Stories, Statistics, and Memory*

Thomas Graeber Christopher Roth Florian Zimmermann

April 7, 2023

Abstract

For most decisions, we encounter relevant information over the course of days, months or years. We consume such information in various forms, including collections of data shown in numbers – statistics – and anecdotes about individual instances – stories. This paper proposes that the information type – story versus statistic – shapes selective memory. In controlled experiments, we document a pronounced story-statistic gap in memory: the average impact of stories on beliefs fades by 33% over the course of a day, but by 73% for statistics. Consistent with a model of similarity and interference in memory, prompting contextual associations with statistics improves recall. A set of mechanism experiments reveals that lower similarity of stories to interfering information is a key force behind the story-statistic gap.

Keywords: Memory; Belief Formation; Stories; Narratives; Statistical Information.

*We thank seminar audiences at the Belief-Based Utility Workshop in Amsterdam, UC Berkeley, the Berlin Behavioral Economics Seminar, Bocconi, the Booth School of Business, the briq Belief Workshop, CERGE-EI Prague, the CESifo behavioral economics conference, Cologne, Cornell, Harvard, Hebrew University, Heidelberg, Innsbruck, MiddExLab, Sciences Po, the Spring School of Behavioral Economics, Stanford and UCLA Anderson for helpful comments. We thank Andrea Amelio, Simon Cordes, Paul Grass, Tsahi Halyo, Apoorv Kanoongo, Emir Kavukcu, Constantin Schesch, and Malin Siemers for excellent research assistance. Funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2126/1-390838866 is acknowledged. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Starting Grant: 948424 - MEMEB). The research described in this article was approved by the Institutional Review Board at the University of Cologne. Graeber: Harvard Business School, tgraeber@hbs.edu. Roth: University of Cologne, ECONtribute, CEPR, briq and MPI for Collective Goods Bonn, roth@wiso.uni-koeln.de. Zimmermann: University of Bonn and briq, florian.zimmermann@briq-institute.org.

1 Introduction

On many economic, political, and cultural issues, we accumulate a wealth of relevant information over the course of time. However, when making a decision, we may often remember only a portion of that information. This paper studies the nature of such imperfect recall and its influence on beliefs. Specifically, we examine the hypothesis that we are more likely to retrieve some *types* of information than others, which systematically shapes our beliefs. In practice, the type of news we are exposed to ranges from collections of information shown in numbers[1] – statistics – to contextualized descriptions of a single or few events – stories. If stories, even when unrepresentative of reality at large, are more easily recalled than statistics, people's beliefs may be disproportionately influenced by information they received through stories. This selective recall can give rise to misperceptions about reality as time passes.

To study the temporal evolution of beliefs in response to different types of information, we run a series of tightly controlled, pre-registered experiments. In our baseline study, a hypothetical product has received a number of reviews, each either positive or negative. We induce a uniform prior over the number of positive reviews. We study different types of information: statistical information about the number of positive reviews in a randomly drawn subsample of reviews (the *Statistic* condition), information about a randomly drawn single review alongside qualitative information describing the experience underlying the review (the *Story* condition), and no additional information. Participants are asked to guess whether a review that is randomly selected from all reviews is positive. In a within-subject design, each participant is presented with three independent product scenarios, and the type of additional information they receive is randomized for each scenario. To examine the role of memory, we elicit beliefs from participants twice: immediately (the *Immediate* condition) and again following a one-day delay (the *Delay* condition).[2] The temporal structure is crucial to our study design. There are numerous distinctions between stories and statistics that could result in different beliefs; however, any such differences *not* associated with memory are accounted for by the immediate belief update. Therefore, since no new information is received in the interim, any change in stated beliefs over time must, by design, be due to memory.[3]

In line with imperfect memory, average beliefs based on either type of information partially revert to the prior as time passes. Our main finding is a pronounced story-statistic gap in the evolution of beliefs: the effect of stories on beliefs decays much more

[1]This corresponds to the Oxford dictionary's definition of a statistic.

[2]We inform participants that the information provided in the initial survey will also be payoff-relevant for a follow-up survey one day later.

[3]We deliberately chose hypothetical products for our baseline study so that there is no relevant additional information that can be gathered outside of the experiment.

significantly than the effect of statistics. Pooling all statistics as well as all stories presented in our baseline study, we find that, on average, the magnitude of belief reversion from the immediate update towards the induced prior is more than twice as large for statistics (73%) as for stories (33%). In fact, we find that the average belief impact (the difference between a stated immediate or delayed belief and the induced prior) is larger for statistics than stories in *Immediate*, but smaller in *Delay*. This means that the relative magnitudes of belief impact *flip* over time.

To provide direct evidence on the role of selective memory, we use a free recall task in the follow-up survey. We find that participants are more accurate at recalling the correct type and valence of the information for the scenario in which they received a story than for the one in which they received a statistic.[4]

Our baseline findings are robust to (i) the extremity and valence of statistical information, (ii) the valence of story content and (iii) the number and type of "decoy" information presented in other scenarios. We also discuss and examine how differential engagement with information, processing time, prolonged deliberation, emotions and outside memories affect the interpretation of the story-statistic gap, showing that either of these factors plays a limited role in our study.

The existence of a story-statistic gap bears implications for how policymakers, managers, and individuals should communicate to persuade their audiences. Yet, any actionable implications depend on its underlying causes. One potential driving force behind the gap may be that stories – as opposed to statistics – are contextualized in nature. To guide our investigation of underlying memory mechanisms and the role of contextualization in particular, we rely on a simple formal framework of selective memory that builds on Bordalo et al. (2022, 2023). This framework focuses on the cue-dependent nature of episodic memory: experiences are organized through associations between memory traces that are activated by contextual cues. The model is based on the principles of similarity and interference in memory. The more similar a target memory trace is to the cue, the higher the chances of retrieving it. The more similar non-target memory traces are to the cue, the lower the chances of retrieving the target trace. Our baseline model makes several core predictions. First, stories are more likely to be successfully retrieved than statistics. Second, adding contextual features to a piece of information decreases the decay of its belief impact over time. Third, the driving force behind the story-statistic gap is *cross-similarity*: the similarity between target and decoy memory traces. Stories exhibit lower similarity to irrelevant, interfering memories than statistics, decreasing the likelihood of interference.

We develop experimental tests of the model's basic predictions as well as of more nuanced predictions targeting the specific levers of selective memory. First, adding –

[4]In follow-up experiments, we confirm these patterns using structured, incentivized recall tasks.

even arbitrary – contextual features to a statistic should boost recall, as this mitigates the interference statistics tend to suffer from. Consistent with this prediction, prompting respondents to imagine a typical review when provided with statistical information increases delayed belief impact, even though immediate updating remains unaffected by the prompt. Put differently, asking participants to add fictional contextual features to a statistic on their own slows the decay of its belief impact.[5] This implies that, if policymakers or individuals want to persistently persuade their audience with statistical information, they should enrich their messaging with contextual associations.[6]

Second, we aim to decompose different potential drivers of selective memory, testing the model's predicted role of cross-similarity as an organizing concept. To this end we conduct a series of mechanism experiments in which we systematically manipulate cross-similarity along various dimensions. We report the following findings, all of which are consistent with the predictions of our framework. As we move from one to three and then to six product scenarios, the magnitude of the story-statistic gap increases. Intuitively, a higher number of product scenarios increases cross-similarity, thereby creating a higher risk of memory interference. The contextualization of stories, however, makes them relatively distinctive and thus less susceptible to this type of interference, widening the gap. This implies that, particularly in information-rich environments with multiple interfering pieces of information, communication through stories is more effective than communication through statistics. Then, zeroing in on the retrieval of stories, we show that higher similarity between the contextual features presented to the same participant across different stories decreases the persistence of belief impact. Our results demonstrate that in environments where many similar but conflicting stories circulate, stories lose their edge over statistics as a communication device.

Finally, we move beyond the relationship between target and non-target memories – cross-similarity – to a second potentially relevant margin: similarity between the target memory and the cue. Intuitively, a story might be particularly likely to be retrieved when its content is semantically related to the cue, i.e. related in meaning. For example, the cue "restaurant" plausibly retrieves stories about restaurants more so than stories about less related topics such as a shopping trip. For statistics, on the other hand, there is generally less of a semantic relationship to cues. This suggests the possibility that the

[5]This further suggests that complementing statistical information with *consistent* qualitative information enhances recall. The effect of receiving a story jointly with contradictory statistical information, however, remains ambiguous as this potentially affects memories at the encoding stage already, e.g., by making the story particularly salient or less credible.

[6]One implication of our results from this experiment is that, in the absence of a prompt to encode additional contextual information, people perform similarly at recalling information about a binary variable as they do at recalling a statistic. This implies that our results are not specific to the recall of statistical information, but instead extend to the recall of simple facts or numbers that are devoid of contextual features.

similarity between the cue and the target memory shapes recall and thereby explains part of the story-statistic gap.[7] To investigate the relevance of cue-target similarity, we design an additional series of mechanism experiments that manipulate the cue-target similarity of both stories and statistics.[8] We document mixed or, at most, weak supportive evidence on delayed belief impact and recall for both stories and statistics.

Taken together, our mechanism experiments highlight the importance of cross-similarity and interference for the story-statistic gap, yet provide comparably little support for the importance of cue-target similarity in this setting. This suggests that to persistently shape beliefs, effective communication should focus on contextual features that boost the distinctiveness of their messages from competing pieces of information.

We conclude with a heuristic decomposition exercise that examines the relative importance of the two potential margins of selective memory: first, at the *extensive margin* of memory, people may fail to retrieve any relevant memories for a given cue; second, at the *intensive margin*, people may successfully recall relevant memory traces but only partially recover the original information content. To this end, we jointly examine our recall and belief data, which allows us to connect the magnitude of belief impact to different recall patterns. Strikingly, we document that conditional on correct recall of the valence and type of information, there is virtually no story-statistic gap. This suggests that the extensive margin of recall, i.e., retrieval failures of any relevant memories, plays a dominant role in shaping the story-statistic-gap in memory. By contrast, there is little information loss on the intensive margin, i.e., whenever memories are successfully retrieved. These analyses on the relative importance of different margins of forgetting provide a novel type of insight as well as an empirical foundation for modeling selective recall as primarily arising from retrieval failures rather than partial information loss.

Our work relates to a nascent literature on stories and narratives in economics (Shiller, 2017, 2020; Michalopoulos and Xue, 2021; Andre et al., 2022a,b; Kendall and Charles, 2022; Morag and Loewenstein, 2021). This literature mostly focuses on the persuasive effects of narratives in the moral or political domains (Bénabou et al., 2018; Eliaz and Spiegler, 2020; Bursztyn et al., 2022, 2023; Alesina et al., 2022). Relatedly, a literature in psychology and management studies the power of stories in influencing people (Fryer, 2003; Monarth, 2014; Bruner, 1987; McAdams, 2011). We add to these literatures by (i) comparing the effect of stories to that of statistics over time, and (ii) providing systematic, theory-guided evidence on mechanisms with a focus on the role of contextualization and

[7]While our baseline model, in following Bordalo et al. (2023), does not explicitly account for cue-target similarity, we develop and extension in Appendix H.1 that accommodates it.

[8]For stories, we vary whether the content of the story is semantically related to the scenario cue (e.g., “restaurant”). There is, by design, no such semantic relationship between scenario and statistic. For statistics, we vary whether the additional information is provided in the same display format (percentage or absolute counts) as the belief task, or not.

interference. Our evidence highlights one mechanism by which narratives are effective: they promote recall and thus more easily come to mind at the time of decision-making.

Our work also ties into a growing literature on the role of attention and memory in economics (Bordalo et al., 2020a, 2021; Gennaioli and Shleifer, 2010; Bordalo et al., 2020b). The model heavily builds on Bordalo et al. (2022, 2023), who provide theoretical frameworks in which agents form beliefs by retrieving experiences from memory based on similarity and interference. On the empirical side, Enke et al. (2022) study the role of associative memory for belief formation and show that it can give rise to overreaction to news. In contrast to our focus on the decay of belief impact over time, Enke et al. (2022) examine the extent to which immediate updating in response to new signals is influenced by the history of previous signals. Afrouzi et al. (forthcoming) experimentally highlight the role of working memory in forecasting experiments. A series of recent papers now also provide evidence on the role of associative recall in field settings, e.g., in finance (Charles, 2022; Jiang et al., 2022; Kwon and Tang, 2023) and a labor market setting (Conlon and Patel, 2022). Consistent with a core insight from this literature, our paper strongly suggests that people do not continuously and permanently update their beliefs every time they receive a piece of information, but instead (partly) construct them on-the-fly. Our paper differs from the previous literature in its focus on how different types of information, statistical versus stories, shape beliefs over time.[9]

More broadly, our work builds on an extensive psychology literature on memory. Schacter (2008), Kahana (2012) and Baddeley et al. (2020) provide overviews. Some previous work in psychology directly relates to the recall of stories, though with a particular focus on the role of scripts (Brewer and Treyens, 1981; Mandler, 1984; Schank and Abelson, 1977; Heath and Heath, 2007), emotions (Kensinger and Schacter, 2008) and mental imagery (Shepard and Cooper, 1986). Bower and Clark (1969) document that students' ability to remember a list of words strongly increases when instructed to create a coherent narrative that contains all of the words.[10] These papers differ from ours in a number of ways. First, they focus on studying the recall of word lists, but do not measure beliefs nor track their evolution over time. Second, they do not compare the dynamics of belief formation based on statistics versus stories. Finally, these experiments do not aim to tightly identify underlying cognitive mechanisms, such as the role of cross-similarity, a crucial ingredient for models of cue-dependent memory (Bordalo et al., 2022, 2023).

[9]Our evidence further speaks to the voluminous literature on biases in belief formation (Enke and Zimmermann, 2019; Graeber, 2022; Enke, 2020; Martínez-Marquina et al., 2019; Hartzmark et al., 2021; Ba et al., 2023).

[10]This relates to techniques for memory enhancement, which use visualizations of familiar spatial environments to improve the recall of information, commonly referred to as memory palaces or method of loci (Foer, 2012).

This paper proceeds as follows: Section 2 presents experiments which demonstrate the existence and robustness of a story-statistic gap in memory. In Section 3, we outline a simple theoretical framework that formalizes the mechanisms underlying the story-statistic gap in memory. Section 4 summarizes our evidence on mechanisms driving the story-statistics gap in memory. In Section 5, we provide a heuristic decomposition of the story-statistic gap and Section 6 discusses the implications of our findings.

2 The Story-Statistic Gap in Memory

2.1 Design

Our baseline study design is guided by the following objectives: (i) panel data on beliefs that allows us to study the evolution of beliefs over time without new information arriving in the meantime; (ii) a measure of immediate updating that captures any differences in the effects of stories and statistics that are not memory-related; (iii) a naturalistic setting in which information both in the form of statistics and stories is common; and (iv) an incentive-compatible belief elicitation. Table A.1 provides an overview of all experimental designs.

Task structure. There are three different hypothetical products. Each of these has received an overall number of reviews, with each review being either positive or negative. For every product, participants' task is to guess whether a randomly selected review is positive. To fix prior beliefs, we truthfully inform them that the actual number of positive reviews would be randomly drawn from a uniform distribution, independently for each product, inducing a flat prior. For each product, participants then receive either a piece of additional information or no additional information, and are subsequently asked to state their guess.

Main treatment variations. We implement two key sources of variation. First, within-subject and across product scenarios, we vary the type of additional information participants are exposed to. For each product, participants receive either statistical information (condition *Statistic*), or anecdotal information (condition *Story*), or no further information. Randomization is blocked such that across scenarios, each individual receives one story, one statistic and once no additional information. Moreover, the order of products is randomized and each individual receives one positive signal and one negative signal.[11] Second, we elicit beliefs twice: once immediately upon receiving the information (condition *Immediate*) and once one day later (condition *Delay*). Our main outcome of interest

[11]Appendix E provides details on the implementation of the randomization.

is respondents' beliefs about the likelihood that a randomly selected review is positive. The belief elicitation is incentivized using a binarized scoring rule (Hossain and Okui (2013)) with a prize of \$30.[12]

We conceptualize statistics as quantitative information about many reviews. In contrast, we define stories as quantitative information about a single review coupled with qualitative information about contextual features. Our design closely adheres to this taxonomy.

Statistical information is communicated as the number of positive reviews for a randomly selected subsample of the population. The fraction of positive reviews is randomly determined, creating rich variation in the extremity of statistics. Below is an example of how statistical information is communicated:

> *13 of the reviews were randomly selected. 4 of the 13 selected reviews are positive, the others are negative.*

A story provides information about whether a single randomly selected review is positive or negative, plus a qualitative description of that review. The description consists of 6-7 sentences recounting the experience underlying the review. We randomize the valence of the statements made in the text between-subjects. For our main analysis, we focus on stories where the valence of the statements made in the text matches the overall review rating.[13] Below is a shortened example of a story accompanying a negative review about a restaurant:[14]

> *One of the reviews was randomly selected. The selected review is negative. It was provided by Justin... The raw fish looked stale and the sushi rolls were falling apart on the plate... The service was poor: his waiter was rude, not attentive and the food was served after a long wait... As they left the restaurant, Justin was very annoyed and thought to himself "I definitely won't be back!"*

A notable feature of stories is that they cannot be accommodated in a Bayesian belief updating framework because the informational content of qualitative statements cannot be quantified in an entirely objective way. For instance, in the above example, the qualitative description of the food arguably allows participants to infer that other reviewers

[12]The precise payment formula is as follows: Probability of winning \$30 (in percent) = $100 - 1/100$ (estimate (in percent) - Truth)^2, where truth = 100 if the randomly selected review is positive, and 0 if not. The binarized scoring rule has been shown to be incentive-compatible, even in the presence of risk aversion. Danz et al. (2022) document that empirically, the binarized scoring rule can lead to systematic bias in reported beliefs. Notice that, even if such bias were present in our experiment, it would not compromise our identification which relies on the comparison of beliefs between *Immediate* and *Delay* for stories and statistics.

[13]In Section 2.3 we consider statements with mixed and neutral valence.

[14]Appendix D.1 reproduces all stories from the baseline experiment.

may have had similar experiences. Because we cannot determine the normatively optimal Bayesian inference from such qualitative information, we rely on our *Immediate* belief measurement to capture how informative participants *perceive* each story to be – including its qualitative statements. Note that this is sufficient for our purposes, as any change in belief impact over the course of one day is then necessarily related to memory.

Recall elicitation. To provide direct evidence on recall of the additional information about product reviews received in the baseline survey, we asked our respondents the following unincentivized open-ended survey question:[15]

> *Please tell us anything you remember about this product scenario. Include as much detail as you can. Most importantly, please describe things in the order they come to mind, i.e., the first thought first, then the next one etc.*

In additional studies that replicate our baseline findings, we include structured incentivized recall tasks instead of the open-ended question and show that they yield very similar results (see Section 2.3).

Hand-coding scheme. To analyze the unstructured text data, we design and implement a hand-coding scheme (see all details in Appendix F). The hand-coding scheme records whether respondents mention the valence and type of information they encountered, and whether they correctly remember these characteristics. It also captures additional features, such as whether (i) respondents in the *Story* condition mention qualitative features, (ii) whether respondents correctly recall the exact statistical information, and (iii) whether respondents recall the belief they stated in the baseline survey. To ensure high quality of the hand-coded data, we proceed as follows. First, we instruct three research assistants on the coding scheme and conduct a series of practice rounds with them. Second, each open text response is independently coded by two of the research assistants. Any potential conflicts are resolved by the third research assistant. We find that the inter-rater reliability is high: for correct recall of type and valence, we find agreement in 94% of the cases.

Procedures, payment and pre-registration. All experiments were conducted online and pre-registered on AsPredicted. See `https://aspredicted.org/e5mw7.pdf` for the pre-registration of the baseline experiment. The pre-registration includes the experimental design, hypotheses, analysis, sample sizes, and exclusion criteria.

Participants were informed in advance that the survey consisted of two parts, with one day in between. We also told participants that the information they receive would

[15] We randomized the order of the belief and recall elicitation in the follow-up survey.

be relevant for payoffs one day later. The average duration of the survey was about 9 minutes for the baseline survey, and 5 minutes for the follow-up survey. We implemented an attention check as well as extensive control questions to verify participants' understanding of the instructions. As pre-registered, participants could only participate in the survey if they passed the attention check and answered all control questions correctly. These control questions ensure high levels of understanding of the payoff incentives as well as the signals and prior distribution of draws.

For the baseline survey, participants received a completion payment of $1.55 and for the follow-up survey they received 90 cents. In addition, participants were truthfully informed that the computer would randomly select 10% of participants to receive a bonus payment that would be based on their responses.[16] To avoid hedging between similar questions in the two parts, one of the three products and one of the two parts for that product (immediate belief, delayed belief) were randomly selected to count for the bonus payment.

We collected data for this experiment on September 8 (baseline) and September 9 (follow-up) 2022. We recruited participants via Prolific, a survey provider commonly used in social science research (Peer et al., 2022). 1,500 respondents completed wave 1 of our experiment. Out of those, 1,437 met our inclusion criteria and were invited for the follow-up survey. 1,035 then completed the follow-up survey. After the pre-specified sample restrictions,[17] our final sample consists of 985 participants, corresponding to a completion rate of 69 percent.[18] The full set of instructions can be found on the following link: `https://raw.githubusercontent.com/cproth/papers/master/SSM_instructions.pdf`.

2.2 Baseline Results

As pre-registered, we start by analyzing stories with content that is consistent with the overall review rating. In Section 2.3, we examine the effect of mixed-valence and neutral stories. The top panel of Figure 1 and Table 1 show the average belief impact in *Immediate* and *Delay*, pooling the data across products and individuals. Belief impact is the signed distance between a stated belief and the prior (50%). For ease of exposition, we reverse-code the belief impact whenever the additional information implied a downward update, i.e., belief impact is signed in the direction of the rational update. Beliefs in *Immediate* serve as a benchmark that captures any difference in the effect of stories

[16] We paid out close to $10,000 in bonuses across all of our data collections.

[17] We pre-specified the exclusion of respondents who indicated having written down the information they received and those updating in the wrong direction in response to statistics.

[18] Given that the key treatment variation is within-person, the attrition rate is not a threat to the internal validity of our findings. For completeness, we report analyses on attrition rates in Appendix Table A.8.

and statistics that is not related to memory.

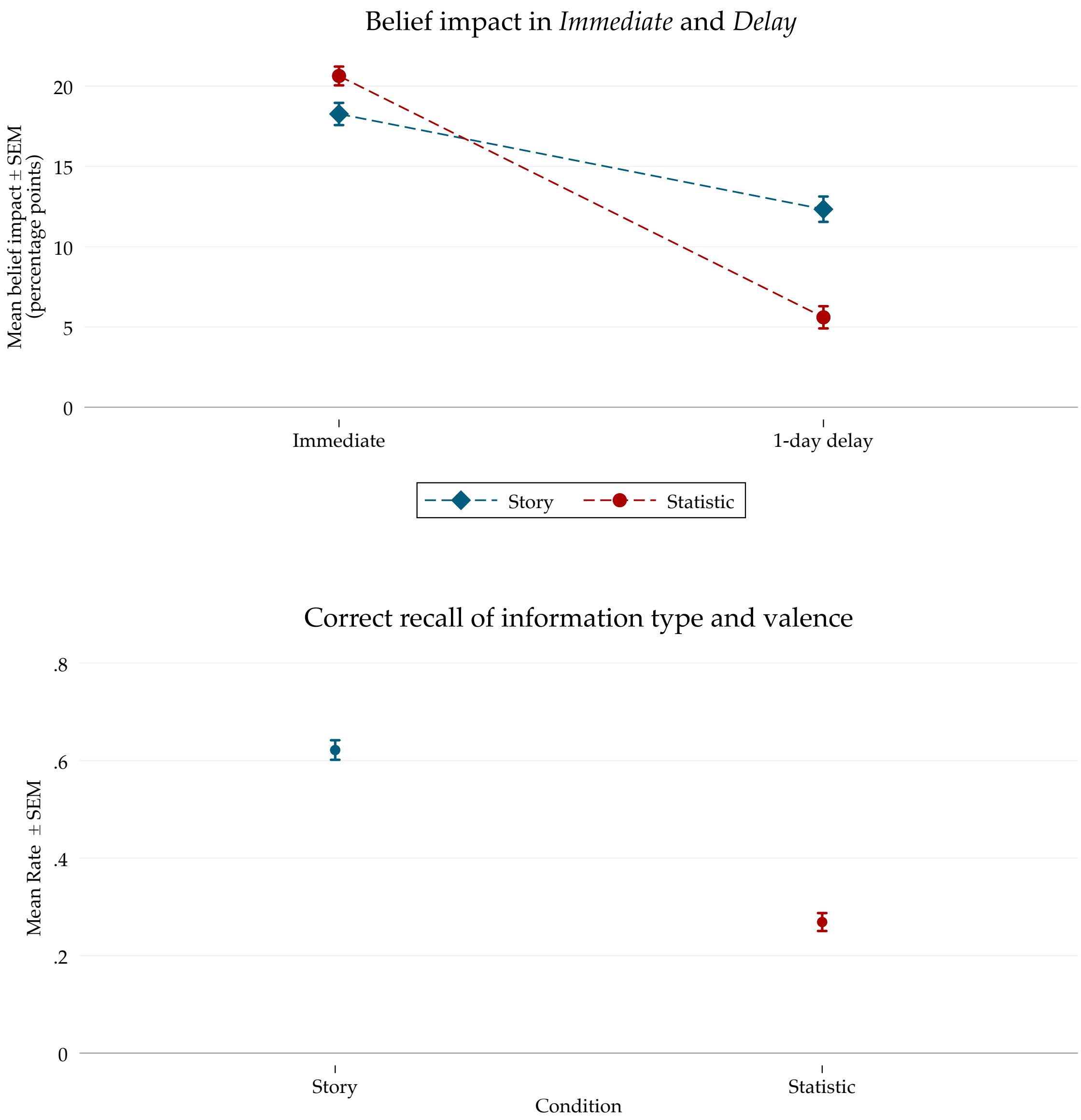


Figure 1: The story-statistic gap in the baseline experiment (984 respondents). The top panel displays belief impact in percentage points, separately for conditions *Immediate* and *Delay*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The bottom panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The red markers illustrate belief impact and recall for statistics, while the blue markers illustrate belief impact and recall for stories. Whiskers indicate one standard error of the mean.

The top panel of Figure 1 reveals that, in line with our hypothesis, the decay in belief impact over time is substantially lower for stories than statistics. This is confirmed by column (3) of Table 1. The difference-in-differences estimate of belief impact between the *Immediate* and *Delay* 1) is highly significant ($p < 0.01$). We next consider point estimates of the belief impact in *Immediate*. Average belief impact in *Immediate* is larger for

Statistic than for *Story*. On average, beliefs moved by 20.63 p.p. (s.e. 0.59) for *Statistic* and by 18.26 p.p. (s.e. 0.69) for *Story*.[19] For the *Delay* condition, by contrast, the top panel of Figure 1 reveals that mean belief impact after one day is substantially more pronounced for *Story* than for *Statistic*. On average, belief impact was 5.60 p.p. (s.e. 0.69) in *Statistic* and 12.33 p.p. (s.e. 0.79) in *Story*. This divergence in belief impact in *Delay* is significantly different from zero ($p < 0.01$). Appendix Figure A.5 underscores these patterns in the cumulative distribution functions of belief impact in *Immediate* and *Delay*, separately for stories and statistics.

To provide direct evidence on the role of memory in shaping the differential decay of belief impact for statistics and stories, we next turn to the hand-coded recall data. To study recall, we examine the fraction of respondents who correctly recall both the type and the valence of the information they were provided.

The bottom panel of Figure 1 shows that correct recall is significantly higher for stories than for statistics ($p < 0.01$). Average correct recall is 62.15 percent for stories and 26.90 percent for statistics. This suggests that the quantitative information in stories is more easily retrieved than the statistical information. Moreover, the richness of the open-ended data reveals several other striking features: (i) A large fraction of respondents (44.91%) mention qualitative features from the story without specifically being prompted to do so; (ii) a very small fraction of respondents (1.32%) correctly recall and indicate the statistic they received; and (iii) only a negligible fraction (4.23%) mention the posterior belief they stated in the baseline wave.

Our first main result can be summarized as follows:

Result 1. *We document a story-statistic gap in memory: following a delay of one day, stories have a stronger effect on beliefs than statistics, even though statistics have stronger immediate effects, on average. Recall accuracy is substantially higher for stories than for statistics.*

2.3 Robustness

In the following we examine the robustness of the story-statistic gap. First, we zero in on our results by examining the gap for different valence and extremity of statistical information. Second, we investigate the sensitivity of the finding to different experimental design choices: (i) the valence of the story content, and (ii) the amount and type of de-

[19] The immediate belief impact is close to the (average) Bayesian benchmark for both statistics (22.0 p.p.) and stories (18.7 p.p.). Note that for stories, we only consider the quantitative information contained in the review to compute the Bayesian benchmark, i.e., we do not factor in the effect of the qualitative information provided. Because of the role of qualitative information, we have no point predictions for beliefs in *Immediate* and therefore do not assign a substantive interpretation to the treatment difference in average point beliefs.

Table 1: The story-statistics gap in memory

	Dependent variable:				
	Belief Impact			Recall combined	
Sample:	Immediate (1)	Delay (2)	Pooled (3)	Consistent (4)	Story (5)
Story	-2.37* (1.23)	6.73*** (1.48)	-2.37** (1.01)	0.35*** (0.03)	
Delay			-15.0*** (0.90)		
Story × Delay			9.10*** (1.28)		
Neutral Story					-0.11*** (0.04)
Mixed Story					-0.026 (0.04)
Control Mean	20.63	5.60	20.63	0.27	0.62
Observations	1168	1168	2336	1168	984
R^2	0.54	0.52	0.43	0.65	0.01

Notes. This Table uses responses from the *Story* and *Statistic* condition. OLS estimates, standard errors clustered at the participant level in parentheses. *Delay* is an indicator taking value 1 for respondents in the follow-up survey, and value 0 for respondents in the baseline survey. *Story* takes value 1 for respondents who received a story for a given product, and zero otherwise. *Statistic* takes value 1 for respondents who received a statistic for a given product, and zero otherwise. Columns (1), (2) and (4) include respondents who received consistent stories. Column (3) pools *Immediate* and *Delay*. Column (5) includes observations who received stories. Columns (1) to (3) display results on belief impact. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. Column (4) and (5) display the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

coy information. We do not aim to disentangle different possible mechanisms underlying the gap here, but defer this discussion to Section 3.

Valence and extremity of statistics. Figure 2 illustrates the heterogeneity of delayed belief impact and correct recall of the type and valence of the information by the extremity of the immediate update. The figure showcases that there is only very muted heterogeneity in correct recall by the extremity of the immediate belief update. Therefore, for all levels of immediate updating, delayed belief impact and correct recall are substantially higher for stories than for statistics.

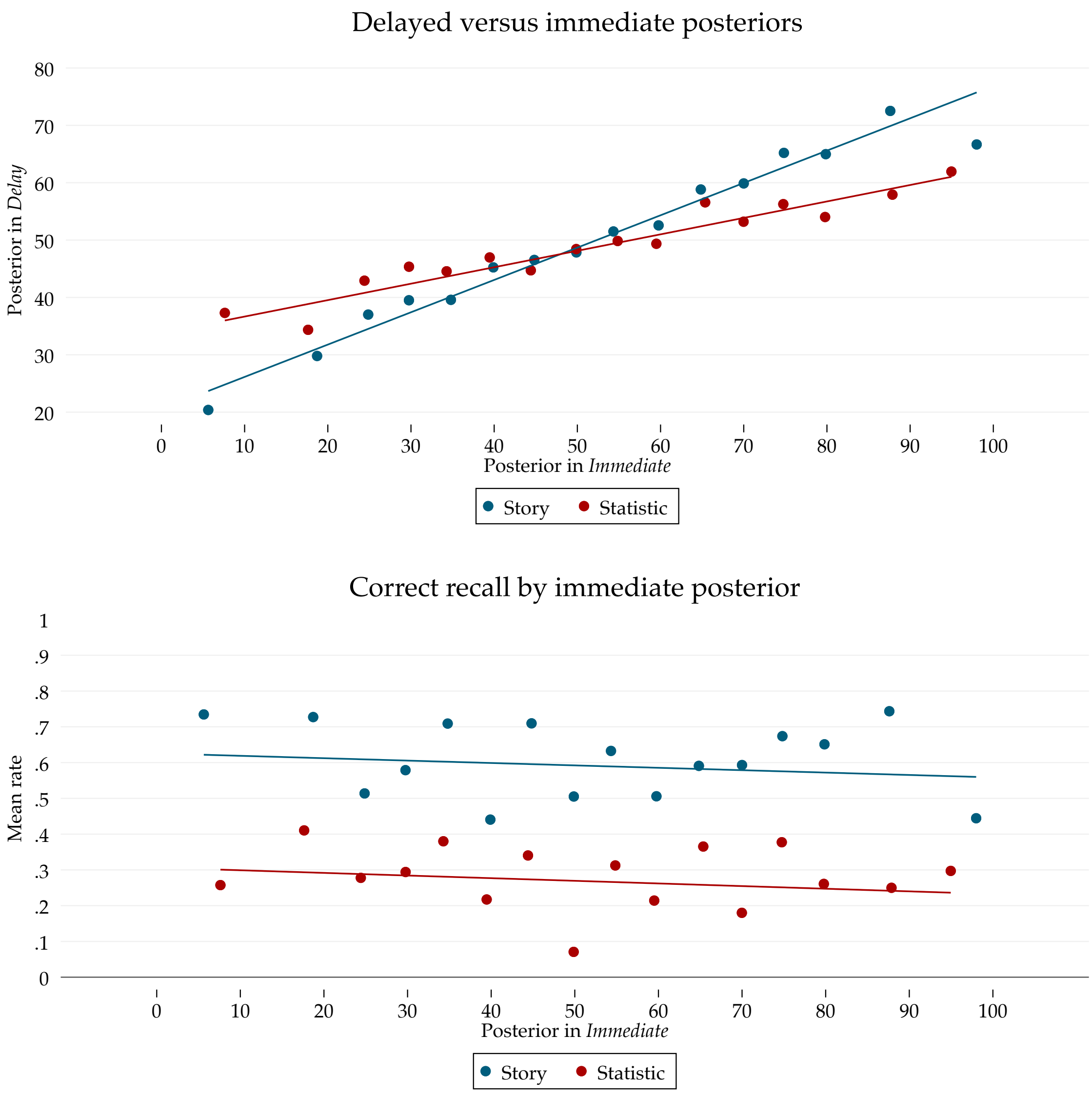


Figure 2: Heterogeneity by extremity of immediate update in the baseline experiment (984 respondents). The top panel displays binned scatterplots regressing beliefs in *Delay* (y-axis) on beliefs in *Immediate*, separately for conditions *Story* and *Statistic*. The bottom panel displays binned scatterplots regressing correct recall of the type and valence of information they received in the baseline survey in *Delay* (y-axis) on beliefs in *Immediate*, separately for conditions *Story* and *Statistic*. The red dots and line illustrate beliefs and recall for statistics, while the blue dots and line illustrate beliefs and recall for stories.

Valence of story content. To examine the importance of the valence of the story content, our baseline experiment cross-randomized whether the contextual information in the stories was (i) consistently positive or negative in line with the review rating, (ii) of mixed valence, or (iii) neutral (see Appendix D for all stories). Figure 3 and Column 5 of Table 1 show that the valence of story content has minor but significant effects.[20] Average correct recall is 62.15 percent in the consistent story condition compared to 59.58

[20]Since we expected the valence manipulation to have potentially strong effects on immediate updating, we pre-registered using recall performance as our main outcome measure.

and 51.20 percent in the mixed and neutral stories treatments, respectively. These levels of recall are substantially higher compared to 26.90 percent for statistics, indicating that the story-statistic gap is robust to variations in the valence of the story content. The patterns for belief impact are consistent with the recall evidence. While belief impact in *Immediate* does indeed depend on the valence of the contextual information, these differences are strongly attenuated in *Delay*.

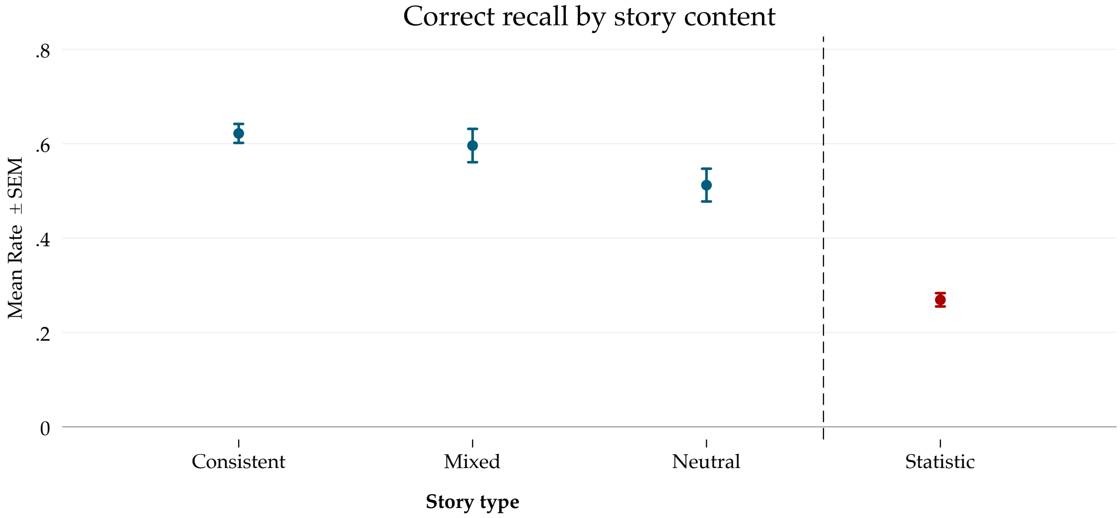


Figure 3: Correct recall of type and valence by story type in the baseline experiment (984 respondents). The figure shows the fraction of correct recall of the type and valence of information received in the baseline survey in *Delay* for respondents in the *Story* condition (blue) and *Statistic* condition (red). Consistent refers to stories with contextual features whose valence was fully consistent with the valence of the review. Mixed refers to stories with contextual features whose valence is mixed. Neutral refers to stories with contextual features whose valence is neutral. Whiskers indicate one standard error of the mean.

Heterogeneity by positive versus negative reviews. Next, we investigate any potential heterogeneity in belief impact and correct recall between positive and negative reviews. Figure 2 illustrates that there is a pronounced story-statistic gap for both positive and negative reviews. In fact, we find no difference in recall performance by whether the reviews are positive or negative (*Story*: $p = 0.328$, *Statistic*: $p = 0.991$). Moreover, there is no heterogeneity in the evolution of belief impact for *Story* by the valence of the quantitative information ($p = 0.860$) . We observe, however, that positive statistics affect beliefs more persistently than negative statistics ($p < 0.001$), consistent with a literature on the hedonic benefits of recalling positive information (Zimmermann, 2020).

Features of decoy information. In our baseline design, each respondent received one statistic and one story across the three scenarios. As a result, any given story was accompanied by a statistic as the “decoy” information in the respective other scenario, whereas any given statistic was accompanied by a story in the decoy scenario. To examine how sensitive the story-statistic gap is to the nature of decoy information, we systematically manipulated the number, type and valence of decoy pieces of information in a separate experiment (see Appendix Section A.1 for details). In a between-subject design, respondents either received two statistics, two stories or twice no information as decoys. As Appendix Section A.1 summarizes, we document a robust story-statistic gap across all conditions. In fact, the estimated story-statistic gap has a similar magnitude, irrespective of the number, type and valence of decoy information, holding fixed the total number of scenarios. This suggests that the story-statistic gap is robust to the basic structure of decoy information in our 3-scenario baseline setting.

2.4 Interpreting the Story-Statistic Gap

There are many differences between stories and statistics, and our baseline evidence leaves open which of these account for the story-statistic gap. In the following, we provide a brief overview of several key differences and how they relate to the story-statistic gap as well as to our model-guided examination of mechanisms in the subsequent parts of this paper. Rather than an in-depth discussion of specific explanations, this section aims at bringing up an array of plausible considerations. Some of these turn out to be of low relevance given our design or have little prominence in our data, which partly motivates the focus of our formal framework and mechanism experiments.

Engagement with additional information and processing time. Differences in the processing time of stories and statistics, which may be indicative of the encoding strength, are a plausible mechanism underlying the story-statistic gap. We find that respondents spend somewhat more time processing stories (median of 42 seconds) than statistics (median of 32 seconds). Appendix Table A.3 examines heterogeneity in belief impact and recall by the time spent processing the information. Correlationally, we find small and insignificant heterogeneity in differential belief impact based on initial processing time. Our mechanism experiments hold the processing time of the target scenario constant, as they only vary cross-similarity, the similarity between a target scenario and other decoy scenarios.

Deliberation. It is conceivable that it may take some time for information to "sink in", and that the beliefs in *Immediate* are elicited before the information has been fully processed. In that case, using the immediate belief as a benchmark may not adequately capture the maximal belief update. This would compromise inference about the story-statistic gap if such deliberation occurs to different degrees for stories and statistics.

Given the nature and content of the scenarios and information, we see little reason for such prolonged deliberation to play much of a role. Empirically, we find that all of our key results on delayed versus immediate belief impact are supported by evidence on recall accuracy, which are not dependent on the immediate updating benchmark.

Emotions and vividness. Research in psychology has established a connection between emotions and memory (e.g., Kensinger and Schacter, 2008). Intuitively, stories are more vivid than statistics and tend to evoke emotions. First, our evidence on the valence of story content partly speaks directly to this mechanism. It suggests that while stories with more consistent qualitative features are recalled at somewhat higher rates than stories with mixed and neutral contextual features, these differences are relatively small, especially compared to the large differences in recall between stories and statistics (see Column 5 of Table 1). Second, while emotions plausibly play a role in driving the baseline story-statistic gap, the bulk of our mechanism evidence focuses on the features of cue-dependent memory, which allows us to hold emotions fixed. In fact, our experimental manipulations on similarity and interference discussed below provide some boundary conditions for the existence of a story-statistic gap that are unrelated to the role of emotions.

Outside memories and sample. Respondents do not enter the experiment with a blank slate but bring in an outside database of memories. This existing database will contain both stories and statistics to some extent, potentially affecting memory of different types of information. Moreover, in the online samples we use, stories and statistics may be differentially surprising or typical given other studies they participate in. We fully embrace these issues and point out that they would also affect the response to stories versus statistics outside of our experiment. Moreover, in Section 4 we will examine mechanisms of cue-dependent memory that operate independently of baseline differences in the background memory database.

Mental representation of stories. A fascinating question is how the story-statistic gap relates to how memories are encoded and retrieved in the brain (Shepard and Cooper, 1986). We explicitly do not claim that any of our evidence directly speaks to cognitive representations of memory in the brain, i.e., we do not claim that stories are an elementary format of information storage in the brain. Instead, while we do not see our evidence as indicating that the brain encodes stories directly, we believe it suggests that stories *facilitate* efficient storage and retrieval of information. Relative to statistics, the story format is plausibly a facilitator of brain processes related to memory, such as mental imagery.

3 Outline of Conceptual Framework

In the following, we outline a model of cue-dependent memory that adapts Bordalo et al. (2022) to accommodate stories and statistics. The model allows us to derive formal predictions for the differential recall of stories and statistics and provides a guiding structure for the empirical analysis of underlying mechanisms. Details and formal derivations are relegated to Appendices G and H.

3.1 Setup

We model a decision maker's (DM) beliefs in two periods. A product p has N total reviews, K of which are positive. The DM is informed that K is drawn uniformly from $\{0, \ldots, N\}$. In the first period, the DM may receive the additional information that there are k positive reviews in a randomly selected subset of $n \leq N$ reviews. The additional information is either a statistic about multiple reviews ($n > 1$) or a story about a single review ($n = 1$). The DM then states a belief about the probability $\pi := K/N$ that a randomly drawn review from the population of N reviews is positive. In the second period, the decision maker again forms their belief about this probability based on the information they retrieve.

Our key object of interest is belief formation in the second period. For simplicity, we assume Bayesian belief formation given the information that is recalled, so that any deviation from the benchmark of optimal belief updating originates from selective recall, rather than non-Bayesian belief formation. We study updating from the quantitative information about the reviews mentioned in the story or statistic, and abstract from inference from the anecdotal information contained in stories. Put differently, in the baseline model the DM treats the qualitative content of the story as uninformative: a story is equivalent to a statistic with $n = 1$.[21]

3.2 The Structure of Episodic Memory

Canonical models of memory structure assume that personal experiences that can be explicitly described are stored in episodic memory (Kahana, 2012). We follow this assumption and formalize how stories and statistics are stored in and retrieved from episodic memory.

[21] The mechanics of the memory model would not be affected by allowing for updating from the qualitative information. Selective recall operates by distorting the update that would be formed without memory constraints – and the model shows why these distortions may be different for stories and statistics. The update that is formed in the absence of memory constraints, however, is not constrained in any way, i.e. it does not need to be Bayesian and based on the quantitative information alone, as is assumed here for tractability.

Memory traces. Episodic memory traces are encoded as vectors of $F \geq 1$ features with values in $V_1, \dots, V_F$. All sets of possible values V_f contain the null value 0 that indicates the absence of said feature. A trace encoding a statistic about a bicycle contains the product type and the value of the statistic, while a trace encoding a story about a restaurant contains the product type, the value of the review and anecdotal features:

$$\begin{aligned} m_{bicycle} =& \textit{(bicycle, 3 out of 7 reviews were positive, 0, 0, \dots)} \\ m_{restaurant} =& \textit{(restaurant, 1 out of 1 review is positive, Justin and his friends had} \\ & \textit{a wonderful experience, they ordered the sushi taster, \dots)} \end{aligned}$$

Our baseline model follows Bordalo et al. (2022) in assuming that each product scenario creates a single memory trace. As detailed below, we model the DM's retrieval process when presented with an external memory *cue*, which here is the prompt to form a belief about a given scenario. We call $C_p = \{m_p\}$ the *cued set*, which consists of the single memory trace that encodes the experience associated with the prompted product scenario p.

Appendix H presents two generalizations. In Section H.2, we present a more general model in which we allow each episode to leave multiple target traces. In Section H.1, we consider an extension that explicitly accounts for the similarity between the external cue and the target trace, i.e., the idea that given the external cue *bicycle*, it is more likely that the target memory is retrieved if its content involves a bicycle than if it involves nothing related to bicycles.

We call M the set of all memories and E the set of memories created during the experiment. $\overline{E} := M \setminus E$ are memories from outside the experiment, and $C_{-p} := E \setminus C_p$ memories of non-cued products, so that C_p, C_{-p} and $\overline{E}$ form a partition of M.

The extensive and intensive margins of memory loss. Given the structure of episodic memory, there are two margins of potential memory loss. First, the target memory trace might not be retrieved: we call this the *extensive margin* of forgetting. Models of episodic memory, including the one by Bordalo et al. (2022) that we adapt here, operate on this extensive margin associated with retrieval failures.

Second, a target memory trace may be successfully retrieved, but might not contain all the information associated with the original experience. We refer to this as the *intensive margin* of forgetting. In other terms, the storage of information in episodic memory traces may be subject to *information loss*. For example, episodic memories may only capture whether a statistic was overall positive or negative, or may encode a coarse version of the true weight and strength of the signal.

As our model operates on the extensive margin of retrieval failures, its main predictions are, in general, unaffected by the degree of information loss in episodic memory traces. The baseline model assumes that the full wealth of information is encoded in episodic memory. Appendix H.3 allows for partial information loss in episodic memory, and shows that our predictions are robust. Empirically, our combination of recall and belief data will allow us to disentangle the extensive and intensive margins. In Section 5, we conduct an empirical analysis to disentangle which aspect of memory loss has a greater impact on the story-statistic gap. Our findings suggest that the gap is indeed almost exclusively driven by the extensive margin of retrieval failures.

3.3 Recall and Similarity

We model episodic memory as cued recall. The prompt to form a belief about a given scenario constitutes the external retrieval cue. Within the baseline model, the scenario cue directly corresponds to the cued set C_p that contains the single target trace. We model the retrieval of memory traces given the cued set.

Memory retrieval is stochastic: the cue triggers the retrieval of a random trace, with a probability distribution that depends on similarities between traces. More precisely, we define a similarity measure over traces $S(m_1, m_2) : M \times M \to [0, 1]$, and require that it is symmetric, increasing in the number of shared features, decreasing in the number of non-shared features, and maximal when $m_1 = m_2$. We define the average similarity of two sets of traces as $S(M_1, M_2) := \frac{1}{|M_1|}\frac{1}{|M_2|}\sum_{m_1 \in M_1}\sum_{m_2 \in M_2} S(m_1, m_2)$. In the simplest version, the decision-maker only samples once, and, following Bordalo et al. (2022), realizes whether they retrieve the target trace or not.[22]

We then assume, in line with Bordalo et al. (2022), that the probability of recalling the target trace m_p when being cued with the scenario of product p is:

$$r(m_p) := \frac{1}{\sum_{m \in M} S(m, m_p)} \tag{1}$$

The probability of recall is driven by *interference*, i.e. the similarity between the cued set and all other memories: the higher this similarity, the likelier it becomes that other, irrelevant memories are accidentally retrieved. Using the partition of memories into a cued, non-cued and out-of-experiment set, we can rewrite (1) as:

$$r(m_p) = \frac{1}{1 + |C_{-p}|\; S(C_{-p}, m_p) + |\overline{E}|\; S(\overline{E}, m_p)} \tag{2}$$

Recall in a given scenario therefore depends on the similarity of that scenario's memory trace with the memory traces of other scenarios, and with memory traces formed outside

[22]If an irrelevant trace is retrieved, the decision-maker discards it and does not update their belief.

the experiment.

3.4 Predictions

In what follows, we outline the core predictions of the model. All proofs are provided in Appendix G.3.

Prediction 1. *There is a story-statistic gap in memory: the effect of a statistic on beliefs decays more rapidly than the effect of a story.*

Retrieval failure is more likely for statistics than for stories because stories tend to be distinct, whereas statistics are generic and hence more similar to non-cued and out-of-experiment traces. Within the model, this difference in the level of distinctiveness occurs because story traces contain more features than statistic traces, so that they exhibit lower similarity S with the interfering traces.

Prediction 2. *Adding contextual features to a piece of information decreases belief decay.*

Contextual features are additional entries in episodic memory traces. This decreases cross-similarity to non-target traces and thereby improves the recall likelihood of the target trace, decreasing the decay of belief impact over time. This prediction follows from the assumption that similarity is driven by the number of shared features.

Prediction 3. *Adding decoy scenarios decreases recall.*

Decoy scenarios increase interference by increasing the number of non-cued memories, thereby weakening recall. This effect applies to both stories and statistics.

3.5 Extensions

Our baseline model is a parsimonious adaptation of Bordalo et al. (2022) to stories and statistics. We explore alternative assumptions as extensions that we briefly outline below.

3.5.1 Cue-target similarity

So far, the target trace itself served as the cue, evoked by the external prompt. However, the recall likelihood might also depend on the strength of the link between the externally provided memory cue and the target memory itself. We formalize this by also representing the cue as a trace, where an additional last entry now encodes the question being asked, e.g.,

$$c_{bicycle} = (\textit{bicycle}, \ldots, \textit{What is the probability of a random review being positive?})$$

Recall then depends on the similarity between the external cue c_p and the target memory m_p. We formalize this similarity relationship in a way that leverages the concept of *semantic memory* (Kahana, 2012), which refers to general knowledge about concepts and categorical relationships between concepts. For example, semantic memory might relate *restaurant* to *food* and *friends*, which are more likely to be present in a trace related to a story than a statistic. Similarity of the cue and the target is then driven by the higher-level features they relate to.

Appendix H.1 formalizes this extension, relying only on general intuitions about cue-target similarity. Any qualitative story features that are inherently related to the cued scenario provide an additional rationalization of the story-statistic gap. We conduct an empirical analysis of this potential memory channel in Section 4.3.

3.5.2 Multiple target traces

In the baseline model, we conceptualize each experience as creating a single trace in the memory database. This implies that the *self-similarity* of the cued set C_p, i.e., the similarity of different memory traces within the cued set, is always 1, shutting down its role in differential recall. Appendix H.2 presents a model in which a single episode can leave multiple memory traces. Our main predictions are preserved in this model variant. The additional force of self-similarity of stories and statistics reflects two opposing forces: although statistics have shorter memory traces which are therefore more similar, stories are more impactful so that they leave more traces. We argue that the net effect favors stories, so that self-similarity also contributes to explaining the story-statistic gap.

4 Mechanisms

The existence of a story-statistic gap is relevant for policymakers, managers and individuals who wish to communicate information in ways that persistently change the beliefs of their audience. However, as noted above, stories and statistics differ along various dimensions (see Section 2.4). To determine the actionable implications of the gap, a better understanding of its specific *sources* is necessary. For example, if contextual associations per se enhance recall, then policymakers aiming to disseminate statistics should supplement them with contextual features that create rich associations. If cross-similarity is an important mechanism, then policymakers should focus on creating distinct messages that are as dissimilar as possible from competing pieces of information. Guided by the predictions spelled out in Section 3, we proceed with our analysis of mechanisms in three steps. First, in Section 4.1, we test Prediction 2 on the power of adding contextual features. Second, we delve into the features of cross-similarity and interference, motivated by Prediction 3, in Section 4.2. Finally, we extend our investigation of the key channels of cue-dependent memory by studying cue-target similarity in Section 4.3.

4.1 The Role of Contextual Associations

Design. To causally examine the role of adding contextual features *while holding the amount of information content provided constant*, we prompt respondents to imagine a typical review for the statistic or for a single review they learn about. This treatment does not provide any objective information, qualitative or quantitative, allowing us to identify the distinct effect of associating obviously fictional contextual features with a piece of information in memory.

We implement four conditions. In *Baseline*, we replicate our main design. The *StatisticPrompt* condition is identical to *Baseline*, except that respondents that receive the statistic are prompted "to imagine how a typical review based on the provided information would look like."

To examine the role of associations for single reviews that do not contain any qualitative contextual features, we design two additional treatments. The *NoStory* condition is identical to *Baseline*, except that instead of a story, respondents receive information about a single review without any contextual information. The *NoStoryPrompt* condition is identical to *NoStory* except that respondents that received information about a single review are asked to imagine what the review might look like, similar to *StatisticPrompt*. The rationale behind these two conditions is to examine what happens when the story provided in the *Story* condition of our main experiment is stripped of its actual content and then replaced by an endogenously generated one.

To complement our open-ended measure of recall from the baseline experiment with an incentivized measure, we use a structured recall task.[23] We ask respondents to indicate whether they (i) received information about a single review, including some additional anecdotal details about the reviewer and their experience with the product, (ii) multiple reviews, (iii) no information or (iv) don't know.[24] Unless respondents indicate that they did not receive any information about this product, we additionally ask them to indicate whether the information they received was positive or negative.[25] Respondents are told that if they correctly recall the information they received, they will receive an additional bonus of $5. To circumvent hedging motives, either beliefs or recall were randomly selected for payment, and one question was randomly chosen to determine the

[23]As before, we randomized the order of our measures of recall and the belief elicitation in the follow-up survey.

[24]Respondents are told that if they choose "don't know", one of the other options will be randomly chosen to determine their payoff.

[25]We tailored the question wording for respondents according to whether they indicated having received a single review, multiple reviews or "don't know".

bonus.

Sample and pre-registration. 1,500 respondents completed wave 1 of our experiment, with 1,442 qualifying for wave 2. Of those, 703 respondents actually completed wave 2. 666 of the final set of respondents satisfied our inclusion criteria, corresponding to a completion rate of 46 percent.[26] The pre-registration for this experiment is available on AsPredicted, see `https://aspredicted.org/v9gk7.pdf`.

Prediction. The decay of belief impact and forgetting is lower in the *Prompt* conditions than in the *No Prompt* conditions.

Results. We start by examining whether the prompt intervention was effective in actually inducing participants to imagine reviews and to write them down. The median (mean) number of words participants wrote to describe an imaginary typical review was 22 (23). The text responses indicate that the vast majority of participants made a significant effort to describe a review, such as in the following excerpt from a response in the *NoStoryPrompt* condition about a negative videogame review:

> *The gameplay was sub-par and glitched randomly. The graphics compared the trailer to the actual gameplay were very different giving the impression that the gameplay will have 3D style graphics while in reality, it had very old-school-style graphics [...].*

For ease of exposition, Figure 4 pools respondents in *NoStoryPrompt* and *StatisticPrompt*, as well as the *NoStory* and *Baseline* conditions.[27] The top panel of Figure 4 shows results on belief impact, while the bottom panel displays results on recall.

Starting with belief impact, we find that, reassuringly, beliefs in *Immediate* are not meaningfully different across the *Prompt* and the *NoPrompt* conditions. Yet, in *Delay*, average belief impact for respondents in the *Prompt* conditions is 7.30 p.p. (s.e. 0.70) compared to only 5.40 p.p. (s.e. 0.68) in *NoPrompt*. This treatment difference in *Delay* is statistically significant ($p < 0.01$). Column (1) of Table A.2 reveals that the difference-in-differences (difference in slopes) is also statistically significant ($p < 0.05$).

These patterns for *Delay* beliefs are underscored by results on recall. The bottom panel of Figure 4 shows that recall accuracy is 43.14 percent for respondents in *Prompt*, compared to only 32.69 percent in the conditions without prompt. Table A.2 reveals that

[26]The completion rate to the follow-up survey does not differ significantly across treatment groups ($p = 0.90$). The somewhat lower completion rate compared to the baseline experiment can be explained by the fact that part of the experiment took place on the weekend.

[27]Table A.2 shows results separately for all 4 conditions and confirms that the disaggregated results are similar.

