 0.080** |
| | (0.055) | (0.055) | (0.040) |
| Actual: *Satisfactory* | 0.231*** | 0.220*** | 0.193*** |
| | (0.056) | (0.057) | (0.039) |
| Actual: *Unsatisfactory* | 0.477*** | 0.458*** | 0.409*** |
| | (0.064) | (0.065) | (0.046) |
| Low honesty | | −0.511** | |
| | | (0.202) | |
| Fall | 0.201*** | 0.201*** | 0.207*** |
| | (0.027) | (0.027) | (0.030) |
| Actual: *Good**Low honesty | | 0.635** | |
| | | (0.277) | |
| Actual: *Satisfactory**Low honesty | | 0.614*** | |
| | | (0.228) | |
| Actual: *Unsatisfactory**Low honesty | | 0.673*** | |
| | | (0.229) | |
| Individual FE | Yes | Yes | Yes |
| Academic year FE | Yes | Yes | Yes |
| Observations | 1,879 | 1,879 | 1,400 |
| $R^2$ | 0.410 | 0.412 | 0.111 |
| Adjusted $R^2$ | 0.199 | 0.199 | 0.103 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table A6: Fall/Spring comparison: robustness check for student honesty

| | *Dependent variable:* | | |
|---|---|---|---|
| | $\mathbb{1}(\tilde{m}_{it} = \textit{Excellent})$ | | |
| | Base | Honesty split | Very honest only |
| | (1) | (2) | (3) |
| Actual: *Good* | $-0.586^{***}$ | $-0.646^{***}$ | $-0.629^{***}$ |
| | (0.047) | (0.037) | (0.040) |
| Actual: *Satisfactory* | $-0.656^{***}$ | $-0.762^{***}$ | $-0.739^{***}$ |
| | (0.047) | (0.033) | (0.037) |
| Actual: *Unsatisfactory* | $-0.665^{***}$ | $-0.787^{***}$ | $-0.778^{***}$ |
| | (0.049) | (0.034) | (0.037) |
| Fall | $0.062^{***}$ | $0.074^{***}$ | $0.065^{***}$ |
| | (0.017) | (0.017) | (0.020) |
| Low honesty | | 0.003 | |
| | | (0.025) | |
| Fall*Low honesty | | 0.020 | |
| | | (0.052) | |
| Individual FE | Yes | No | Yes |
| Academic year FE | Yes | Yes | Yes |
| Observations | 1,879 | 1,879 | 1,400 |
| $R^2$ | 0.586 | 0.357 | 0.361 |
| Adjusted $R^2$ | 0.438 | 0.352 | 0.356 |
| *Note:* | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |