Memory Rehearsal and Belief Biases

John J. Conlon

First version: May 28, 2024

This version: August 20, 2024*

Abstract

We rely on memories to form beliefs, but we also frequently revisit memories in conversation and private reflection. I show experimentally that such rehearsal of past experiences generates systematic belief biases. Participants are given a set of experiences and then randomized to have conversations about a subset of them, either ones that reflect well or poorly on them. Such rehearsal has large effects on which of the original experiences participants can recall a week later. Crucially, participants appear naive about rehearsal effects: they take what they remember at face value when later incentivized to form accurate beliefs about the full set of original experiences. Rehearsal therefore distorts not only future recall but also future beliefs. Participants also make rehearsal choices without regard to their later distortionary effects. Intrinsic preferences for thinking about certain experiences instead drive rehearsal choices and therefore belief biases: in particular, a preference to reflect on positive experiences unintentionally generates a positivity bias in future recall and beliefs. This mechanism provides a new non-strategic channel through which seemingly motivated beliefs arise and generates novel predictions in a range of economic domains.

*Carnegie Mellon University: jconlon@andrew.cmu.edu. I thank Chiara Aina, Kai Barron, Lukas Bolte, Lucas Coffman, Ben Enke, Thomas Graeber, Spencer Kwon, Muriel Niederle, Dev Patel, Pete Robertson, Chris Roth, Peter Schwardmann, and Andrei Shleifer for helpful comments.

1 Introduction

Inaccurate beliefs generate choice mistakes across a wide array of economic settings, leading to increased interest in how and why biased expectations arise. Growing evidence suggests memory plays a key role. For example, biases in people's expectations about the what will happen to them in the future often come paired with corresponding biases about what *has* happened to them in the past, such as in beliefs about workplace productivity, sophistication regarding self-control problems, and stock market returns (Huffman et al. 2022, Sial et al. 2023, Gödker et al. 2024). There is thus growing recognition that memories are a crucial input into expectation-formation, and therefore that systematic frictions in what people recall can generate mistakes in beliefs and thereby in choices (Bordalo et al. 2023b, Graeber et al. 2023, Enke et al. 2024, Bohren et al. 2024).

But our reasons for retrieving and attending to our experiences go well beyond using them to construct beliefs or inform some immediate decision. Memory is a ubiquitous part of people's social and inner lives: we constantly relive our pasts, recounting our experiences in conversation with others, telling and re-telling stories, privately reflecting on past episodes, and so on. Often such reminiscing serves no particular purpose besides the inherent enjoyment of thinking back on previous experiences. This fact has received much attention in psychology (Bluck & Alea 2002, Walker et al. 2009), but what implications does it have for economics?

In this paper, I study experimentally how these two forces—memory as a source for inference and reminiscing as an activity with independent utility consequences—interact to generate belief biases. Psychologists have long known that when trying to memorize something, it helps to *rehearse* it: remembering an experience at one time makes it easier to recall in the future (Kahana 2012). Systematic patterns in the experiences people choose to rehearse therefore have the potential to distort the set of experiences at their disposal when later using memory to form beliefs. To formalize and explore these ideas, I first describe a simple model of endogenous memory rehearsal.[1] The agent derives in-the-moment utility depending on which experiences she spends more time thinking about, but doing so also alters which experiences come to mind in the future when she needs to form beliefs from memory. I show that the *magnitude* of belief biases that rehearsal generates depends on whether, and when, agents are sophisticated or naive about its effect on their memory. *Which* biases arise, and their welfare consequences, depends on the motivations underlying rehearsal patterns.

I design an experiment to measure the consequences for belief biases of naturalistic re-

[1] Mullainathan (2002) and Bodoh-Creed (2020) incorporate rehearsal-like effects in models of exogenous memory retrieval, which is used in their models to generate autocorrelation in beliefs and thereby lasting distortions originally arising from other memory forces (i.e., associativity).

hearsal of rich experiences, while maintaining tight control of the underlying environment. A first survey asks participants to list and then describe classes they took in college/high school that they felt they either did well or poorly in (henceforth, "successes" and "failures"). The number of successes vs failures is pre-determined, with the experiment eliciting a fixed number of each one-by-one in a random order. A second survey a week later then asks participants to guess, of these classes from the first survey, what share were successes vs failures. For a standard "rational" agent, this belief elicitation is trivial: participants directly provided the full set of classes and so could simply report the true share. In practice, however, they must rely on what they can remember about the first survey in order to form their beliefs, and the average participant's belief is about 15 percentage points away from the truth. In addition, these beliefs (and errors therein) are tightly linked to participants' memory: they are separately asked to free-recall (by listing the names of) these classes, and this measure of recall is highly correlated with their beliefs.

To study the effect of rehearsal, participants in the first survey have conversations with a large language model about three of the classes they named. These conversations are randomized to focus either on three successful classes (the *Successes* treatment) or three unsuccessful classes (the *Failures* treatment), and participants are aware that this is how the conversation topics are selected. These conversations are incentivized: participants' responses are graded according to how engaged they were in the conversation, and their bonus payment increases (in expectation) in this score. This method of incentivization, as well as the structure of the conversations, was chosen to generate organic but focused rehearsal of the classes being discussed. To my knowledge, this natural-language method of keeping certain experiences top-of-mind is novel in economics experiments, though related methods are sometimes used to measure what is already top-of-mind (Haaland et al. 2024).

These conversations have large effects on recall a week later. Participants in the second survey recall 45.9% of the classes they do not have conversations about. In contrast, they recall 72.1% of the classes they were randomized to have conversations about, a 26.2 percentage point increase ($p < 0.01$). There are therefore significant corresponding differences across treatments in the number of successes and failures that participants recall. In the *Successes* treatment, participants on average recall 2.9 successes and 1.9 failures, while those in the *Failures* treatment recall 2.2 successes and 2.8 failures (all differences significant at the $p < 0.01$ level. Thus, of the classes participants do recall, 62.7% are successes in the *Successes* treatment but only 41.8% are successes in the *Failures* treatment (a 20.9 percentage point difference, $p < 0.01$).[2]

[2]These treatment effects are almost entirely on the extensive margin: i.e., whether classes originally classified as successes vs failures are recalled at all in the second survey. I find only very small effects on

I then turn to my primary question: the implications of these rehearsal effects for belief biases. The model I describe makes the intuitive point that rehearsal can only bias beliefs if agents are naive about its effect on their memory when they are retrieving experiences *ex post* to inform their beliefs. Such an agent takes what she recalls at face value, failing to account for the fact her recall is distorted by earlier rehearsal. "Biases" in rehearsal (e.g., only having conversations about successes) therefore produce biases in her beliefs. In contrast, an *ex post* sophisticated agent, who realizes that some experiences are more likely to come to mind due to rehearsal, can "back out" accurate beliefs even from a biased memory. *A priori*, it is unclear whether we should expect sophistication about rehearsal. On the one hand, agents appear naive *ex post* about other forces in memory such as cueing and associativity when forming beliefs (Bordalo et al. 2023b). On the other hand, rehearsal is perhaps the most well-known force that shapes human recall, as any student using flashcards to prepare for an exam can attest. Indeed, rehearsal is sometimes defined in explicitly sophisticated terms. For example, Kahana (2012) writes, "By rehearsal, we mean the *strategy* of repeating previously presented items ... to keep those items in mind until the time of test" (pg 71, italics mine). The question is whether people understand the magnitude (and duration) of these effects or realize that they shape memory even in contexts where agents may not be explicitly attempting to memorize anything.

I find that participants appear almost wholly naive *ex post* about rehearsal's effect on their memory. Those in the *Successes* treatment, who had conversations about only successful classes, believe that 61.8% of the original classes were successes. This average belief is 16.5 percentage points lower ($p < 0.01$) among participants in the *Failures* treatment whose conversations instead focused on unsuccessful classes. This treatment effect is not statistically distinguishable from the 20.9 percentage point effect on recall, so I cannot reject that participants are fully naive *ex post*. In addition to the incentivized belief about the share of classes that are successes vs failures, the second survey also asks participants to make a binary bet on whether a randomly selected class is a success vs a failure. Participants in the *Successes* treatment are 47.6 percentage points more likely ($p < 0.01$) to bet that the randomly selected class is a success than those in the *Failures* treatment. In a sense, this *ex post* naivete represents a internal form of selection neglect (Enke 2020, Jin et al. 2021, Barron et al. 2024, Farina et al. 2024): participants take what they remember at face value, failing to account for the fact that factors beyond objective frequency contribute to recall.

My experiment exogenously varies rehearsal patterns to identify their effect on partici-

the intensive margin (whether, conditional on recalling the name of a class, it is (mis)remembered as being a success vs a failure). This is partly because participants are extremely internally consistent in classifying individual classes as successes vs failures (they agree across surveys in 93% of cases). Thus the partly subjective nature of the success vs failure classification does not drive any results.

pants' memory and beliefs. But which experiences do participants choose to rehearse, and therefore which belief biases would arise endogenously? To answer these questions, I ask participants to express a preference for which classes to have conversations about. They can choose to discuss only successes, only failures, or a randomly selected set of classes. To make this elicitation incentive-compatible, I actually implement the choice of a small fraction of participants. Using the estimated treatment effect of rehearsing positive experiences, I then compute what I call participants' "endogenous belief": the belief they would have formed had their rehearsal choice been implemented. I find that participants are dramatically more likely to choose to discuss successes than failures (44.6% vs 10.0%, $p < 0.01$). This "bias" toward rehearsing positive experience, combined with *ex post* naivete, implies that participants' endogenous beliefs are overly "optimistic": that is, they prefer rehearsal patterns that on average lead them to believe that they did well in more classes than they in fact did.[3]

What drives these rehearsal choices and therefore endogenous belief biases? Open-ended questions asking participants the reasons behind their rehearsal decisions allow me to provide qualitative evidence on this question. Two factors appear to largely determine participants' choices. First, financial incentives appear to matter: 37.0% of participants mention preferring to discuss classes about which they could more easily have thoughtful conversations (recall that participants' bonus payment depended on how thoughtful their responses in the conversations were graded as being). Second, many participants report an intrinsic preference for thinking about times they were successful and for avoiding ruminating on failures: over half of those who choose to discuss successes mention this consideration.

In contrast, almost no participants mention that their choice of which experiences to rehearse might affect their memory or beliefs in the future. A randomly selected half of participants were warned immediately before making their rehearsal choice that they would later face monetary incentives for accurate beliefs about the fraction of classes that were successes vs failures. Of these, only 3.3% appear to mention anything about how their choice of conversation might affect these future beliefs or their future memory. Further, this manipulation does not have any effect on participants' rehearsal choices, nor does it attenuate the effect of rehearsal on later beliefs. These results suggest that, in addition to being naive *ex post* about the effect of rehearsal on their memory, participants are also naive *ex ante*: they do not take into consideration effects on their later memory/beliefs when deciding which experiences to rehearse. In-the-moment factors, either instrumental motives (financial

[3]Of course, because these beliefs are about the number of successful vs unsuccessful classes that the first survey asked about (not about, say, how well participants did overall academically in school), these apparently "overoptimistic" beliefs are likely divorced from ego-relevant considerations about ability and self-confidence. This was an intentional design choice: seemingly "overconfident" beliefs can be generated through rehearsal choices even in a domain that is plausibly ego-irrelevant.

incentives) or intrinsic preferences (for thinking about positive experiences), instead drive rehearsal choices and therefore belief biases within my experiment.

This mechanism—intrinsic preferences for thinking about positive experiences unintentionally biasing beliefs—is to my knowledge a novel explanation in economics for overoptimistic beliefs. The typical story told in the literature on motivated reasoning is inherently sophisticated: agents with belief-based or anticipatory utility optimize their future beliefs, achieving an inflated opinion of themselves by strategically managing the evidence they collect or remember (e.g., Bénabou & Tirole 2002, Brunnermeier & Parker 2005).[4] My evidence shows that belief-optimization is not necessary to explain seemingly motivated memory: preferences over which experiences to rehearse, plus *ex post* naivete about their effects on memory, are sufficient. Further, I show that while this latter mechanism can make similar predictions as a sophisticated optimized-beliefs story about what biases agents will come to hold (and even about the means by which they arrive at these misperceptions), they have starkly different welfare implications: for example, a sophisticated agent with belief-based utility is *harmed* by an information intervention correcting her beliefs, while a naive agent with intrinsic rehearsal preferences is helped. Disentangling these two mechanisms is therefore relevant for policy makers considering whether de-biasing beliefs will tend to be beneficial.

Finally, I explore through the lens of a simple model how rehearsal-based overoptimism about the past leads to belief biases about the future. I first show that an agent who enjoys thinking about positive experiences grows increasingly overoptimistic about past time periods. This is because she has had more opportunities to differentially rehearse positive experiences that happened longer ago. Thus, while the agent is on average weakly overoptimistic about all periods in the past, she is especially so about the more distant past. Consistent with this result, Americans overwhelmingly tend to report in surveys that the best time in the US along many dimensions, including arts and music, happy family lives, moral society, and political harmony, happens to coincide with their own childhood. Conversely, people across generations tend to agree that the worst time in the country's history along all these dimensions happens to be now (Dam, 2024).

[4]The evidence provided for this optimization channel is typically that overoptimism is sometimes found to decrease as the monetary incentives for accurate beliefs increase. However, such evidence shows that agents can respond to incentives for accuracy, but not that their biases absent monetary incentives were chosen strategically. For example, suppose that agents know they have imperfect/noisy memory but also, as in my experiment, fail to understand how acting on their intrinsic rehearsal preferences biases their recall. With sufficient incentives for accurate beliefs, agents may take other steps to improve their memory (e.g., writing down everything that happens during the experiment to ensure no forgetting later on, or spending much more time and effort trying to remember everything they can at the retrieval stage). Such agents might "improve" their beliefs in response to greater incentives, but this is consistent with never having *attempted* to have biased beliefs in the first place.

How does the agent interpret her increasingly rosy view of the past when making forecasts about the future? I assume she has priors about both the expected value of the initial period and how this expected value changes over time. She uses Bayes rule to form posteriors about these two parameters from what she (seems to) remember, which she then uses to forecast the expected value of future periods. I show that whether her rehearsal-induced overoptimism about more distant past periods leads her to become overoptimistic about the future depends crucially on what she knows about whether the expected value of periods is changing over time. Suppose first that the agent knows there is not any true time trend: for example, she might know that her true expected productivity compared to her coworkers is relatively constant over time. She will then interpret the last few periods (which she tends to remember less overoptimistically than more distant periods) as a string of somewhat bad luck and expect mean-reversion to an overly high expected value. Such an agent is therefore overconfident about the future because she is overoptimistic about the past. But suppose instead that the agent suspects there may be a time trend but is unsure: perhaps, for example, she wonders whether the country is changing for the better or not. This agent will interpret her decreasing overoptimism about the more recent past as evidence that things are getting worse, and she will therefore become overly pessimistic about the future. Such an agent looks nostalgic for the past and skeptical that the negative apparent trend will reverse itself.

I conclude by briefly applying this basic insight to three economic environments. First, I show that this force predicts that voters will repeatedly sour on incumbent politicians, whom they blame for the deterioration of affairs that their rehearsal leads them to perceive. Second, rehearsal leads agents to be excessively hesitant to adopt, and overly quick to abandon, self-improvement technologies like therapy. Their overoptimism about the past leads them to understate their initial need for improvement, while they blame the fact that more recent periods seem less rosy than more distant ones on the self-improvement investment failing to work. Finally, rehearsal can produce a status quo bias through a form of information trap. Suppose an agent repeatedly consumes a good (e.g., a similar vacation every year) whose perceived average utility serves as a signal of its expected quality. Rehearsal will lead her to believe that her previous purchases were better than they in fact were. She can thus become stuck with her initial decision, failing to switch to a better option even with an infinite stream of mediocre signals about her current choice. These applications point toward how a greater understanding of rehearsal-based belief biases can generate rich new predictions about economic contexts.

This paper contributes to a growing literature in economics studying memory distortions. There is by now ample evidence that people's memories of their own pasts are often system-

atically distorted. Models explaining memory distortions have largely emphasized the role of similarity, both between experiences themselves as well as between experiences and the retrieval context or cue, in explaining a wide variety of belief biases and choice anomalies (Mullainathan 2002, Bordalo et al. 2023b, Graeber et al. 2023, Bordalo et al. 2024, Enke et al. 2024, Bohren et al. 2024). Similarity alone, however, does not provide an obvious explanation for the well-documented pattern in both economics and psychology that people tend to hold overoptimistic or self-serving beliefs about their own past: e.g., overestimating one's own previous productivity, academic performance, consistency, healthy behaviors, and financial returns (Saucet & Villeval 2019, Adler & Pansky 2020, Chew et al. 2020, Zimmermann 2020, Müller 2022, Huffman et al. 2022, Amelio & Zimmermann 2023, Sial et al. 2023, Gödker et al. 2024). Instead the literature on motivated memory focuses on sophisticated explanations whereby agents strategically manage their future beliefs by intentionally manipulating their own memory (Bénabou & Tirole 2002).[5] To my knowledge, the explanation my experiment provides evidence for—that intrinsic preferences for thinking about positive experiences unintentionally generate overoptimistic beliefs when acted on by naive agents—is novel to this literature.[6]

Next, this paper contributes to a literature on attention in economics (see Loewenstein & Wojtowicz 2023 for a review). Existing work has largely focused on the forces that attract attention as well how mistakes can arise in decisions when too much/little attention is devoted to features of those decisions (e.g., Bordalo et al. 2022, 2023a, Koszegi & Szeidl 2013, Bushong et al. 2021, Link et al. 2023, Conlon 2024). There is a also a small literature on attention-based utility that relates to this paper. For example, Bolte & Raymond (2024) model how preferences for attending to past or future good states affect choices, though their agent is fully rational (i.e., forward looking with correct expectations), precluding beliefs biases like those I study. Quispe-Torreblanca et al. (2022) show that people have a preference for attending to positive information (about stock market portfolios) even when it

[5]A related but distinct literature studies "asymmetric updating," a tendency to react more to positive news than negative news. Often such updates are measured immediately after participants receive signals, and so memory *per se* is unlikely to underlie this form of apparently motivated cognition, though more general attention-based utility rather than beliefs-based utility could. See, among others, Eil & Rao (2011), Möbius et al. (2022), Drobner (2022), Engelmann et al. 2024, and Bolte & Fan (2024).

[6]Perhaps relatedly, many papers find that people tasked with convincing others of some fact (e.g., a legal or political case) tend to also convince themselves in a way that seems "self serving" (e.g., Babcock et al. 1995, Schwardmann & van der Weele 2019, Schwardmann et al. 2022). Such effects are sometimes explained by suggesting that agents try to convince themselves of an argument in order to better convince others. Like with belief-based utility, this explanation is inherently sophisticated. My results suggest a different, more naive mechanism may also operate: participants tasked with arguing for one side of an argument may simply pay more attention to (i.e., rehearse more) the arguments that favor their assigned side. If agents are naive about the fact that this will make such arguments more memorable to themselves later on, they will accidentally convince themselves of the merits of their case.

has no effect on their beliefs, a finding they connect with information-avoidance behavior à la the "ostrich effect" (Karlsson et al. 2009). To my knowledge, my paper is the first provide direct experimental evidence that attention allocation (which rehearsal requires) at one time can have longer-run cognitive effects, through memory, on beliefs at later times.

Finally, this paper speaks to a literature on rehearsal effects in psychology that goes back at least to the nineteenth century. Most such work focuses on exploring how different variables (e.g., lag time, spacing, elaboration) increase or decrease the effect of rehearsal on later recall (Madigan 1969, Craik & Lockhart 1972, Roediger & Karpicke 2006, Cepeda et al. 2006, Bartsch et al. 2018) and how other patterns in memory (e.g., primacy effects) are related to rehearsal (Rundus 1971, Brodie & Murdock 1977, Tan & Ward 2000). The psychology literature tends to focus on participants actively studying word lists, unlike my experiment where experiences and rehearsal episodes are more naturalistic. In addition, to my knowledge none of the work in psychology speaks to naivete about rehearsal nor connects memory distortions to belief biases.

2 Model

Setup I first describe a simple model of endogenous memory rehearsal to fix ideas and motivate the experimental design. Assume there are three periods $t \in \{0, 1, 2\}$. In periods $t = 1$ and $t = 2$, the agent has a memory database M comprised of individual experiences m that occurred in the prior period $t = 0$. Each experience $m \in \{1, 2, ..., M\}$ has some characteristic y_m. For example, M might represent various assignments, tests, or classes a student took in school, and y_m might capture how successful or unsuccessful she was in each. In period $t = 1$ the agent decides how much to think about each experience m. Let $a = (a_1, a_2, ..., a_M)$ denote the vector of her rehearsal choices, which I will also interchangeably call her attention vector. For simplicity, I assume that during this period, the agent has perfect memory: she can decide to think more about any of her experiences, without any uncertainty about or constraint on whether she can recall them.

Utility In period $t = 1$, the agent derives some in-the-moment utility u_a depending on the rehearsal vector a that she chooses: in particular, I assume $u_a = \sum_m \nu_m a_m - \frac{\phi}{2} a_m^2$. That is, the agent derives a constant marginal utility ν_m from thinking about m (e.g., perhaps some experiences are simply pleasant or unpleasant to think about, or perhaps she faces decisions in $t = 1$ that require thinking back on some experiences more than others) but with a quadratic cost ϕ. A natural way to interpret this assumption is that $a_m = 0$ is a baseline level of rehearsal, with the agent paying a cost to distort her attention away from

this baseline.

In the final period $t = 2$, I assume the agent derives utility from two potential sources. First, she may have belief-based utility $u_b = \alpha E_a[\bar{y}]$ depending on her subjective expectation of the average value of the characteristic y_m in her memory database (where $\alpha = 0$ nests the standard case without belief-based utility). For example, the agent may have an intrinsic preference for her future self to have high beliefs about her ability. Second, the agent derives instrumental utility u_c (c for *c*orrect) that decreases in how far her belief $E_a[\bar{y}]$ is from the true average value of y_m. I assume a simple quadratic loss function, $u_c = -\frac{\beta}{2}(E_a[\bar{y}] - \bar{y})^2$, where β governs the costs of incorrect beliefs.[7]

Equation 1 summarizes these assumptions about the agent's utility:

$$u = \underbrace{\sum_m \nu_m a_m - \frac{\phi}{2} a_m^2}_{\text{Rehearsal Utility } u_a} + \underbrace{\alpha E_a[\bar{y}]}_{\text{Belief-Based Utility } u_b} - \underbrace{\frac{\beta}{2}(E_a[\bar{y}] - \bar{y})^2}_{\text{Instrumental Utility } u_c} \tag{1}$$

Beliefs by Sampling How does the agent form her beliefs $E_a[\bar{y}]$ about the average value of y_m in period 2? I assume she does so by *sampling* experiences from her memory database M. She then takes a weighted average of the y_m values of the experiences she samples. I assume she takes N iid samples (with replacement), and I analyze her beliefs in the limit where $N \to \infty$. Let $r(m, a)$ be the probability that the agent recalls experience m given rehearsal vector a. Finally, let $\omega(m, a)$ be the weight the agent gives m in the weighted average conditional on recalling that experience. With these assumptions, the agent's beliefs are given by equation 2:

$$E_a[\bar{y}] = \frac{1}{N}\sum_{n=1}^{N} y_{m(n)} \cdot \omega(m(n), a) \quad \xrightarrow{P} \quad \sum_{m \in M} y_m \cdot r(m, a) \cdot \omega(m, a) \tag{2}$$

Rehearsal and *Ex Post* Sophistication I assume that the probability $r(m, a)$ of recalling experience m in period 2 is higher the more attention a_m that experience received in period 1 and the less other experiences received. For tractability, I assume $r_m \propto \exp\{\gamma a_m\}$, where γ governs the strength of these rehearsal effects. A crucial question is whether the agent realizes the extent to which earlier rehearsal alters her probability of recalling experiences. Let $\tilde{r}_2(m, a)$ be the agent's belief in period 2 about her probability of recalling m, which I assume takes the following form: $\tilde{r}_2(m, a) \propto \exp\{\lambda_2 \gamma a_m\}$. The sophistication parameter λ_2 represents the agent's (dogmatic) belief about the strength of rehearsal effects,

[7] A simple way to microfound this would be to assume the agent takes some action x with a quadratic loss function. i.e., if $u_c(x) = -\frac{\beta}{2}(x - \bar{y})^2$, then the agent will choose $x^* = E_a[\bar{y}]$. In that case, we can rewrite this utility as $u_c(E_a[\bar{y}]) = -\frac{\beta}{2}(E_a[\bar{y}] - \bar{y})^2$

relative to their true strength. The *ex post* rational, full-sophistication benchmark is $\lambda_2 = 1$, while full naivete would correspond with $\lambda_2 = 0$.

I assume that the agent attempts to correct for the effect she believes rehearsal has on her memory when she is forming beliefs. In particular, she employs an inverse probability weighting rule to back out the true average $\bar{y}$ of her experiences even in the presence of rehearsal effects. Let $\omega(m, a) = \frac{1}{M\tilde{r}_2(m,a)}$, meaning she down-weights experiences she believes she was more likely to recall and up-weights experiences that she believes were less likely to come to mind. Then Proposition 1 follows (all proofs in Appendix B):

Proposition 1 *Rehearsal distorts beliefs if and only if it boosts recall of more attended-to experiences ($\gamma > 0$) and agents are at least partially naive ex post ($\lambda_2 < 1$):*

$$E_a[\bar{y}] \approx \bar{y} + (1 - \lambda_2) \cdot \gamma \cdot Cov(a_m, y_m) \tag{3}$$

The approximation is a first-order Taylor expansion around $a = 0$.[8]

Note that I have interpreted naivete as a belief by the agent that rehearsal affects her recall less than it truly does. An equivalent interpretation (in the sense that Proposition 1 would still follow) would be that, though the agent in a sense is aware of rehearsal, she simply fails (to an extent $1 - \lambda_2$) to correctly apply this knowledge when forming her beliefs. For example, perhaps it simply does not come to mind that one's recall is distorted, or perhaps the agent does not know which weighting strategy would correct for this fact. I leave these interpretations aside as they all lead to similar conclusions about how rehearsal would distort beliefs.[9]

Rehearsal Choice and *Ex Ante* Sophistication We have seen that rehearsal distorts an *ex post* naive agent's beliefs in period 2. How does such an agent choose how much to rehearse each experience in period 1? The answer depends in part on whether the agent is sophisticated *ex ante* about rehearsal: does she realize that her choice of which experiences to think more about today will distort her recall tomorrow? Assume now that the agent

[8] This approximation assumes away two forces. First, of course, it eliminates non-linearity in the effect of attending to an experience on whether that experience is recalled later. This will be especially awkward for very negative values of a_m as of course the probability of recall cannot be negative. More substantively, the approximation assumes that increasing rehearsal of one experience boosts recall of it by distracting *equally* from all other experiences. For example, boosting rehearsal of one experience should disproportionately reduce recall rates of experiences that one is already rehearsing more, a dynamic that is shut down by linearizing.

[9] An important place where these interpretations would come apart is when thinking about what interventions might help an agent to correct for her distorted recall. If the only problem is that agents don't think to "undo" rehearsal effects when forming beliefs, then a simple nudge toward doing so could be helpful.

in period 1 believes that her period-2 recall probability is given by $\tilde{r}_1(m,a) \propto \exp \lambda_1 \gamma a_m$. That is, the agent believes, when choosing which experiences to attend to, that rehearsal effects are only a λ_1 fraction of their true magnitude. For $\lambda_1 = 1$ the agent is fully *ex ante* sophisticated, whereas if $\lambda_1 = 0$ she is fully naive. I also make the simplifying assumption the agent knows about her future level of *ex post* sophistication λ_2, such that she treats her future self as believing that recall probabilities are proportional to $\exp \lambda_1 \lambda_2 \gamma a_m$. Proposition 2 then characterizes the agent's rehearsal choice and ultimate beliefs, which I then interpret through a series of corollaries

Proposition 2 *The agent chooses rehearsal vector a^* in period 1 according to equation 4 and forms beliefs in period 2 according to equation 5:*

$$a^*_m = \frac{1}{\phi}\left[\nu_m + \lambda_1(1-\lambda_2)\frac{\gamma}{M}(y_m - \bar{y})\Big(\alpha - \lambda_1\beta\big(E_{a^*}[\bar{y}] - \bar{y}\big)\Big)\right] \tag{4}$$

$$\textit{and} \qquad E_{a^*}[\bar{y}] = \bar{y} + (1-\lambda_2)\frac{\gamma}{M}\frac{Cov(y,\nu) + \alpha\lambda_1(1-\lambda_2)\frac{\gamma}{M}Var(y)}{\frac{\phi}{M} + \beta\Big(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\Big)^2 Var(y)} \tag{5}$$

Corollary 1 *Intrinsic rehearsal preferences distort later beliefs if and only if rehearsal distrorts recall ($\gamma > 0$) and agents are at least partially naive ex post ($\lambda_2 < 1$). With incentives for accurate beliefs ($\beta > 0$), ex ante sophistication ($\lambda_1 > 0$) moderates but does not eliminate the effect of rehearsal preferences on belief biases:*

$$\frac{dE_{a^*}[\bar{y}]}{d\nu_m} = \frac{(1-\lambda_2)\gamma(y_m - \bar{y})}{\phi M + \beta\big[\lambda_1(1-\lambda_2)\gamma\big]^2 Var(y)} \tag{6}$$

The intuition behind Corollary 1 is straightforward. If the agent enjoys thinking about experience m, she will do so more and therefore m will come to mind more readily in period 2. Assuming she will be *ex post* naive about rehearsal, her later beliefs therefore increase or decrease depending on whether y_m is especially high or low. If the agent is also naive *ex ante* about rehearsal ($\lambda_1 = 0$), she will not realize that boosting rehearsal of m comes at the cost of biasing her beliefs in the future. Importantly, however, even a fully *ex ante* sophisticated agent ($\lambda_1 = 1$) reacts to her intrinsic rehearsal preferences. Her concern for her future beliefs attenuates but does not eliminate her rehearsal response and therefore the effect her preference ends up having on her future beliefs. Thus, the necessary and sufficient condition for rehearsal preferences to bias beliefs is that the agent be *ex post*, but not necessarily *ex ante*, naive.

Proposition 1 makes clear that we can test for *ex post* naivete about rehearsal by looking at the effect on beliefs of inducing different covariances between rehearsal and y_m. How can

we test for *ex ante* sophistication? As is intuitive, Corollary 2 below states that we can do so by varying the incentives for accurate beliefs that the agent in period 1 is aware she will face in period 2. An *ex ante* sophisticated agent will then alter their rehearsal choices to reduce belief biases in the future.

Corollary 2 *Incentives for accurate beliefs affect rehearsal choices (and therefore beliefs) if and only if rehearsal distorts recall ($\gamma > 0$), agents are at least partially naive ex post ($\lambda_2 < 1$), and agents are at least partially sophisticated ex ante ($\lambda_1 > 0$):*

$$\frac{dE_{a^*}[\bar{y}]}{d\beta} = -(E_{a^*}[\bar{y}] - \bar{y}) \frac{\left(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\right)^2 Var(y)}{\frac{\phi}{M} + \beta\left(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\right)^2 Var(y)} \tag{7}$$

Welfare Implications We have seen above how to test for rehearsal effects, *ex post* sophistication, and *ex ante* sophistication. I now turn briefly to the question of the welfare implications of the different answers we might find. I consider three benchmarks, the first two being the most common assumptions in the literature, and the third being a new (but, as we shall see, empirically relevant) assumption. The first benchmark is a fully rational agent (subject to memory distortions) who is both *ex post* and *ex ante* sophisticated.

My second benchmark is what I call a pure belief-based utility agent. This agent has a strict preference for self-serving beliefs ($\alpha > 0$) but no intrinsic preferences for thinking about experiences ($\nu_m = 0$ for all m). Critically, this agent is naive *ex post* about rehearsal ($\lambda_2 = 0$) but sophisticated *ex ante* ($\lambda_1 = 1$). That is, this agent knows she can use rehearsal in period 1 to manipulate her period-2 self's beliefs in the future. Further, her belief-based utility gives reason to do so. This benchmark represents the standard assumptions in the motivated reasoning literature (e.g., Bénabou 2015).[10]

Finally, I consider what I call a pure rehearsal-based utility agent. This agent has no belief-based utility ($\alpha = 0$). Instead, she simply intrinsically prefers thinking about positive experiences, which I capture by assuming that $\nu_m = \nu(y_m - \bar{y})$, where ν captures the intensity of this preference. Further, this agent is naive both *ex post* ($\lambda_2 = 0$) and *ex ante* ($\lambda_1 = 0$).

Corollary 3 gives the rehearsal choices and ultimate beliefs for each of these benchmarks. It also looks at how period-2 utility would change if agents were credibly informed about the true value of $\bar{y}$. I compute this by looking at how u_b and u_c would change were the agent's beliefs changed to the truth. This holds fixed any rehearsal-based utility u_a.

[10] Actually, oftentimes in this literature agents are supposed to be *ex post* sophisticated as well. This greatly complicates the agent's task of manipulating her future self, as a rational sophisticated agent must satisfy Bayes rule and therefore many belief biases are ruled out (e.g., about expected values).

Corollary 3 *Let $\gamma > 0$. Define Δu as the change in utility in period 2 from changing the agent's beliefs to be correct.*

1. ***Rationality**: If $\lambda_1 = \lambda_2 = 1$, then*

$$a_m^* = \frac{\nu_m}{\phi} \quad \textit{and} \quad E_{a^*}[\bar{y}] = \bar{y} \quad \textit{and} \quad \Delta u = 0$$

2. ***Belief-based utility**: If $\lambda_1 = 1$, $\lambda_2 = 0$, $\nu_m = 0$, and $\alpha > 0$, then*

$$a_m^* = \frac{\alpha\gamma(y_m - \bar{y})}{\phi M + \beta\gamma^2 Var(y_m)} \quad \textit{and} \quad E_{a^*}[\bar{y}] = \bar{y} + \frac{\alpha\gamma^2 Var(y_m)}{\phi M + \beta\gamma^2 Var(y_m)} \quad \textit{and} \quad \Delta u < 0$$

3. ***Rehearsal-based utility**: If $\lambda_1 = \lambda_2 = \alpha = 0$, and $\nu_m = \nu y_m$, then*

$$a_m^* = \frac{\nu}{\phi} y_m \quad \textit{and} \quad E_{a^*}[\bar{y}] = \bar{y} + \frac{\gamma\nu}{\phi} Var(y) \quad \textit{and} \quad \Delta u > 0$$

Corollary 3 makes three points. First, as we have seen, rehearsal only matters for beliefs if agents are naive *ex post* about them. The rational agent can "myopically" optimize her rehearsal preferences in period 1, leaving it to her future period-2 self to back out accurate beliefs despite "biased" recall.

Second, in a static setting both the belief-based utility and rehearsal-based utility benchmarks can yield similar predictions. For example, an *ex ante* sophisticated agent who wants to have positive beliefs about $\bar{y}$ can achieve those beliefs by spending more time rehearsing better-than-average experiences. A naive agent who simply prefers thinking about those experiences will make a similar choice and therefore (by accident) end up with similar beliefs.

Third, though the belief-based utility and rehearsal-based utility benchmarks can make similar predictions about rehearsal patterns and ultimate beliefs, they have starkly different welfare implications. Consider an information intervention designed to help agents make better informed decisions by telling them in period 2 the true value of $\bar{y}$. The belief-based utility agent is *harmed* by this policy. The intuition is a simple revealed preference argument: this agent *chose* her misspecified beliefs rather than remain correct and so must prefer her bias to accurate beliefs. The rehearsal-based utility agent, in contrast, never intended to bias her beliefs. She is helped by this policy because she no longer loses utility due to erroneous beliefs (and still reaps the rehearsal-based utility from period 1 that came from focusing on positive experiences). Thus, it is important from a policy perspective to disentangle why agents make the rehearsal decisions they do, and therefore which belief biases they end up having.

3 Experimental Design

The experiment was designed to mimic the setting of the model described above. In particular, it first elicits from participants a memory database, then provides an opportunity to differentially rehearse some of the experiences included in it, and then later measures beliefs about and recall of these experiences. I describe each component of this design in turn.

Memory database The experiment first elicits from participants a "memory database" for the purpose of the study. Oftentimes memory experiments define such a database by providing participants a set of new (typically short and abstract) experiences, such as a word list or a set of Raven's matrices problems, and then test their memory of these experiences later on. My experiment instead has participants describe nine of their existing experiences in a first survey, and then tests in a second survey a week later what participants can recall about these nine experiences. In particular, it asks both whether participants can remember which experiences they described in the first survey (in a free-recall task) as well as their beliefs about summary statistics regarding this set of experiences (e.g., what share of them had particular characteristics). This design allows me to exploit richer experiences—about which activities like conversation and reflection are natural and which participants may have pre-existing preferences for thinking about—than typical memory experiments that provide very brief and abstract experiences can.

More precisely, in the first of two surveys, participants are asked to name and describe nine classes they took during college or high school (depending on whether they have a college degree or not). They are asked one-by-one to name classes that they felt they did well or poorly in "compared to your normal academic performance," which I henceforth call "successes" and "failures." In a second survey a week later, participants are asked to recall which nine classes they named in the first survey as well as the share of these classes that were successes vs failures. This binary success vs failure variable represents the y_m characteristic from the model in Section 2. The ground truth for this variable is predetermined: the survey randomly asks either for four successes and five failures or for five successes and four failures. These elicitations occur in a random order. In addition to the name of the class, the survey also asks a few followup questions about each class, including the gender of the professor, the year in school participants took it, whether it involved a final project or paper as a major part of the grade, whether it was part of their major (for college classes), and whether it was an elective class (for high school classes).

Rehearsal of Experiences After listing the nine classes in the first survey, participants were informed that they would be having conversations about three of the classes they had just named. Their conversation partner was described as a "chatbot" and was in fact GPT-4, a large language model. These conversations consisted of GPT-4 asking natural-language questions about the class in question, with participants answering in their own words and then followup questions being asked in response to their answers. Each conversation stopped after either 7.5 minutes or 12 question-and-answer pairs, whichever was shortest. The average participant wrote 209 words per conversation, while on average GPT-4 wrote 189 words per conversation.

These conversations were incentivized so as to encourage participants to be engaged and pay attention to the classes they were about. In particular, participants were told (truthfully) that their responses in each of the three conversations would be graded "on a scale from 0 to 100 according to how thoughtful they were". If a conversation was the randomly chosen response that determined their bonus, they would earn a bonus with a percent chance equal to the grade they received for it. In practice, this grade was provided by GPT-4, though participants were not told this.

After learning they would be having these conversations, participants were asked whether they would prefer to have conversations about three classes they did well in ("successes", though they were not framed this way to participants), three classes they did poorly in ("failures"), or a randomly selected three classes. To make these elicitations incentive-compatible, 5% of participants had their choice actually implemented. For the remaining 95% of participants, the computer simply flipped a coin to decide whether they would discuss three successes or three failures. Participants knew this procedure and were also told whether their preference was implemented and, if not, whether they were randomized to have conversations about three classes they did well or poorly in. If participants were randomized to discuss three successes (the *Successes* treatment), the classes they discussed were randomly selected from the four or five successes in the memory database. Similarly, the three discussed classes in the *Failures* treatment were randomly chosen from the set of failures in the memory database.

Participants were randomly either warned that they would later face incentives for recall and accurate beliefs or were not warned. For the half of participants who were not warned, nothing in the experiment suggested that the study was about memory.[11] The other half of participants, directly before being asked their conversation preferences, were told (truthfully) that "later, we'll ask you to remember some of the answers you just gave (e.g., in how many

[11]Participants were aware that there was a short second survey a week later, though its purpose was not disclosed during the first survey.

classes you named did you have a male professor, do well vs poorly, etc.). You'll have a better chance of winning the bonus if you answer those questions right." They then had to correctly answer a comprehension question reiterating these facts. Thus, these participants were aware both that the study was about memory and also were informed of the exact question they would be asked (about the share of classes that were successes vs failures).[12]

Beliefs and Recall After having their three conversations, the first survey of the study ended. A week later, participants were invited back to take a short second survey that elicited their beliefs about and recall of the original set of nine classes. The second survey began by asking participants to "Think back to all the classes we had you name during the first survey you took, including **both** the classes you had conversations about **and** the ones you did not have conversations about." A block of questions then asked participants to make incentivized binary bets about whether a randomly selected one of these classes had each of the binary characteristics that the first survey asked about, including the main variable indicating whether they felt they did well or poorly in it. If one of these questions was the randomly chosen response that determined their bonus, participants' earned a $2.00 bonus if their bet was correct. A separate block of questions asked probabilistic versions of these same questions: e.g., "What do you think is the percent chance that the randomly selected class is one you said you did... well in? poorly in?" If one of these questions was the randomly chosen response that determined their bonus, participants' earned a $2.00 bonus if their answer was within five percentage points of the true percentage. These two blocks occurred in a random order.

After these two belief-elicitation blocks, participants faced an incentivized free-recall task. In particular, they were asked to write down the names of as many classes that they could remember listing in the first survey. If the free-recall question was chosen to determine participants' bonus, each correct response increased their chances of winning by 10 percentage points, while each incorrect response reduced this chance by 10 percentage points. Of course, their chances of winning could not go above 100% or below 0%. Their free-recall answers were graded by having GPT-4 assess whether each class they listed in the recall task matched any of the classes they originally listed in the first survey (they need not have word-for-word matched).

Note that collecting separate high-quality free-recall data in addition to beliefs is crucial for attributing biased beliefs to biased memory. My reliance on such data as the primary measure of what participants remember is, to my knowledge, novel to the economics literature

[12]Participants were also asked beliefs about the other elicited characteristics of classes, such as the share of classes they said had a male professor. This part of the warning was thus also truthful.

on self-serving memory. Many studies use as their measure of recall what participants *guess* about the number of various experiences they had in the past. Absent corroborating free-recall evidence, however, such measures can easily be misleading. As an example, suppose an agent has an exaggerated prior about how often she is trustworthy for reasons orthogonal to memory. An experiment has her play prisoners' dilemma games and then after a delay asks how often she chose to cooperate in them. Suppose her memory is completely unbiased: for each game (no matter what she chose) there is a fixed probability per unit time that it will be forgotten. Then the agent will initially be unbiased; she can perfectly remember the games immediately after playing them, so her biased priors plays no role. Over time, however, she will appear to have a "biased memory": her guess of how many games she cooperated in will revert to her prior. To see this, note that in the limit where she has completely forgotten everything about the experiment, her belief will simply be her original prior, which by assumption was biased for non-memory reasons. Thus separate measures of beliefs and recall are crucial to attributing biases in expectations to biases in memory.

Logistics A total of 439 participants completed the first survey.[13] From these, I first drop 29 participants whose rehearsal preferences were randomly chosen to be implemented, because for such respondents I do not have random variation in rehearsal. I then drop 24 participants who gave any duplicate class names in the first survey. Finally, I drop 16 participants who did not complete the second survey, reflecting a 96% rate of taking the second survey. This set of 370 remaining participants constistutes the final sample. The median respondent spent 38 minutes on the first survey and 6 minutes on the second survey. One of the incentivized questions (across the two surveys) was randomly selected to determine participants' bonus.

4 Results

Beliefs and Recall I begin by describing some basic patterns in the beliefs and recall data. For all participants, the true share of classes from the first survey that were successes was either four or five out of nine: that is, 44% or 56%. The left panel of Figure 1 shows the distribution of participants' beliefs about this share in the second survey. We see a wide range of beliefs, indicating that participants cannot perfectly think back on the first survey and report the actual distributions of experiences from it. Indeed, in the free-recall

[13]The original planned sample size was 600 participants. Fewer than that number successfully completed the first survey due to a server crash that made the chatbot unresponsive for some participants. I include only respondents who did not encounter any technical issues stemming from this because those who did encounter this issue were not invited to take the second survey.

task the average participant only remembers 4.9 of the nine classes.[14] The middle panel of Figure 1 calculates for each participant the share of the classes they recalled that were successes vs failures, and then plots the distribution of this variable. We see wide variation in this measure of what participants recall. The right panel of Figure 1 shows that these two variables—beliefs and recall—are tightly linked in the cross-section: participants who recall more successes than failures in turn believe the original set of experiences contained a greater share of successes.

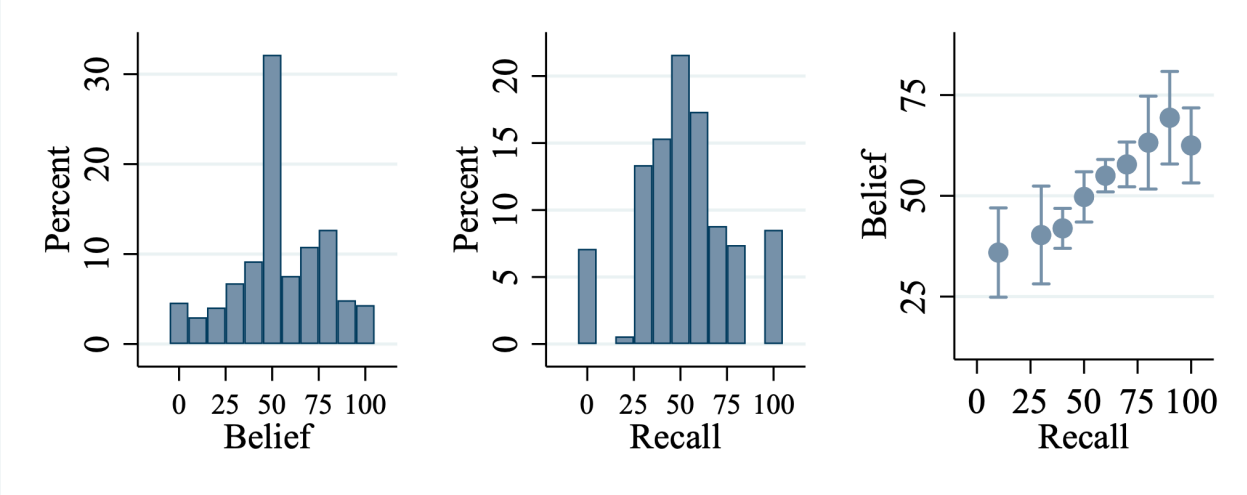

Figure 1: Beliefs and Recall

Notes: The left panel shows the distribution of participants' beliefs in the second survey about the percent of classes they listed in the first survey that were successes vs failures. The middle panel shows the distribution of the share of recalled classes that are successes. The right panel shows a binscatter correlating beliefs and recall.

Rehearsal Next, I ask to what extent the rehearsal manipulation in the first survey affected recall in the second survey. Note that all participants listed nine classes in the first survey. Of these, either four or five were successes, and therefore five or four were failures. The rehearsal manipulation involved conversations about three classes, all successes in the *Successes* treatment or all failures in the *Failures* treatment. Thus, every participant had at least one success and one failure that they did *not* have a conversation about.

Recall that there are two sources of randomization that the experimental design allows me to exploit. First, participants randomly discuss either three successes or three failures

[14]Throughout the results, unless otherwise noted, I focus on the classes that participants recall correctly. For 4.6% of entries in the free-recall task, participants wrote a class name that does not (according to GPT-4) correspond to one of the classes they listed in the first survey. Further, among recalled classes that did appear in the first survey 93.2% are correctly identified as being a success/failure. None of the results presented are sensitive to including/excluding these "false memories."

depending on their treatment status. Second, within treatment, which classes the conversations focus on are chosen randomly from among all the successes (in the *Successes* treatment) or all the failures (in the *Failures* treatment). I start by looking at this second source of variation, comparing recall rates within treatment across experiences.

Panel A of Figure 2 shows recall rates within each treatment for different types of classes. First, we see that within the *Successes* treatment, participants recall 47.5% of the successes they do *not* have conversations about but 72.3% of the successes they do have conversations about. This difference of 24.8 percentage points is significant at the $p < 0.01$ level. We see very similar within-treatment differences in recall rates for the *Failures* treatment. Such participants recall 46.4% of the failures they did not have conversations about but 71.9% of those they did have converastions about. This 25.5 percentage point effect is significantly distiguishable from zero ($p < 0.01$) and indistinguishable from the 24.8 percentage point effect on successes in the *Successes* treatment ($p = 0.89$).

My primary focus in the recall data is on whether participants remember classes or not, but I also observe the *order* in which which classes were recalled. To analyze these data, I define the first class a participant writes down in the free-recall task as having an order of one, the second an order of two, and so on. Panel B of Figure 2 shows, among classes that participants recalled, within-treatment differences in the average order in which they were recalled. We see that within the *Successes* treatment, classes participants had conversations about are recalled 1.1 places sooner than successes they did not have conversations about ($p < 0.01$). Similarly, in the *Failures* treatment, participants recall failures they had conversations about 0.96 places sooner than failures they did not have conversations about ($p < 0.01$). These two treatment effects are not statistically distinguishable from each other ($p = 0.66$). Thus, rehearsing experiences makes them more likely to come to mind in the future but also to come to mind *faster* in the future.

The above results showed effects of conversations on later recall at the class-by-class level. Of course, the two treatments differed in whether these conversations focused on successes or failures. We should therefore expect treatment effects at the individual level on the total number of successes and failures that participants recall. Indeed, the first pair of bars in Figure 3 shows that those in the *Successes* treatment recall on average 2.9 successes, 0.6 more than those in the *Failures* treatment ($p < 0.01$). Similarly, the second pair of bars shows that those in *Failures* recall 2.8 failures, 0.9 more than those in *Successes* ($p < 0.01$). Transforming these numbers, the third pair of bars shows that 63% of the classes participants in *Successes* recall are successes, 20.9 percentage points more than those in *Failures* ($p < 0.01$). Thus, shifting rehearsal toward successes in turn systematically shifts the composition of participants' recall toward successes.

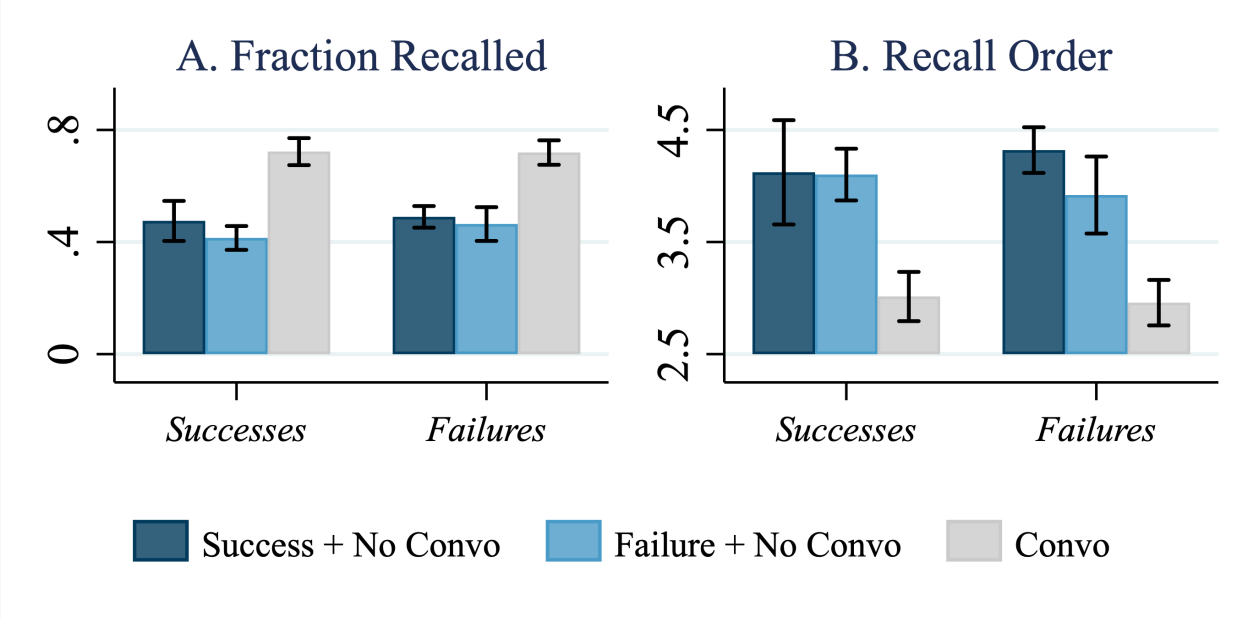

Figure 2: Effects of Rehearsal on Recall

Notes: Panel A shows recall rates by treatment status. Panel B shows the average order of recalled classes by treatment status, where lower numbers indicate earlier recall. "Success + No Convo" refers to successful classes that participants did not have conversations about. "Failure + No Convo" refers to unsuccessful classes that participants did not have conversations about. "Convo" refers to classes participants did have conversations about, which were all successes in *Successes* and failures in *Failures*. Whiskers show 95% confidence intervals. Table 1 shows the underlying numbers for this figure.

My primary measure of recall includes only those classes that participants name in the second survey *and* correctly remember whether they were initially classified as a success or a failure in the first survey. A natural question is whether the rehearsal manipulation in my experiment led participants to *mis*remember successes as failures or vice versa. Table A.I shows only limited such effects, however. Only 6.8% of recalled classes are misclassified, with participants being somewhat more likely to misidentify failures as successes than vice versa. Conversations improve misclassification rates, but these effects are much smaller than the effects on recall described above: participants are 3.4 p.p. less likely to mislabel successes as failures and 4.7 p.p. less likely to mislabel failures as successes. Correspondingly, Table A.II shows that treatment variation across participants only has a small effect on the number of misclassified successes and failures. The effect on the number of "recalled" false successes and false failures is only 13.1% and 7.5% the effect on true successes and failures, respectively. I conclude that the subjective classification of classes into successes vs failures is therefore not driving any of the main results, in part because participants are extremely internally consistent in labeling classes within my experiment.

***Ex Post* Sophistication vs Naivete** All of the results presented thus far are consistent with a fully sophisticated agent reacting rationally to her memory constraints when forming beliefs. I now turn to discussing whether rehearsal biases beliefs in addition to recall. That is, do the distortions of recall from rehearsal in turn distort participants' beliefs about the share of classes that were successes vs failures? Proposition 1 stated that belief biases arise from rehearsal effects if and only if agents are naive *ex post* about them. That is, a sophisticated agent who understands that what she can recall partly reflects what she has rehearsed more can nonetheless form accurate beliefs (in expectation). A naive agent will instead take what she recalls at face value: i.e., as representative of the true distribution of her experiences.

The fourth pair of bars in Figure 3 shows participants' average beliefs about the share of the original set of classes that were successes rather than failures. Those in *Successes* on average believe 61.8% of classes were successes, while those in *Failures* believe only 45.3% were successes. This treatment effect of 16.5 percentage points ($p < 0.01$) is 79% of the 20.9 percentage point effect on recall that we saw above (third pair of bars in Figure 3) and is statistically indistinguishable from it ($p = 0.16$).[15] I therefore cannot reject that agents are fully naive *ex post* about rehearsal's effect on their recall. Rehearsal therefore biases beliefs in addition to memory.

Recall that in addition to eliciting probabilistic beliefs, the experiment also asked participants to make a binary bet on whether a randomly selected class was a success or failure. The final pair of bars in Figure 3 shows very large treatment effects of the rehearsal manipulation on these bets. Those in *Successes* are 47.6 percentage points ($p < 0.01$) more likely to be the random class is a success than those in *Failures*.

In addition to stating beliefs and making bets on whether experiences were successes or failures, participants were also asked about four other attributes of the classes they listed.[16] Table A.III shows similar effects on recall, beliefs, and bets of being randomly assigned to have conversations about classes with more or fewer attributes. That is, being randomly assigned to have conversations about more classes with a female instructor, for example, makes participants more likely to recall such classes in the second survey as well as boosts their beliefs and willingness to bet on whether the randomly selected class had a female instructor.

Endogenous Rehearsal Choices Thus far, I have focused on the effect of exogenously assigning rehearsal patterns to participants in order to measure their effects on recall and

[15] This p-value is from a seemingly unrelated regression comparing treatment effects on each variable.

[16] These included the gender of the instructor, whether they took the class in their first two years of high school/college, whether it involved a final project, whether it was part of their major (for college classes), and whether it was an elective class (for high-school classes).

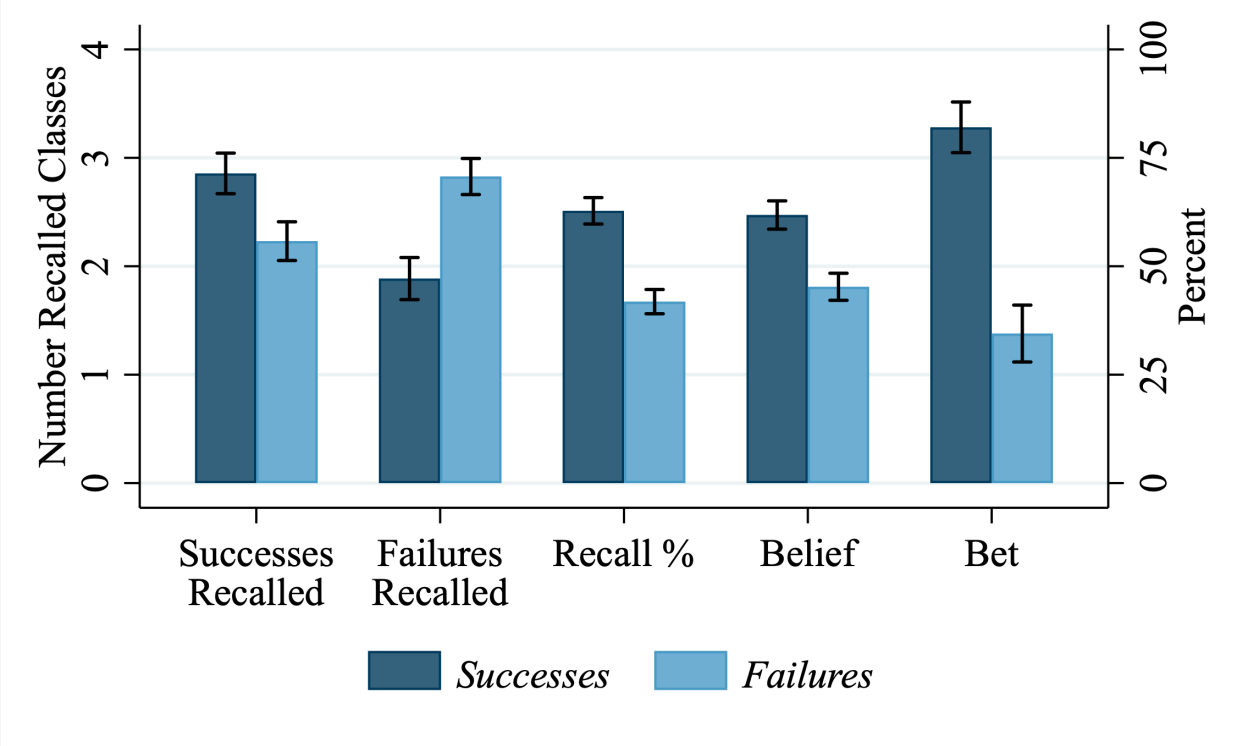

Figure 3: Effects of Rehearsal on Recall and Beliefs

Notes: The first two pairs of bars from the left show the number of successes and failures that participants correctly recalled in the free-recall task. The third pair of bars shows that average percent of recalled classes that were successes. The fourth pair shows the average believe about the percent of classes that were successes. The final pair shows the percent of participants who bet the randomly selected class was a success. Table 2 shows the underlying numbers. Whiskers show 95% confidence intervals.

beliefs. I now consider participants' rehearsal preferences to understand what memory-based belief biases (if any) arise endogenously. Recall that participants could express a preference for having conversations about either three successes, three failures, or three randomly selected classes. The light blue bars of figure 4 shows the share of participants choosing each of these options among those who were *not* aware that they would later face monetary incentives to form accurate beliefs. We see that participants are over four times more likely (44.6% vs 10.0%) to choose to discuss successes vs failures, with the remainder choosing to discuss randomly selected classes.

Because many more participants choose to discuss successes than choose to discuss failures, and doing so has large impacts on beliefs, participants are in effect (if not intentionally) choosing to have biased beliefs on average. To make this claim precise, I calculate what I call participants' "endogenous" beliefs: the beliefs they would have formed had their rehearsal preferences been implemented.[17] Figure 4 shows that on average participants choose "overop-

[17]More precisely, define $ChosenBelief_i = Belief_i + (ChosenSuccess_i - Successes_i) \cdot TE/3$, where $Belief_i$

timistic beliefs", meaning they make rehearsal choices that predictably lead them to believe that more of the original experiences were successes.

The dark blue bars of Figure 4 show rehearsal choices and endogenous beliefs among those participants who were warned that they would later face incentives for accurate beliefs. We see no statistically significant differences in any of these variables when comparing those who were vs were not warned about incentives. Recall from Corollary 2 that incentives for accurate beliefs induce the agent to change her rehearsal strategy (and therefore her eventual beliefs) only if she is sophisticated *ex ante* about its effects on recall. In particular, such agents should be less willing to make rehearsal choices that bias their beliefs. Choosing to rehearse three randomly selected classes mechanically induces no correlation between rehearsal and success, in contrast to rehearsing only successes or only failures, which therefore should lead to unbiased beliefs (see Proposition 1). However, we see no increase in the fraction of participants selecting this option, with if anything slightly fewer choosing the random option among those warned about incentives. These data therefore do not provide any evidence of *ex ante* sophistication about rehearsal effects.

While Figure 4 shows no evidence of *ex ante* sophistication regarding the rehearsal choice I can measure (which experiences to have conversations about), in principle *ex ante* sophisticated agents could take other steps to mitigate the effect this choice would have on their beliefs. For example, such an agent, upon being told that she will have conversations about successes, might choose to otherwise rehearse or improve her memory of failures to offset the foreseen effect the conversations will have on her memory of successes. If so, we should expect attenuated treatment effects among participants who were warned of the incentives for belief accuracy. To study this possibility, Figure 5 shows treatment effects on recall, beliefs, and the binary bet, splitting the sample by whether they were warned or not about the incentives for belief accuracy. We see no significant differences in treatment effects on the number of successes or failures recalled, on the share of recalled experiences that were successes, on the mean belief about the share of successes, or on the fraction of participants who bet the random class is a success. I therefore find no evidence of *ex ante* sophistication either directly in rehearsal choices or indirectly in later recall and beliefs.

What Drives Rehearsal Choices Because participants appear naive both *ex post* and *ex ante* about rehearsal effects, strategic motives related to managing their future beliefs seem not to determine their rehearsal choices in my experiment. To provide some qualitative evidence on what does drive these choices, the experiment simply asked participants

is i's actual belief about the percent of successes, $ChosenSuccess_i$ is the number of successes they preferred to have conversations about, $Successes_i$ is the actual number of successes they were randomly selected to discuss, and TE is the estimated treatment effect (from Figure 3) on beliefs of being assigned to *Successes*.

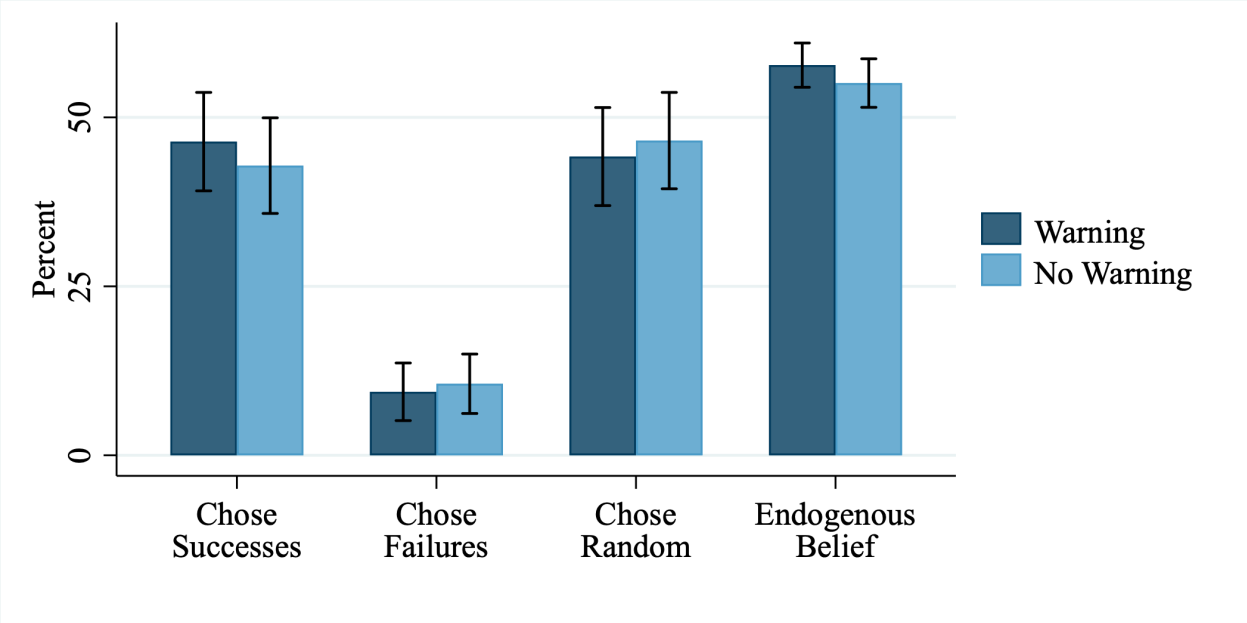

Figure 4: Effects of Incentives Warnings on Rehearsal Choices

Notes: The first three pairs of bars from the left show the fraction of participants who expressed that they wanted to have conversations about three successes, three failures, or a randomly selected three classes. The rightmost pair of bars shows the average implied "chosen" belief: the belief participants would have held about the share of successes had their rehearsal choice been implemented. "Warning" and "No Warning" refer to participants who did and did not receive a warning that they would face incentives for accurate beliefs. Table 3 shows the underlying numbers from this figure. Whiskers show 95% confidence intervals.

to explain in their own words why they selected the rehearsal option they did. I then have GPT-4o, a large language model, code whether their response mentions any of five different considerations (described more below). These five considerations were chosen in an *ex post* fashion, but I also allow GPT-4o to respond that the participant mentions any other consideration. A total of 92.7% of participants are coded as mentioning at least one of these five considerations, and only 14.1% mention anything else (6.8% mention both one of the five considerations as well as something else). These five considerations therefore appear to comprise the very large majority of what participants mention.

Figure 6 shows the fraction of participants who are coded as mentioning each consideration, split by which rehearsal option they selected. We see that 25.7% of participants mention that it is "more enjoyable to discuss positive memories than negative ones". Perhaps unsurprisingly, almost all of these participants had selected to discuss successes. This consideration was meant to capture in-the-moment utility from thinking back on positive experiences (rather than, say, a desire to produce self-confidence in the future), and many of participants' responses that are coded this way appear in line with this interpretation. For example, some such responses include "It's more pleasant to think about times I performed

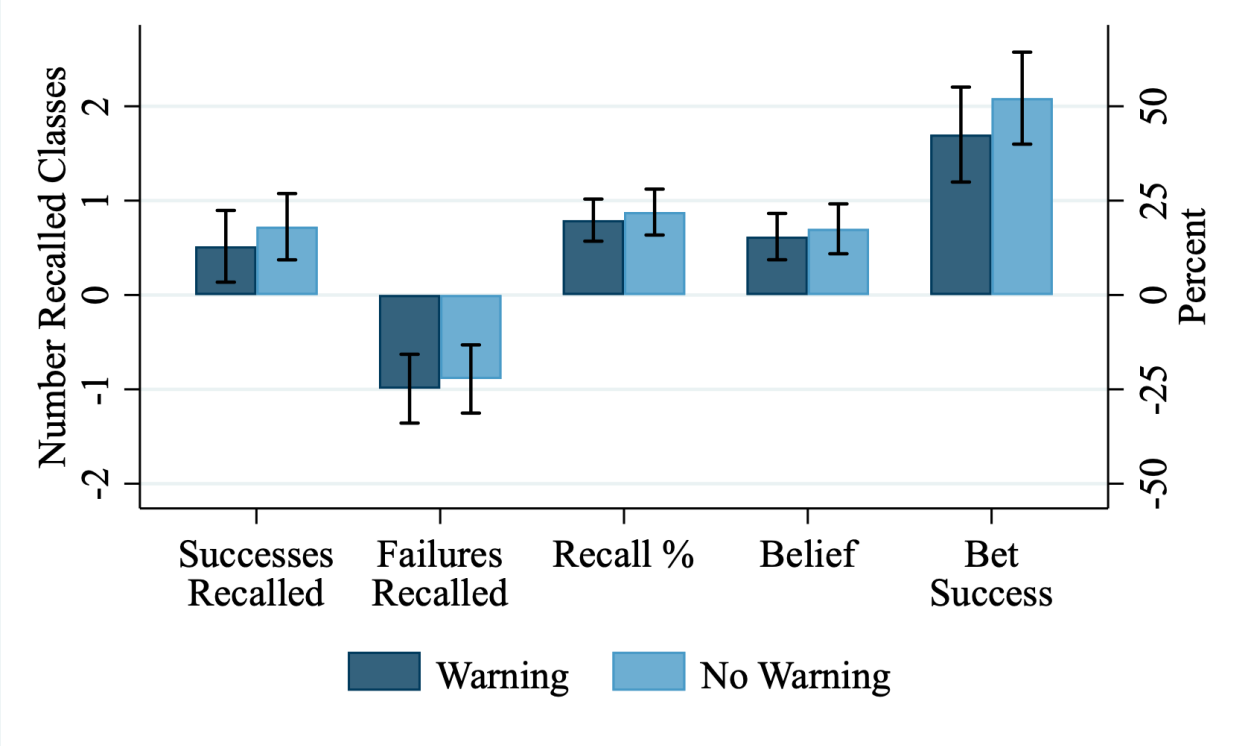

Figure 5: Effects of Rehearsal on Recall and Beliefs

Notes: This figure shows treatment effects of being assigned to the *Successes* treatment, splitting the sample by those who were or were not warned about later incentives for accurate beliefs. The first two pairs of bars from the left show effects on the number of successes and failures that participants correctly recalled in the free-recall task. The third pair of bars shows effects on the percent of recalled classes that were successes. The fourth pair shows effects on beliefs about the percent of classes that were successes. The final pair shows effects on whether participants bet that the randomly selected class was a success. Table 4 shows the underlying regressions. Whiskers show 95% confidence intervals.

well than times I did not...", "I feel good about how well I did in the classes and prefer to focus on the positive...", and "They bring back great memories."

Next, 37.0% of participants mention "how easy it would be to engage thoughtfully in conversations". This category was chosen to reflect the monetary incentives participants faced in the conversations, where their responses were graded according to how thoughtful they were. For example, some such responses include "I feel that these classes are the easiest to remember to have a informed discussion about" and "I believe it will be easier for me to respond thoughtfully." These respondents also tend to favor discussing successes.

Next, 30.0% of participants mention being indifferent about the conversation topics, and perhaps unsurprisingly almost all choose to discuss a random set of classes. Another substantial driver of choosing random classes is that is "more enjoyable to discuss a wide variety of things". Many such participants mention wanting to keep the conversations interesting or

mention the value of surprise in conversation topics.

One natural concern with the analysis above showing that rehearsal choices do not respond to incentives for accurate recall is that the incentives in the experiment may simply not be large enough. For example, many participants might be aware of the distortionary impact of rehearsal on their future beliefs, but they might nonetheless choose to discuss successes because the in-the-moment intrinsic utility stakes dwarf any concern about the effect on beliefs. My results of course cannot conclusively rule out such a story, but I can investigate whether many participants mention such a tradeoff while explanating their choice. In fact, almost no participants (3.2%) are coded as mentioning "anything about how their conversations will have an effect on what they will later remember or believe in the study," which I included to capture any expressed sophistication about rehearsal. Even among these 12 responses, none expresses the logic of rehearsal in a clear way.[18] Thus, if participants are in fact sophisticated about rehearsal but simply do not face high enough incentives, none expresses this sentiment. Further, Table 5 shows that the share of participants mentioning each consideration does not meaningfully vary depending on whether they were warned about incentives for belief accuracy. This is despite the incentive warning occurring immediately prior to rehearsal choices (to maximize its salience during this decision). Finally, participants do seem motivated by the monetary incentives in the experiment, as evidenced by many participants choosing conversation topics that they felt would better allow them to earn a high score (and the associated better chance of a bonus payment) that came from more thoughtful engagement in the conversations.

To conclude, two factors (besides indifference) seem to explain the majority of participants' rehearsal choices in this context. First, they respond to monetary incentives, tending to select rehearsal options they feel will help them earn more.[19] Second, they enjoy thinking about positive experiences that reflect well on them. These combine in my context to produce overoptimistic beliefs a week later by distorting which experiences come to mind.

[18]The closest such participant writes "Less chance of me hyper focusing on a specific, better chance at memory recall in my opinion".

[19]Of course, participants may have an intrinsic preference as well for engaging conversations.

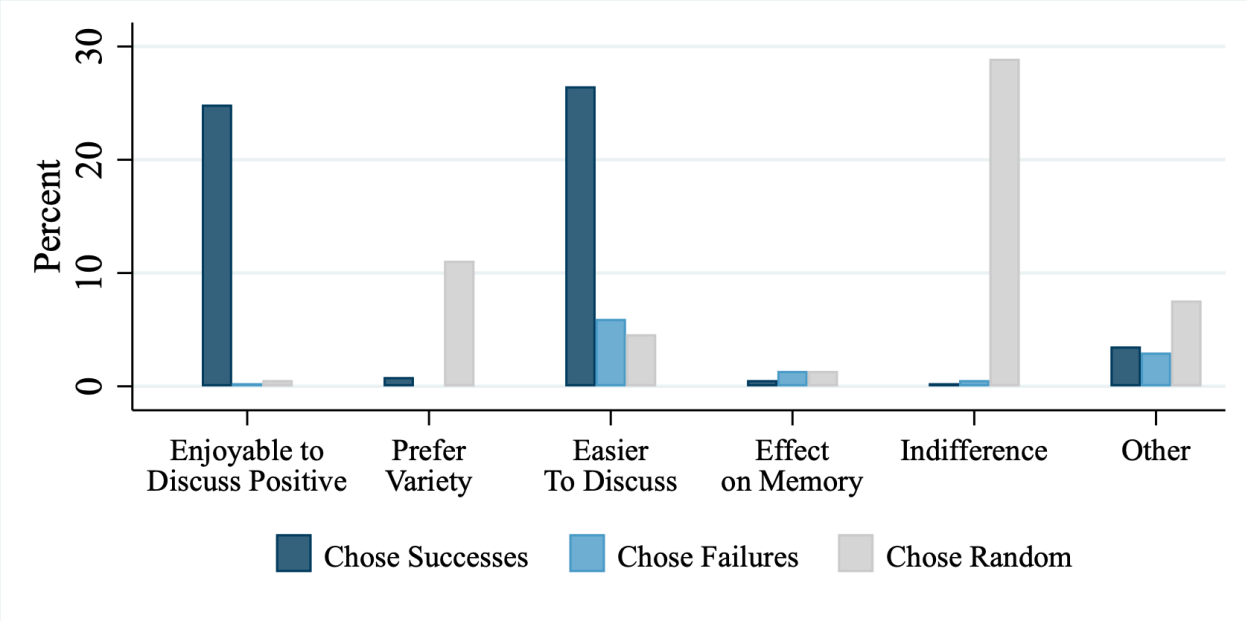

Figure 6: Reasons behind Rehearsal Choices

Notes: The figure shows the share of participants mentioning each of five considerations in their explanation of how they made their choice about which rehearsal option to select. The first panel of Table 5 shows the numbers underlying this figure.

5 Implications: Optimism and Pessimism

We have seen that a preference for thinking about positive experiences biases beliefs through rehearsal. In this section, I briefly explore the implications of these results for whether agents become overly optimistic about the future and for the choices they might make in response. Following the results from the experiment, I focus on the rehearsal-based utility benchmark described in Section 2. Recall that such an agent is naive both *ex post* and *ex ante* about rehearsal ($\lambda_2 = \lambda_1 = 0$), has no belief-based utility ($\alpha = 0$), and intrinsically prefers to rehearse experiences with a high y_m value ($u_a = \nu \sum a_m y_m$), which I interpret as positive experiences.

I also generalize the model from Section 2 to allow more than a single period of experiences and rehearsal. In particular, now assume there are $T+2$ periods $t \in \{0, 1, ..., T, T+1\}$. In each period $t \leq T$, the agent has a set of experiences M_t. I assume that in each period the characteristic y_m has the same variance σ_y^2 but with a true mean $\bar{y}_t$ that is random. In particular, let $\bar{y}_t = \mu + \kappa t + \epsilon_t$, where $\epsilon \sim N(0, \sigma_\epsilon^2)$: that is, in the first period $t = 0$ in expectation the average of y_m is μ with a variance in that average of σ_ϵ^2. Further, this expectation potentially evolves in a linear way, changing by κ each period.

In addition to having experiences in each period, the agent also chooses much attention

$a_{m,t}$ to pay to each of her previous experiences $m \in \bigcup_{t=0}^{t-1} M_t$. As before, I assume that in these rehearsal periods, the agent has no memory constraints (adding this complication would simply amplify the effect of rehearsal). In period $T+1$, the agent must then use her memory to form beliefs $E_a[\bar{y}_t]$ about the average value of y_m within each previous time period. I assume that rehearsal effects take the following form: $r(m,a) \propto \exp\{\gamma \sum_{t=1}^{T} a_{m,t}\}$. Proposition 3 then follows.

Proposition 3 *Beliefs about the most recent period are unbiased. The agent is increasingly overoptimistic about periods further in the past. As a result, on average she perceives the time trend in $\bar{y}_t$ to be more negative (or less positive) than the true trend κ.*

$$E_{a^*}[\bar{y}_t] = \bar{y}_t + (T-t)\frac{\gamma\nu}{\phi}\sigma_y^2$$
$$E\Big[E_{a^*}[\bar{y}_t] - E_{a^*}[\bar{y}_{t-1}]\Big] = \kappa - \frac{\gamma\nu}{\phi}\sigma_y^2$$

The intuition behind Proposition 3 is straightforward. The agent has had more opportunities to rehearse experiences that happened longer ago. Her tendency to focus on positive experiences each period therefore creates larger distortions in her memory about more distant experiences. In contrast, she has had no opportunity to differentially rehearse experiences in the most recent period T, so her beliefs about that period are correct. An implication of this increasing optimism about the more distant past is that from the agent's perspective, the time trend in the perceived average value of experiences tends to be less steeply positive (or more steeply negative) than the true trend κ. If in fact there is no underlying time trend ($\kappa = 0$), the agent tends to perceive things to be getting worse, with the earliest period she can remember on average seeming best and the most recent period seeming worst. Indeed, in surveys, people overwhelmingly tend to report that the best time in the US, in terms of arts and music, happy family lives, moral society, and political harmony, happened to coincide with their own childhood. Conversely, people across generations tend to agree that the worst time in America along all these dimensions happens to be now (Dam, 2024).

What implications does this increasing over-optimism regarding the past have for the agent's beliefs about the future? The answer depends on what the agent believes about the dynamics of $\bar{y}_t$. Suppose the agent's priors for both μ (the expected average y_m in period $t=0$) and κ (the time trend) are normally distributed:

$$\begin{pmatrix}\mu\\ \kappa\end{pmatrix} \sim N\left(\begin{pmatrix}\tilde{\mu}\\ \tilde{\kappa}\end{pmatrix}, \begin{pmatrix}\sigma_\mu^2 & 0\\ 0 & \sigma_\kappa^2\end{pmatrix}\right)$$

Then she can use the apparent values of $\bar{y}_t$ in the past to update her beliefs about μ and

κ and therefore her forecast about the future. More precisely, Proposition 4 follows.

Proposition 4 *Let $E_{a^*}[\bar{y}_{T+1}]$ be the agent's forecast of the average value of y_m in period $T+1$. Assume the agent is sufficiently uncertain about the initial expected value of experiences ($\sigma_\mu \to \infty$) and that her priors are unbiased ($\tilde{\mu} = \mu$ and $\tilde{\kappa} = \kappa$).*

Then rehearsal ($\gamma > 0$) makes her overoptimistic about the future ($E_{a^*}[\bar{y}_{T+1}] > \bar{y}_{T+1}$) on average if and only if she is sufficiently confident she understands how $E[\bar{y}_t]$ evolves over time: i.e., if and only if σ_κ is sufficiently small.

To see the intuition behind Proposition 4, consider a case where in fact the distribution of experiences is constant over time. Figure 7 shows an example of the actual past history of $\bar{y}$ (gray dots) along with the agent's rehearsal-enhanced belief about it (blue dots). Suppose first that the agent knows the value of κ for sure ($\sigma_\kappa = 0$), which is the case portrayed in the middle panel of Figure 7. For example, the agent might know that her expected ability is roughly constant over time but not know it's true level. Though she tends to perceive more recent periods as having been less positive than more distant periods, her knowledge that there is no underlying trend leads her to attribute this fact to recent bad luck. She concludes that the average experience is overly positive which, absent a belief that this average might be changing, pushes her to believe that her ability in the future will be higher than it in fact will be. That is, she expects reversion to an excessively high mean. In this case, overoptimism about the past leads to overconfidence about the future.

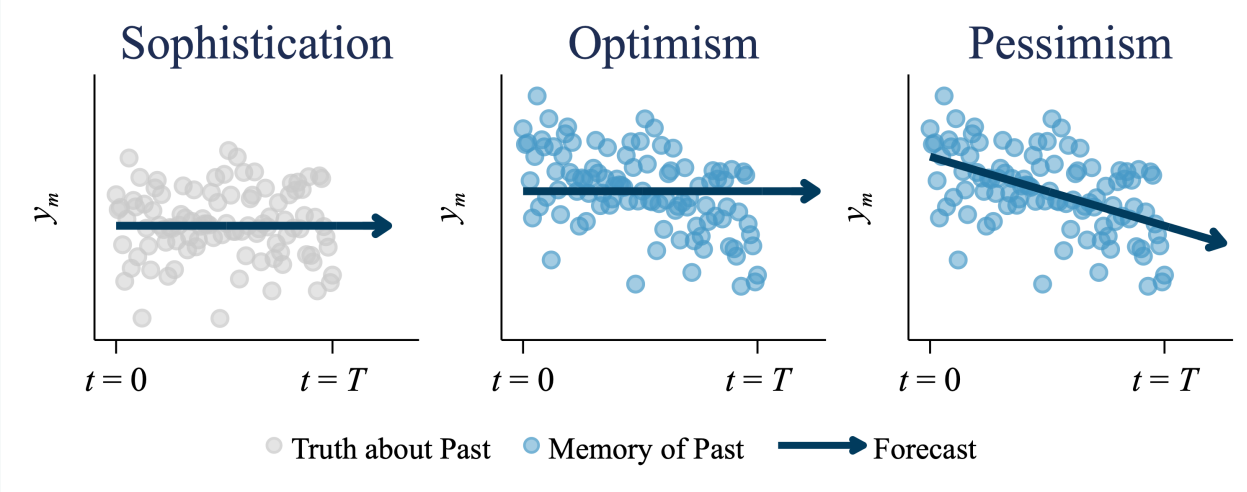

Figure 7: Optimism vs Pessimism

Next, consider a case where the agent is sufficiently uncertain about the extent to which the average experience is changing over time (σ_κ large enough), such as in the right panel

of Figure 7. This agent will then partially interpret the fact that more recent periods seem worse than earlier ones as evidence of a negative (or less positive) trend. Her beliefs about the most recent period $t = T$ are still unbiased, and therefore her underestimation of the time trend makes her overly pessimistic or underconfident about the future despite the fact that she is weakly overoptimistic about every previous period. Such an agent looks nostalgic for the past and skeptical that the negative apparent trend will reverse itself.

I conclude by sketching how applications of Proposition 4 can generate predictions about how rehearsal might distort choices in various economic domains. First, consider a citizen deciding whether to vote out her country's incumbent leaders. We can analyze this case by relabeling terms above. Let $t = 0$ denote the current administration's first term in office and $\bar{y}_t$ denote quality of life during time t. We can then interpret κ as the incumbents' skill in governing (i.e., more skilled leaders on average improve quality of life). Rehearsing good times in the past will gradually lead voters to believe that things were better in the past than they truly were: i.e., that things are getting worse under the current administration. If σ_κ is large enough (sufficient initial uncertainty about the new leaders' ability), then voters will gradually attribute this apparent deterioration to their current leaders' poor governing abilities. Such voters would then be willing to replace their current leaders for even below-average ($\tilde{\kappa} < \kappa$) new ones, whom they will of course in turn vote out in due time. This result resonates with survey evidence showing that the US public's approval of presidents tends to decrease while they are in office and then increase in the years after they leave office (Beck et al. 2012, Jones 2023).

Next, suppose an agent is considering investing in self-improvement, such as by beginning to go to therapy. Let $\bar{y}_t$ be on average how she was feeling at time t. Suppose absent the investment, the agent knows her average well-being is not changing over time ($\sigma_\kappa = 0$). The investment involves a continuing per-period cost, which I assume would boost κ by an uncertain amount only if her average well-being is below some threshold (e.g., therapy is more helpful if one has certain mental health conditions). First, such an agent will be overly hesitant to start investing in self-improvement: she believes her average well-being is higher than it truly is because she tends to rehearse positive memories. Though she has a realistic assessment of more recent periods, she will view these as being unlucky deviations from a high mean. She will thus tend to underestimate the benefits of self-improvement.

In addition to being unlikely to invest in self-improvement to begin with, however, such an agent will also be excessively likely to *abandon* it once starting. By assumption, the self-improvement investment has an uncertain impact on κ. The agent's tendency to be increasingly overoptimistic about the past will thus lead her to conclude, after beginning to invest in self-improvement, that it must not be very effective (else things would be getting

better, not worse). She is therefore excessively likely to quit using it. Of course, she would have perceived a decline even absent the investment, but in that case (as we have seen) she would instead have attributed it to temporary bad luck. Thus, the agent tends to reject and abandon investments that have uncertain impacts on her well-being.

Finally, Proposition 4 can lead to a form of status-quo bias through a type of information trap. Suppose a consumer repeatedly purchases a product that provides a mix of good and bad experiences with an unknown but non-time-varying expected value ($\sigma_\kappa = 0$). For example, she might choose to go on a similar vacation every year. Each purchase provides a new signal about the expected value of the good. But in addition, by rehearsing positive experiences the agent eventually comes to believe that her previous purchases were better than they were: i.e., her perception of her existing signals grows more optimistic over time. Suppose the product is in fact only mediocre (the true μ is small) but that agent's first few purchases happened to be better than average, and so she continued to buy the good for several periods. An agent like this might never switch to purchasing a different good (with a new expected value μ) despite receiving an infinite stream of informative (and typically negative) signals. She thus exhibits a form of status-quo bias, sticking with the same products, job, city, relationships, or habits despite plenty of evidence that a change is in order.

6 Conclusion

In this paper, I show experimentally that rehearsing experiences generates belief biases by distorting future memory. Further, agents appear to have intrinsic preferences over which experiences to rehearse: they enjoy thinking about past experiences that reflect well on them. Acting on this preference, combined with naivete about rehearsal, generates overoptimistic beliefs. These results provide a novel non-strategic mechanism behind seemingly motivated beliefs. Further, we saw that while rehearsal-based overoptimism about the past often naturally produces optimism about the future, uncertainty can reverse this prediction, with novel applications in a range of simple economic settings.

More speculatively, these results open the way to asking new questions about memory and belief biases. First, they raise the question of what other preferences and incentives might drive rehearsal patterns and therefore produce belief biases. Most obviously, people naturally need to attend to their own immediate environments and experiences more than those of others. To the extent that individual environments and experiences differ, this fact will naturally create polarization in beliefs, through memory. People will react more to prices or signals they personally have to think about more (Simonsohn & Loewenstein 2006, Georganas et al. 2014, D'acunto et al. 2021, Hartzmark et al. 2021), will believe that

their own (or their family's) wages and jobs are more typical than they are (Cruces et al. 2013, Jäger et al. 2023, Conlon & Patel 2023), or will believe that information or evidence they personally uncovered (and therefore spent more time thinking about) outweighs what others have found (Thompson & Loewenstein 1992, Schwardmann & van der Weele 2019; Schwardmann et al. 2022, Conlon et al. 2022). Of course, many explanations for such polarization are possible, but my results suggest that a simple rehearsal-based story—things we spend more time thinking about later come to mind more easily—may partly contribute.

Second, while this study focuses on preferences for thinking about experiences, bottom-up forces play a key role in determining what grabs our attention independently of our preferences (Bordalo et al. 2022). Exploring how such factors might produce later belief biases by altering rehearsal patterns remains an important question for future work. One natural hypothesis is that extreme past events—which contrast highly with other experiences—may attract attention and therefore loom larger in the future than their true frequency would suggest. In addition, some mental health conditions like depression are partly defined as an inability to avoid ruminating on negative experiences, which (combined with rehearsal effects and naivete) could produce low self-confidence for affected individuals.

Finally, the participants in my experiment appear almost fully naive about rehearsal effects both *ex ante* and *ex post*, but this cannot be universally true. Students use flashcards to study for tests, leveraging rehearsal effects to boost their academic performance. When is the effect of attention on recall clear enough that agents take it into account, either exploiting it to improve/change their memories or accounting for it *ex post*? Answering this question would aid both in predicting when (and which) belief biases arise as well as in understanding how to design policies to counteract rehearsal's distortionary effects.

References

Adler, O., & Pansky, A. (2020, 1). A "rosy view" of the past: Positive memory biases. In (p. 139-171). Elsevier. doi: 10.1016/B978-0-12-816660-4.00007-6

Amelio, A., & Zimmermann, F. (2023). Motivated memory in economics-a review. *Games*. Retrieved from www.econtribute.de

Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. (1995). Biased judgments of fairness in bargaining. *American Economic Review*, *85*, 1337-1343.

Barron, K., Huck, S., & Jehiel, P. (2024, 8). Everyday econometricians: Selection neglect and overoptimism when learning from others. *American Economic Journal: Microeconomics*, *16*, 162-198. Retrieved from https://pubs.aeaweb.org/doi/10.1257/mic.20200030 doi: 10.1257/mic.20200030

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018, 7). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory and Cognition*, *46*, 796-808. doi: 10.3758/s13421-018-0805-9

Beck, J. W., Carr, A. E., & Walmsley, P. T. (2012, 10). What have you done for me lately? charisma attenuates the decline in u.s. presidential approval over time. *Leadership Quarterly*, *23*, 934-942. doi: 10.1016/j.leaqua.2012.06.002

Bluck, S., & Alea, N. (2002). Exploring the functions of autobiographical memory: Why do i remember the autumn? In J. D. Webster & B. K. Haight (Eds.), *Critical advances in reminiscence work: From theory to application* (pp. 61–75). New York: Springer.

Bodoh-Creed, A. L. (2020). Mood, memory, and the evaluation of asset prices. *Review of Finance*, *24*, 227-262. doi: 10.1093/rof/rfz001

Bohren, A., Hascher, J., Imas, A., Ungeheuer, M., & Weber, M. (2024). *A cognitive foundation for perceiving uncertainty.* Retrieved from https://ssrn.com/abstract=4706147

Bolte, L., & Fan, T. Q. (2024, 1). Motivated mislearning: The case of correlation neglect. *Journal of Economic Behavior and Organization*, *217*, 647-663. doi: 10.1016/j.jebo.2023.11.020

Bolte, L., & Raymond, C. (2024). *Emotional inattention.*

Bordalo, P., Burro, G., Coffman, K., Gennaioli, N., & Shleifer, A. (2024). Imagining the future: Memory, simulation, and beliefs. *Review of Economic Studies*.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2023a). How people use statistics. *NBER Working Paper No. 31631*. Retrieved from https://www.socialscienceregistry.org/trials/11166

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2023b, 12). Memory and probability. *The Quarterly Journal of Economics*, *138*, 265-311. doi: 10.1093/qje/qjac031

Bordalo, P., Gennaioli, N., & Shleifer, A. (2022). Salience. *Annual Review of Economics*, 1-42.

Brodie, D. A., & Murdock, B. B. (1977). Effect of presentation time on nominal and functional serial-position curves of free recall. *Journal of Verbal Learning and Verbal Behavior*, *16*, 5-7.

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal expectations. *American Economic Review*.

Bushong, B., Rabin, M., & Schwartzstein, J. (2021). A model of relative thinking. *Review of Economic Studies*.

Bénabou, R. (2015). The economics of motivated beliefs. *Revue d'Économie Politique*, *125*(5), 665-685. doi: 10.3917/redp.255.0665

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

Cepeda, N. J., Pashler, H., Vul, E., Rohrer, D., & Wixted, J. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*.

Chew, S. H., Huang, W., & Zhao, X. (2020). Motivated false memory. *Journal of Political Economy*, *128*, 3913-3939. doi: 10.1086/709971

Conlon, J. (2024). Attention, Information, and Persausion. *Working paper*.

Conlon, J., Mani, M., Rao, G., Ridley, M. W., & Schilbach, F. (2022). Not learning from others. *NBER Working Paper 30378*. Retrieved from http://www.nber.org/papers/w30378

Conlon, J., & Patel, D. (2023). *What jobs come to mind? stereotypes about fields of study.*

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research 1. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.

Cruces, G., Truglia, R. P., & Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*. Retrieved from http://hdl.handle.net/10419/51652

Dam, A. V. (2024). America's best decade, according to data. *The Washington Post*. Retrieved from https://www.washingtonpost.com/business/2024/05/24/when-america-was-great-according-data/

Drobner, C. (2022, 3). Motivated beliefs and anticipation of uncertainty resolution. *American Economic Review: Insights*, *4*, 89-105. doi: 10.1257/aeri.20200829

D'acunto, F., Malmendier, U., Ospina, J., & Weber, M. (2021). Exposure to grocery prices and inflation expectations. *Journal of Political Economy*. doi: 10.1086/713192

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*, 114-138. doi: 10.1257/mic.3.2.114

Engelmann, J. B., Lebreton, M., Salem-Garcia, N. A., Schwardmann, P., & Weele, J. J. V. D. (2024). Anticipatory anxiety and wishful thinking. *American Economic Review*.

Enke, B. (2020, 8). What you see is all there is. *Quarterly Journal of Economics*, *135*, 1363-1398. doi: 10.1093/qje/qjaa012

Enke, B., Schwerter, F., & Zimmerman, F. (2024). *Associative memory and belief formation.*

Farina, A., Fréchette, G. R., Lizzeri, A., & Perego, J. (2024). The selective disclosure of evidence: An experiment. *Working paper*.

Georganas, S., Healy, P. J., & Li, N. (2014, 4). Frequency bias in consumers' perceptions of inflation: An experimental study. *European Economic Review*, *67*, 144-158. doi: 10.1016/j.euroecorev.2014.01.014

Graeber, T., Roth, C., & Zimmermann, F. (2023). *Stories, statistics, and memory.*

Gödker, K., Jiao, P., & Smeets, P. (2024). Investor memory. Retrieved from https://ssrn.com/abstract=3348315

Haaland, I., Roth, C., Stantcheva, S., & Wohlfart, J. (2024). Measuring what is top of mind. *ECONtribute Discussion Paper No. 298*. Retrieved from www.econtribute.de

Hartzmark, S. M., Hirshman, S. D., & Imas, A. (2021, 8). Ownership, learning, and beliefs. *Quarterly Journal of Economics*, *136*, 1665-1717. doi: 10.1093/qje/qjab010

Huffman, D., Raymond, C., & Shvets, J. (2022, 10). Persistent overconfidence and biased memory: Evidence from managers. *American Economic Review*, *112*, 3141-3175. doi: 10.1257/aer.20190668

Jin, G. Z., Luca, M., & Martin, D. (2021). Is no news (perceived as) bad news? an experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, *13*, 141-173. doi: 10.1257/mic.20180217

Jones, J. M. (2023). Retrospective approval of jfk rises to 90%; trump at 46%. *Gallup News*. Retrieved from https://news.gallup.com/poll/508625/retrospective-approval-jfk-rises-trump.aspx

Jäger, S., Roth, C., Roussille, N., & Schoefer, B. (2023). Worker beliefs about outside options. *NBER Working Paper 29623*. Retrieved from http://www.nber.org/papers/w29623

Kahana, M. J. (2012). *Foundations of human memory*. OUP USA.

Karlsson, N., Loewenstein, G., & Seppi, D. (2009, 4). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, *38*, 95-115. doi: 10.1007/s11166-009-9060-6

Koszegi, B., & Szeidl, A. (2013). A Model of Focusing in Economic Choice. *Quarterly Journal of Economics*, 53–104. doi: 10.1093/qje/qjs049.Advance

Link, S., Peichl, A., Roth, C., & Wohlfart, J. (2023). Attention to the macroeconomy. *ECONtribute Discussion Paper No. 256*. Retrieved from www.econtribute.de

Loewenstein, G., & Wojtowicz, Z. (2023). *The economics of attention.* Retrieved from https://ssrn.com/abstract=4368304

Madigan, S. A. (1969). *Intraserial repetition and coding processes in free recall* (Vol. 8).

Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*. Retrieved from https://academic.oup.com/qje/article/117/3/735/1932979

Möbius, M. M., Niederle, M., & Niehaus, P. (2022). Managing self-confidence. *Management Science*.

Müller, M. W. (2022). *Selective memory around big life decisions.*

Quispe-Torreblanca, E., Gathergood, J., Loewenstein, G., & Stewart, N. (2022). Attention utility: Evidence from individual investors.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, *17*.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63-77.

Saucet, C., & Villeval, M. C. (2019, 9). Motivated memory in dictator games. *Games and Economic Behavior*, *117*, 250-275. doi: 10.1016/j.geb.2019.05.011

Schwardmann, P., Tripodi, E., & van der Weele, J. J. (2022, 4). Self-persuasion: Evidence from field experiments at international debating competitions. *American Economic Review*, *112*, 1118-1146. doi: 10.1257/aer.20200372

Schwardmann, P., & van der Weele, J. (2019, 10). *Deception and self-deception* (Vol. 3). Nature Research. doi: 10.1038/s41562-019-0666-7

Sial, A. Y., Sydnor, J. R., & Taubinsky, D. (2023). Biased memory and perceptions of self-control. *NBER Working Paper 30825*. Retrieved from http://www.nber.org/papers/w30825

Simonsohn, U., & Loewenstein, G. (2006). Mistake 37: The effect of previously encountered prices on current demand. *The Economic Journal*, 175-199. Retrieved from https://academic.oup.com/ej/article/116/508/175/5087768

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1589-1625.

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Proceses*, *51*, 176-197.

Walker, W. R., Skowronski, J. J., Gibbons, J. A., Vogl, R. J., & Ritchie, T. D. (2009). Why people rehearse their memories: Frequency of use and relations to the intensity of emotions associated with autobiographical memories. *Memory*, *17*, 760-773. doi: 10.1080/09658210903107846

Zimmermann, F. (2020). *The dynamics of motivated beliefs* (Vol. 110). American Economic Association. doi: 10.1257/aer.20180728

Table 1: Effects of Rehearsal on Recall

	Recalled Class			Order of Recall		
	Pooled	*Successes*	*Failures*	Pooled	*Successes*	*Failures*
	(1)	(2)	(3)	(4)	(5)	(6)
Conversation	0.721	0.723	0.719	2.984	3.014	2.959
	(0.016)	(0.025)	(0.022)	(0.076)	(0.112)	(0.104)
Success + No Conversation	0.487	0.475	0.490	4.280	4.122	4.320
	(0.017)	(0.037)	(0.020)	(0.096)	(0.238)	(0.104)
Failure + No Conversation	0.428	0.414	0.464	4.047	4.102	3.919
	(0.018)	(0.022)	(0.031)	(0.098)	(0.118)	(0.175)
Observations	3,330	1,503	1,827	1,819	792	1,027
Individuals	370	167	203	359	162	197
p-value: Convo = Success + No Convo	0.00	0.00	0.00	0.00	0.00	0.00
p-value: Convo = Failure + No Convo	0.00	0.00	0.00	0.00	0.00	0.00
p-value: No Convo, Success = Failure	0.01	0.08	0.43	0.09	0.94	0.06

Notes: Table shows average recall rates, with standard errors in parentheses. Columns 1 and 4 show rates for the sample as a whole, columns 2 and 5 restrict to the *Successes* treatment, and columns 3 and 6 restrict to the *Failures* treatment. "Conversation" shows recall rates for the classes participants had conversations about, while "Success + No Conversation" and "Failure + No Conversation" show rates for the successes and failures that participants did not have conversations about. *, **, and *** show significance at the $p < 0.10$, 0.05, and 0.01 levels respectively.

Table 2: Effect of Rehearsal on Recall and Beliefs

	Successes Recalled (1)	Failures Recalled (2)	Recall % (3)	Belief (4)	Bet (5)
Successes	0.625***	-0.941***	0.209***	0.165***	0.476***
	(0.132)	(0.131)	(0.021)	(0.023)	(0.045)
Constant	2.232***	2.828***	0.418***	0.453***	0.345***
	(0.091)	(0.085)	(0.014)	(0.016)	(0.033)
Observations	370	370	359	370	370

Notes: Table shows OLS regressions. The dependent variables are the number of successes participants correctly recalled (column 1), the number of failures they correctly recall (column 2), the fraction of recalled classes that were successes (column 3), participants' beliefs about the share of classes that were successes (column 4), and whether participants' bet that the randomly selected classes was a success (column 5). *Successes* is an indicator variable for having conversations about three successes (rather than three failures). Robust standard errors in parentheses. *, **, and *** show significance at the $p < 0.10$, 0.05, and 0.01 levels respectively.

Table 3: Effect of Knowing about Belief Incentives on Rehearsal Choices

	Chose Successes (1)	Chose Random (2)	Chose Failures (3)	"Chosen" Belief (4)
Warned about Incentives	0.036	-0.012	-0.024	0.027
	(0.052)	(0.031)	(0.052)	(0.025)
Constant	0.429***	0.106***	0.466***	0.551***
	(0.036)	(0.022)	(0.036)	(0.018)
Observations	370	370	370	370

Notes: Table shows OLS regressions. The dependent variables are indicators for whether participants' chose to discusses three successes, a random three classes, or three failures (columns 1-3, respectively), and the implied belief from their rehearsal choice had it been implemented (column 4). The latter is calculated as $ChosenBelief_i = Belief_i + (ChosenSuccess_i - Successes_i) \cdot TE/3$, where $Belief_i$ is i's actual belief about the percent of successes, $ChosenSuccess_i$ is the number of successes they preferred to have conversations about, $Successes_i$ is the actual number of successes they were randomly selected to discuss, and TE is the estimated treatment effect (from Figure 3) on beliefs of being assigned to *3 Successes*. Robust standard errors in parentheses. *, **, and *** show significance at the $p < 0.10$, 0.05, and 0.01 levels respectively.

Table 4: Effects of Rehearsal on Recall and Beliefs, Depending on Warning about Belief Incentives

	Successes Recalled		Failures Recalled		Recall %		Belief		Bet	
	NW	W	NW	W	NW	W	NW	W	NW	W
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Successes	0.724***	0.516***	-0.890***	-0.993***	0.220***	0.198***	0.175***	0.155***	0.521***	0.425***
	(0.179)	(0.194)	(0.184)	(0.186)	(0.031)	(0.028)	(0.034)	(0.031)	(0.062)	(0.064)
Constant	2.086***	2.388***	2.819***	2.837***	0.404***	0.434***	0.437***	0.469***	0.276***	0.418***
	(0.123)	(0.134)	(0.112)	(0.129)	(0.019)	(0.021)	(0.023)	(0.021)	(0.044)	(0.050)
Observations	189	181	189	181	187	172	189	181	189	181
p-value: NW = W		0.43		0.69		0.61		0.65		0.28

Notes: Table shows OLS regressions. The dependent variables are the number of successes participants correctly recalled (column 1-2), the number of failures they correctly recall (column 3-4), the fraction of recalled classes that were successes (column 5-6), participants' beliefs about the share of classes that were successes (column 7-8), and whether participants' bet that the randomly selected classes was a success (column 9-10). *Successes* is an indicator variable for having conversations about three successes (rather than three failures). "NW" and "W" indicate whether the sample is restricted to those who were not warned or who were warned, respectively, about future incentives for accurate beliefs. Robust standard errors in parentheses. *, **, and *** show significance at the $p < 0.10$, 0.05, and 0.01 levels respectively.

Table 5: Reasons for Rehearsal Choices

	Chose Successes	Chose Failures	Chose Random
	Full Sample		
Enjoy Discuss Success	0.249	0.003	0.005
Prefer Variety	0.008	0.000	0.111
Easier to Discuss	0.265	0.059	0.046
Affect Later Memory	0.005	0.014	0.014
Indifferent	0.003	0.005	0.289
Other	0.035	0.030	0.076
	No Warning about Belief Incentives		
Enjoy Discuss Success	0.217	0.005	0.011
Prefer Variety	0.011	0.000	0.116
Easier to Discuss	0.280	0.063	0.058
Affect Later Memory	0.005	0.011	0.016
Indifferent	0.000	0.011	0.275
Other	0.048	0.026	0.095
	Warning about Belief Incentives		
Enjoy Discuss Success	0.282	0.000	0.000
Prefer Variety	0.006	0.000	0.105
Easier to Discuss	0.249	0.055	0.033
Affect Later Memory	0.006	0.017	0.011
Indifferent	0.006	0.000	0.304
Other	0.022	0.033	0.055

Notes: Table shows the share of participants who select each rehearsal option and mention each consideration when justifying their choice. The top panel shows data for the whole sample, while the lower two panels restrict to those who were not warned or who were warned about future incentives for accurate beliefs.

A Additional Tables and Figures

Table A.I: Misremembering is Rare

	Misclassified		
	Pooled	*Successes*	*Failures*
	(1)	(2)	(3)
Conversation	0.040	0.016	0.058
	(0.008)	(0.007)	(0.013)
Success + No Conversation	0.058	0.050	0.060
	(0.009)	(0.020)	(0.011)
Failure + No Conversation	0.126	0.135	0.105
	(0.017)	(0.021)	(0.026)
Observations	1,952	853	1,099
Individuals	363	165	198
p-value: Convo = Success + No Convo	0.11	0.12	0.89
p-value: Convo = Failure + No Convo	0.00	0.00	0.10
p-value: No Convo, Success = Failure	0.00	0.00	0.12

Notes: Table shows average rates at which classes were recalled but incorrectly (i.e., as a success if initially a failure or *vice versa*), with standard errors in parentheses. Column 1 shows rates for the sample as a whole, column 2 restricts the sample to the *Successes* treatment, and column 3 restricts the sample to the *Failures* treatment. "Conversation" shows recall rates for the classes participants had conversations about, while "Success + No Conversation" and "Failure + No Conversation" show rates for the successes and failures that participants did not have conversations about.

Table A.II: Small Treatment Effects on Misremembering

	False Successes (1)	False Failures (2)
Successes	0.082	-0.071**
	(0.061)	(0.033)
Constant	0.212***	0.143***
	(0.038)	(0.026)
Observations	370	370

Notes: Table shows OLS regressions. The dependent variables are the number of failures participants incorrectly recalled as being successes (column 1) and the number of successes they incorrectly recalled as being failures (column 2). *Successes* is an indicator variable for having conversations about three successes (rather than three failures). Robust standard errors in parentheses. *, **, and *** show significance at the $p < 0.10$, 0.05, and 0.01 levels respectively.

Table A.III: Effects of Rehearsal on Recall and Beliefs about Other Attributes

	Recalled w/ Attribute (1)	Recalled w/o Attribute (2)	Recall % (3)	Belief (4)	Bet (5)
# Conversations	0.216***	-0.295***	0.059***	2.271**	0.048***
	(0.051)	(0.056)	(0.008)	(0.954)	(0.018)
True #	0.473***	-0.536***	0.094***	5.069***	0.070***
	(0.028)	(0.025)	(0.003)	(0.442)	(0.008)
Constant	0.138**	5.489***	-0.004	27.000***	0.165***
	(0.060)	(0.127)	(0.007)	(1.470)	(0.025)
Observations	1,480	1,480	1,452	1,480	1,480
Individuals	370	370	363	370	370

Notes: Table shows OLS regressions. Each observation is an individual-by-class-attribute pair, where the class attributes are whether its instructor was female, whether it occurred in the first two years of high school/college, whether it had a final paper/project, whether it was an elective (for high school classes), and whether it was for participants' major (for college classes). The dependent variables are the number of classes with each attribute that participants correctly recalled (column 1), the number without each attribute they correctly recall (column 2), the fraction of recalled classes that had each attribute (column 3), participants' beliefs about the share of classes with each attribute (column 4), and whether participants' bet that the randomly selected classes had each attribute (column 5). "# Conversations" indicates how many of the classes participants were randomized to have conversations about had that attribute, and "True #" is the total number of classes with that attribute. Robust standard errors, clustered at the individual level, in parentheses. *, **, and *** show significance at the $p < 0.10$, 0.05, and 0.01 levels respectively.

B Proofs

I start by solving for the period-1 agent's beliefs about what her period-2 self's beliefs will be, which I denote by $\tilde{E}[\bar{y}|a]$. Recall that $\tilde{r}_t(m,a)$ is the agent's belief in period t about her probability of recalling m given attention a. Let and that $\tilde{r}_{1,2}(m,a)$ be her belief at time $t=1$ about what her future self at $t=2$ will believe this probability to be, which as in the main text I assume takes the form $\tilde{r}_{1,2}(m,a) \propto \exp \lambda_1\lambda_2\gamma a_m$. Then,

$$
\begin{aligned}
\tilde{E}[\bar{y}|a] &= \sum_m y_m \cdot \tilde{r}_1(m,a) \cdot \frac{1}{M\tilde{r}_{1,2}(m,a)} \\
\implies \frac{d\tilde{E}[\bar{y}|a]}{da_l} &= \sum_m y_m \left[\frac{d\tilde{r}_1(m,a)}{da_l} \frac{1}{M\tilde{r}_{1,2}(m,a)} + \tilde{r}_1(m,a) \frac{d\frac{1}{M\tilde{r}_{1,2}(m,a)}}{da_l} \right] \\
&= \sum_m y_m \left[\frac{1}{M\tilde{r}_{1,2}(m,a)} \frac{d}{da_l} \frac{\exp\{\lambda_1\gamma a_m\}}{\sum_k \exp\{\lambda_1\gamma a_k\}} + \frac{1}{M}\tilde{r}_1(m,a) \frac{d}{da_l} \frac{\sum_k \exp\{\lambda_1\lambda_2\gamma a_k\}}{\exp\{\lambda_1\lambda_2\gamma a_m\}} \right] \\
&= y_l \left[\frac{1}{M\tilde{r}_2(l,a)} \frac{d}{da_l} \frac{\exp\{\gamma\lambda_1 a_l\}}{\sum_k \exp\{\lambda_1\gamma a_k\}} + \frac{1}{M}\tilde{r}_1(l,a) \frac{d}{da_l} \frac{\sum_k \exp\{\lambda_1\lambda_2\gamma a_k\}}{\exp\{\lambda_1\lambda_2\gamma a_l\}} \right] \\
&+ \sum_{m\neq l} y_m \left[\frac{1}{M\tilde{r}_{1,2}(m,a)} \frac{d}{da_l} \frac{\exp\{\lambda_1\gamma a_m\}}{\sum_k \exp\{\lambda_1\gamma a_k\}} + \frac{1}{M}\tilde{r}_1(m,a) \frac{d}{da_l} \frac{\sum_k \exp\{\lambda_1\lambda_2\gamma a_k\}}{\exp\{\lambda_1\lambda_2\gamma a_m\}} \right] \\
&= y_l \left[\frac{1}{M\tilde{r}_2(l,a)} \frac{\sum_k \exp\{\lambda_1\gamma a_k\}\lambda_1\gamma \exp\{\lambda_1\gamma a_l\} - \exp\{\lambda_1\gamma a_l\}\lambda_1\gamma \exp\{\lambda_1\gamma a_l\}}{\sum_k \exp\{\gamma\lambda_1 a_k\}^2} \right. \\
&\left. + \frac{1}{M}\tilde{r}_1(l,a) \frac{\exp\{\lambda_1\lambda_2\gamma a_l\}\lambda_1\lambda_2\gamma \exp\{\lambda_1\lambda_2\gamma a_l\} - \sum_k \exp\{\lambda_1\lambda_2\gamma a_k\}\lambda_1\lambda_2\gamma \exp\{\lambda_1\lambda_2\gamma a_l\}}{\exp\{\lambda_1\lambda_2\gamma a_l\}^2} \right] \\
&+ \sum_{m\neq l} y_m \left[\frac{1}{M\tilde{r}_2(l,a)} \frac{-\exp\{\lambda_1\gamma a_m\}\lambda_1\gamma \exp\{\lambda_1\gamma a_l\}}{\sum_k \exp\{\lambda_1\gamma a_k\}^2} \right. \\
&\left. + \frac{1}{M}\tilde{r}_1(m,a) \frac{\exp\{\lambda_1\lambda_2\gamma a_m\}\lambda_1\lambda_2\gamma \exp\{\lambda_1\lambda_2\gamma a_l\}}{\exp\{\lambda_1\lambda_2\gamma a_l\}^2} \right]
\end{aligned}
$$

Note that when $a=0$, $\tilde{r}_1(m,a) = \tilde{r}_{1,2}(m,a) = \frac{1}{M}$ for all m. Thus,

$$
\begin{aligned}
\frac{d\tilde{E}[\bar{y}|a]}{da_k}|_{a=0} &= y_l \left[\frac{M\lambda_1\gamma - \lambda_1\gamma}{M^2} + \frac{1}{M^2} \frac{\lambda_1\lambda_2\gamma - M\lambda_1\lambda_2\gamma}{1} \right] + \sum_{m\neq l} y_m \left[-\frac{\lambda_1\gamma}{M^2} + \frac{1}{M^2}\frac{\lambda_1\lambda_2\gamma}{1} \right] \\
&= \frac{\lambda_1\gamma}{M}(1-\lambda_2) \left[y_l - \frac{1}{M}\sum_m y_l \right] \\
&= \lambda_1(1-\lambda_2)\frac{\gamma}{M}(y_l - \bar{y})
\end{aligned}
$$

Then, taking a first-order approximation around $a = 0$,

$$\begin{aligned}
\tilde{E}[\bar{y}|a] &\approx \bar{y} + \lambda_1(1-\lambda_2)\frac{\gamma}{M}\sum_m (y_m - \bar{y})a_m \\
&= \bar{y} + \gamma\lambda_1(1-\lambda_2)E[(y_m - \bar{y})a_m] \\
&= \bar{y} + \gamma\lambda_1(1-\lambda_2)\Big(E[y_m a_m] - E[y_m]E[a_m]\Big) \\
&= \bar{y} + \gamma\lambda_1(1-\lambda_2)\text{Cov}(y,a)
\end{aligned}$$

We can then quickly prove Proposition 1 by supposing that the period-1 self is sophisticated ($\lambda_1 = 1$) since such an agent will be correct about what her future self will believe. That is, recall that

$$E[\bar{y}|a] = \sum_m y_m \cdot r(m,a) \cdot \frac{1}{M\tilde{r}_2(m,a)}$$

Note that, when $\lambda_1 = 1$, $\tilde{r}_1(m,a) = r(m,a)$ and $\tilde{r}_{1,2}(m,a) = \tilde{r}_2(m,a)$. Therefore, we can simply redo the above derivation but letting $\lambda_1 = 1$ to get that

$$E[\bar{y}|a] \approx \bar{y} + \gamma(1-\lambda_2)\text{Cov}(y,a)$$

which is Proposition 1.

Next, I prove Proposition 2. The first-order conditions for a_m are

$$\nu_m = \phi a_m^* - \Big(\alpha - \beta(\tilde{E}[\bar{y}|a^*] - \bar{y})\Big)\frac{d\tilde{E}[\bar{y}|a^*]}{da_m} \tag{8}$$

$$\nu_m = \phi a_m^* - \Big(\alpha - \beta\lambda_1(1-\lambda_2)\frac{\gamma}{M}\sum_k (y_k - \bar{y})a_k^*\Big)\frac{\gamma}{M}\lambda_1(1-\lambda_2)(y_m - \bar{y})$$

$$\nu_m = \phi a_m^* - \alpha\frac{\gamma}{M}\lambda_1(1-\lambda_2)(y_m - \bar{y}) + \beta\left[\lambda_1(1-\lambda_2)\frac{\gamma}{M}\right]^2 (y_m - \bar{y})\sum_k (y_k - \bar{y})a_k^*$$

$$\phi a_m^* = \nu_m + (y_m - \bar{y})\left[\alpha\frac{\gamma}{M}\lambda_1(1-\lambda_2) - \beta\Big(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\Big)^2\sum_k (y_k - \bar{y})a_k^*\right]$$

$$\phi\sum_m a_m^*(y_m - \bar{y}) = \sum_m \nu_m(y_m - \bar{y}) + \sum_m (y_m - \bar{y})^2\left[\alpha\frac{\gamma}{M}\lambda_1(1-\lambda_2) - \beta\Big(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\Big)^2\sum_k (y_k - \bar{y})a_k^*\right]$$

$$\sum_m a_m^*(y_m - \bar{y}) = \frac{\sum_m \nu_m(y_m - \bar{y}) + \sum_m (y_m - \bar{y})^2\left[\alpha\frac{\gamma}{M}\lambda_1(1-\lambda_2)\right]}{\phi + \beta\sum_m (y_m - \bar{y})^2\Big(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\Big)^2}$$

$$\sum_m a_m^*(y_m - \bar{y}) = \frac{\mathrm{Cov}(y,\nu) + \mathrm{Var}(y)\left[\alpha\frac{\gamma}{M}\lambda_1(1-\lambda_2)\right]}{\frac{\phi}{M} + \beta\mathrm{Var}(y)\Big(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\Big)^2}$$

We can then plug this expression into the formula for $\tilde{E}[\bar{y}|a]$ to get that

$$\tilde{E}[\bar{y}|a] = \bar{y} + \lambda_1(1-\lambda_2)\frac{\gamma}{M}\frac{\mathrm{Cov}(y,\nu) + \alpha\mathrm{Var}(y)\left[\frac{\gamma}{M}\lambda_1(1-\lambda_2)\right]}{\frac{\phi}{M} + \beta\mathrm{Var}(y)\Big(\lambda_1(1-\lambda_2)\frac{\gamma}{M}\Big)^2}$$

This, plus a simple rearrangement of equation 8, yields Proposition 2. Corollaries 1 to 3 then follow straightforwardly from Proposition 2.

Next, Proposition 3 follows directly from Corollary 3 by considering the attention-based utility benchmark described there.

Next I prove Proposition 4. Let $\bar{y} = \begin{pmatrix}\bar{y}_0 & \bar{y}_1 & \dots & \bar{y}_T\end{pmatrix}'$. Let $\xi = \begin{pmatrix}\mu \\ \kappa\end{pmatrix}$ be the true

parameter vector. Let

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{pmatrix}$$

.

Then $\bar{y} = X\xi + \epsilon$, where $\epsilon = \begin{pmatrix} \epsilon_0 & \epsilon_1 & ... & \epsilon_T \end{pmatrix}'$.

Let Σ_T and ξ_T be the agents' posteriors about Σ and ξ after observing $\bar{y}$.

This is then a problem of Bayesian linear regression. The standard solution is then $\Sigma_T^{-1} = \Sigma^{-1} + \frac{X^T X}{\sigma_\epsilon^2}$ and

$$\begin{aligned} \xi_T &= \Sigma_T \left(\Sigma^{-1} \tilde{\xi} + \frac{X^T \bar{y}}{\sigma_\epsilon^2} \right) \\ &= \Sigma_T \begin{pmatrix} \frac{\tilde{\mu}}{\sigma_\mu^2} + \frac{1}{\sigma_\epsilon^2} \sum_t \bar{y}_t \\ \frac{\tilde{\kappa}}{\sigma_\kappa^2} + \frac{1}{\sigma_\epsilon^2} \sum_t t\bar{y}_t \end{pmatrix} \end{aligned}$$

We can then calculate Σ_T^-1:

$$\begin{aligned} \Sigma_T^{-1} &= \begin{pmatrix} \frac{1}{\sigma_\mu^2} & 0 \\ 0 & \frac{1}{\sigma_\kappa^2} \end{pmatrix} + \frac{1}{\sigma_\epsilon^2} \begin{pmatrix} 1 & 1 & ... & 1 \\ 0 & 1 & ... & T \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{pmatrix} \\ \Sigma_T^{-1} &= \begin{pmatrix} \frac{1}{\sigma_\mu^2} & 0 \\ 0 & \frac{1}{\sigma_\kappa^2} \end{pmatrix} + \frac{1}{\sigma_\epsilon^2} \begin{pmatrix} T+1 & \sum_{t=0}^T t \\ \sum_{t=0}^T t & \sum_{t=0}^T t^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sigma_\mu^2} + \frac{T+1}{\sigma_\epsilon^2} & \frac{T(T+1)}{2\sigma_\epsilon^2} \\ \frac{T(T+1)}{2\sigma_\epsilon^2} & \frac{1}{\sigma_\kappa^2} + \frac{T(T+1)(2T+1)}{6\sigma_\epsilon^2} \end{pmatrix} \end{aligned}$$

To simplify expressions, I use ψ to denote the corresponding precision for each σ parameter: i.e., $\psi_X \equiv \frac{1}{\sigma_X^2}$. Then,

$$\Sigma_T^{-1} = \begin{pmatrix} \psi_\mu + \psi_\epsilon(T+1) & \psi_\epsilon \frac{T(T+1)}{2} \\ \psi_\epsilon \frac{T(T+1)}{2} & \psi_\kappa + \psi_\epsilon \frac{T(T+1)(2T+1)}{6} \end{pmatrix}$$

$$\implies \Sigma_T = \frac{1}{det(\Sigma_T^{-1})} \begin{pmatrix} \psi_\kappa + \psi_\epsilon \frac{T(T+1)(2T+1)}{6} & -\psi_\epsilon \frac{T(T+1)}{2} \\ -\psi_\epsilon \frac{T(T+1)}{2} & \psi_\mu + \psi_\epsilon(T+1) \end{pmatrix}$$

$$\begin{aligned} det(\Sigma_T^{-1}) &= \Big(\psi_\mu + \psi_\epsilon(T+1)\Big)\Big(\psi_\kappa + \psi_\epsilon \frac{T(T+1)(2T+1)}{6}\Big) - \psi_\epsilon^2 \frac{T^2(T+1)^2}{4} \\ &= \psi_\mu\psi_\kappa + \psi_\epsilon\psi_\kappa(T+1) + \psi_\mu\psi_\epsilon \frac{T(T+1)(2T+1)}{6} + \frac{1}{12}\psi_\epsilon^2 T(T+1)^2(T+2) > 0 \end{aligned}$$

We can now compute ξ_T:

$$\begin{aligned} \xi_T &= \Sigma_T \begin{pmatrix} \frac{\tilde{\mu}}{\sigma_\mu^2} + \frac{1}{\sigma_\epsilon^2}\sum_t \bar{y}_t \\ \frac{\tilde{\kappa}}{\sigma_\kappa^2} + \frac{1}{\sigma_\epsilon^2}\sum_t t\bar{y}_t \end{pmatrix} \\ &= \frac{1}{det(\Sigma_T^{-1})} \begin{pmatrix} \Big(\frac{\tilde{\mu}}{\sigma_\mu^2} + \frac{1}{\sigma_\epsilon^2}\sum_t \bar{y}_t\Big)\Big(\psi_\kappa + \psi_\epsilon \frac{T(T+1)(2T+1)}{6}\Big) - \Big(\frac{\tilde{\kappa}}{\sigma_\kappa^2} + \frac{1}{\sigma_\epsilon^2}\sum_t t\bar{y}_t\Big)\psi_\epsilon \frac{T(T+1)}{2} \\ \Big(\frac{\tilde{\kappa}}{\sigma_\kappa^2} + \frac{1}{\sigma_\epsilon^2}\sum_t t\bar{y}_t\Big)\Big(\psi_\mu + \psi_\epsilon(T+1)\Big) - \Big(\frac{\tilde{\mu}}{\sigma_\mu^2} + \frac{1}{\sigma_\epsilon^2}\sum_t \bar{y}_t\Big)\psi_\epsilon \frac{T(T+1)}{2} \end{pmatrix} \\ &= \frac{1}{det(\Sigma_T^{-1})} \begin{pmatrix} \Big(\psi_\mu\tilde{\mu} + \psi_\epsilon\sum_t \bar{y}_t\Big)\Big(\psi_\kappa + \psi_\epsilon \frac{T(T+1)(2T+1)}{6}\Big) - \Big(\psi_\kappa\tilde{\kappa} + \psi_\epsilon\sum_t t\bar{y}_t\Big)\psi_\epsilon \frac{T(T+1)}{2} \\ \Big(\psi_\kappa\tilde{\kappa} + \psi_\epsilon\sum_t t\bar{y}_t\Big)\Big(\psi_\mu + \psi_\epsilon(T+1)\Big) - \Big(\psi_\mu\tilde{\mu} + \psi_\epsilon\sum_t \bar{y}_t\Big)\psi_\epsilon \frac{T(T+1)}{2} \end{pmatrix} \end{aligned}$$

The predictive posterior should then be

$$\begin{aligned} \hat{y}_{T+1} &= \mu_T + (T+1)\kappa_T \\ &= \frac{1}{det(\Sigma_T^{-1})}\Bigg[\Big(\psi_\mu\tilde{\mu} + \psi_\epsilon\sum_t \bar{y}_t\Big)\Big(\psi_\kappa + \psi_\epsilon \frac{T(T+1)(2T+1)}{6}\Big) - \Big(\psi_\kappa\tilde{\kappa} + \psi_\epsilon\sum_t t\bar{y}_t\Big)\psi_\epsilon \frac{T(T+1)}{2} \\ &+ (T+1)\Bigg(\Big(\psi_\kappa\tilde{\kappa} + \psi_\epsilon\sum_t t\bar{y}_t\Big)\Big(\psi_\mu + \psi_\epsilon(T+1)\Big) - \Big(\psi_\mu\tilde{\mu} + \psi_\epsilon\sum_t \bar{y}_t\Big)\psi_\epsilon \frac{T(T+1)}{2}\Bigg)\Bigg] \end{aligned}$$

Recall however that when the agent thinks about periods $t = 0$ to $t = T$, she misperceives $\bar{y}_t$ to be $E_{a^*}[\bar{y}_t] = \bar{y}_t + (T-t)\frac{\gamma\nu}{\phi}\sigma_y^2$. Replacing $\bar{y}_t$ with these values in the above equation then yeilds the subjective predictive posterior:

$$
\begin{aligned}
E_{a^*}[\bar{y}_{T+1}] = \frac{1}{det(\Sigma_T^{-1})}\Bigg[&\Big(\psi_\mu\tilde{\mu} + \psi_\epsilon \sum_t \Big(\bar{y}_t + (T-t)\sigma_y^2\gamma\frac{\nu}{\phi}\Big)\Big)\Big(\psi_\kappa + \psi_\epsilon\frac{T(T+1)(2T+1)}{6}\Big) \\
&- \Big(\psi_\kappa\tilde{\kappa} + \psi_\epsilon \sum_t t\Big(\bar{y}_t + (T-t)\sigma_y^2\gamma\frac{\nu}{\phi}\Big)\Big)\psi_\epsilon\frac{T(T+1)}{2} \\
&+ (T+1)\Bigg(\Big(\psi_\kappa\tilde{\kappa} + \psi_\epsilon \sum_t t\Big(\bar{y}_t + (T-t)\sigma_y^2\gamma\frac{\nu}{\phi}\Big)\Big)\Big(\psi_\mu + \psi_\epsilon(T+1)\Big) \\
&- \Big(\psi_\mu\tilde{\mu} + \psi_\epsilon \sum_t \Big(\bar{y}_t + (T-t)\sigma_y^2\gamma\frac{\nu}{\phi}\Big)\Big)\psi_\epsilon\frac{T(T+1)}{2}\Bigg)\Bigg]
\end{aligned}
$$

We can then take the derivative of $E_{a^*}[\bar{y}_{T+1}]$ with respect to γ to learn the effect of rehearsal on this posterior:

$$
\begin{aligned}
\frac{dE_{a^*}[\bar{y}_{T+1}]}{d\gamma} &= \frac{1}{det(\Sigma_T^{-1})}\Bigg[\psi_\epsilon \sum_t (T-t)\sigma_y^2\frac{\nu}{\phi}\Big(\psi_\kappa + \psi_\epsilon\frac{T(T+1)(2T+1)}{6}\Big) - \psi_\epsilon\sum_t t(T-t)\sigma_y^2\frac{\nu}{\phi}\psi_\epsilon\frac{T(T+1)}{2} \\
&+ (T+1)\Bigg(\psi_\epsilon\sum_t t(T-t)\sigma_y^2\frac{\nu}{\phi}\Big(\psi_\mu + \psi_\epsilon(T+1)\Big) - \psi_\epsilon\sum_t (T-t)\sigma_y^2\frac{\nu}{\phi}\psi_\epsilon\frac{T(T+1)}{2}\Bigg)\Bigg] \\
&= \frac{\psi_\epsilon\sigma_y^2\frac{\nu}{\phi}}{det(\Sigma_T^{-1})}\Bigg[\sum_t (T-t)\Big(\psi_\kappa + \psi_\epsilon\frac{T(T+1)(2T+1)}{6}\Big) - \sum_t t(T-t)\psi_\epsilon\frac{T(T+1)}{2} \\
&+ (T+1)\Bigg(\sum_t t(T-t)\Big(\psi_\mu + \psi_\epsilon(T+1)\Big) - \sum_t (T-t)\psi_\epsilon\frac{T(T+1)}{2}\Bigg)\Bigg] \\
&= \frac{\psi_\epsilon\sigma_y^2\frac{\nu}{\phi}}{det(\Sigma_T^{-1})}\Bigg[\sum_t (T-t)\Big(\psi_\kappa + \psi_\epsilon\frac{T(T+1)(2T+1)}{6} - \psi_\epsilon\frac{T(T+1)^2}{2}\Big) \\
&+ (T+1)\Bigg(\sum_t t(T-t)\Big(\psi_\mu + \psi_\epsilon(T+1) - \psi_\epsilon\frac{T}{2}\Big)\Bigg)\Bigg] \\
&= \frac{\psi_\epsilon\sigma_y^2\frac{\nu}{\phi}}{det(\Sigma_T^{-1})}\Bigg[\sum_t (T-t)\Big(\psi_\kappa - \psi_\epsilon(T+1)T\frac{T+2}{6}\Big) \\
&+ (T+1)\Bigg(\sum_t t(T-t)\Big(\psi_\mu + \psi_\epsilon\frac{T+2}{2}\Big)\Bigg)\Bigg] \\
&= \frac{\psi_\epsilon\sigma_y^2\frac{\nu}{\phi}T(T+1)}{2det(\Sigma_T^{-1})}\Bigg[\Big(\psi_\kappa - \psi_\epsilon(T+1)T\frac{T+2}{6}\Big) + (T+1)\frac{T-1}{3}\Big(\psi_\mu + \psi_\epsilon\frac{T+2}{2}\Big)\Bigg] \\
&= \frac{\psi_\epsilon\sigma_y^2\frac{\nu}{\phi}T(T+1)}{2det(\Sigma_T^{-1})}\Bigg[\psi_\kappa - \psi_\epsilon(T+1)(T+2)\frac{1}{6} + \frac{(T+1)(T-1)}{3}\psi_\mu\Bigg]
\end{aligned}
$$

Proposition 4 assumes that σ_μ is sufficiently large, so we can drop the third term in the final line above (i.e., set $\psi_\mu = 0$). This derivative is therefore positive if and only if ϕ_κ is large enough (equivalently, σ_κ is small enough).

Proposition 4 also assumes that the agent's priors are unbiased: $\tilde{\mu} = \mu$ and $\tilde{\kappa} = \kappa$. In that case, absent rehearsal the agent's posteriors will also be unbiased (on average). Thus her posteriors will on average be too high or too low depending on whether ψ_κ is large enough or not. This proves Proposition 4.