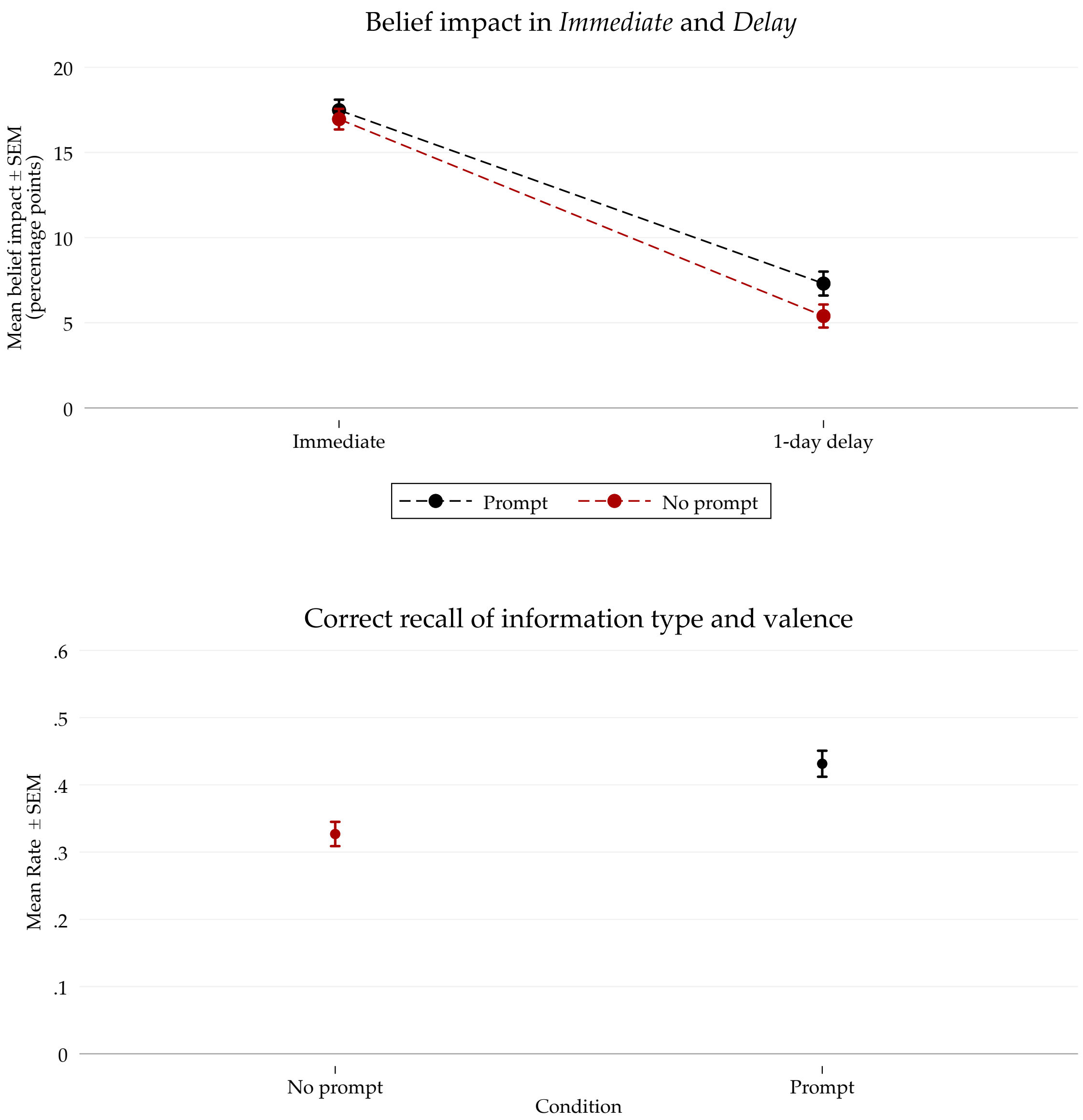


Figure 4: Belief impact and recall in Mechanism Experiment 1 (666 respondents). The top panel displays belief impact in percentage points, separately for conditions *Immediate* and *Delay*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The bottom panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The red markers illustrate belief impact and recall for *No Prompt*, while the black markers illustrate belief impact and recall for *Prompt*. Whiskers indicate one standard error of the mean.

these differences are highly statistically significant when comparing respondents in the *StatisticPrompt* and *Baseline* conditions, as well as when comparing respondents in the *NoStoryPrompt* and *NoStory* conditions.

Our second main result is given as follows:

Result 2. *The addition of contextual features causes a more pronounced belief impact in* *delay and facilitates more accurate recall of information.*

Recall of binary quantitative information. One key result that emerges from this experiment is that in the absence of a prompt to encode additional contextual information, people perform similarly at recalling information about a binary variable as they do at recalling a statistic. Specifically, Table A.2 reveals that correct recall among respondents in the *NoStoryPrompt* condition is 16 percent and thus, if anything, lower compared to correct recall of statistical information in the *Baseline* condition (22 percent). This implies that our results are not specific to the recall of statistical information, but instead extend to the recall of simple facts or numbers that are devoid of contextual features.

Complementing statistics with stories. This experiment was deliberately designed with the aim of keeping the amount of objectively provided information constant. An alternative approach is to provide *both* a story and a statistic within the same scenario. The interaction between these two types of pieces of information will depend on their interplay both at the encoding and retrieval stages. When the story and the statistic provide a consistent message, stories may boost the retrieval of statistical information, as suggested by the evidence presented in this section. When the story and the statistic provide conflicting messages, various mechanisms may play a role. The salient contrast may attract attention, enhance overall encoding and boost recall of the statistic nonetheless, in the sense that participants distinctly remember receiving contradicting evidence in a scenario. Alternatively, the conflict might lead to memory competition, resulting confusion and therefore dampen correct recall of the statistic, as participants may sometimes only recall the story and (incorrectly) infer a compatible statistic. Such conflicts and their effect on encoding and recall provide a fruitful area for future research, but are outside of the scope of the present empirical exercise and guiding framework.

4.2 Cross-similarity

We present three experiments that jointly aim to examine the importance of cross-similarity in a comprehensive fashion. We investigate the role of (i) the number of product scenarios presented within the experiment, (ii) the similarity of different pieces of information, and (iii) the similarity of product cues.

4.2.1 The Number of Product Scenarios

A key prediction of our model is that increases in cross-similarity through a higher number of product scenarios tend to more strongly impede the recall of statistics than stories. The rationale for more muted effects of this variation in cross-similarity on stories is that the richness of anecdotal content makes stories distinct and hence less similar to additional product scenarios.

Design. The design broadly follows the structure of the main experiment. The key difference is that we vary, between-subjects, whether there are one, three or six product scenarios. In the *1-product* treatment, there is a single scenario and participants only received one piece of information, either a story or a statistic. Identical to the baseline experiment, participants in the *3-product* treatment see three scenarios and receive two pieces of information, one story, one statistic and once no information. In the *6-product* treatment, participants see six scenarios overall and also receive two pieces of information (one story and one statistic), as well as four times no information. This means that the comparison between the *3-product* and *6-product* design allows us to cleanly study the effects of the number of product scenarios, while holding the total pieces of information constant.[28]

To keep incentives exactly constant between the different conditions, participants in all treatments complete a total of six payoff-relevant tasks in both *Immediate* and *Delay*: the additional filler tasks are incentivized dot estimation tasks. Respondents in the 1-product treatment arm complete 5 dot estimation tasks, while respondents in the 3-product treatment arm complete 3 dot estimation tasks, and respondents in the 6-product treatment only face product-related tasks. The experimental instructions for the dot estimation task, in which participants have to guess the number of dots displayed in a box for a short period of time, can be found on the following link: `https://raw.githubusercontent.com/cproth/papers/master/SSM_instructions.pdf`.

Sample and pre-registration. We recruited 1500 respondents. 1404 respondents qualified for the follow-up survey. After the pre-specified sample restrictions, our final sample consists of 1018 respondents, corresponding to a completion rate of 73 percent.[29]

The pre-registration for this experiment can be found on AsPredicted, see `https://aspredicted.org/as7i7.pdf`.

Prediction. The magnitude of the story-statistic gap both in the decay of belief impact as well as recall accuracy increases with the number of scenarios.

Results. Figure 5 and Table A.5 illustrate changes in belief impact between *Immediate* and *Delay* as well as recall for stories and statistics across the different number of product

[28]The comparison between the *1-product* and *3-product* condition jointly identifies the effects of increasing the total number of products and increasing the pieces of information.

[29]The completion rate to the follow-up survey does not differ significantly across treatment groups ($p = 0.37$).

scenarios. The top panel depicts the change in belief impact between *Immediate* and *Delay* across the three treatment arms, separately for stories and statistics. We find that, overall, the change in belief impact tends to become more pronounced as we increase the number of product scenarios. This effect is relatively small for stories. In fact, the *6-product* treatment does not lead to a more pronounced decay of belief impact than the *3-product* and *1-product* versions. At the same time, the effect of more scenarios on the decay of belief impact is quantitatively large for statistics. As a consequence, and in line with our model, the story-statistic gap widens with the number of product scenarios.[30]

This pattern is strongly supported by the recall data, see the bottom panel of Figure 5. Recall accuracy of statistics drastically decreases as we move from 1 to 3 to 6 scenarios, while recall accuracy of stories remains comparably stable.

Viewed through the lens of the model, these findings suggest that the differential effect of the number of product scenarios on stories versus statistics arises from differences in cross-similarity rather than memory load. The rationale for muted effects of cross-similarity on stories is that the richness of anecdotal content makes stories distinct and hence less similar to other product scenarios.

4.2.2 The Similarity of Story Content

One key difference between stories and statistics is that statistics are intrinsically more similar to one another than stories: intuitively, the numbers 73% and 82%, for example, are generally less distinctive than two arbitrary stories about different products. This higher cross-similarity in turn increases interference. To study the role of cross-similarity, we conduct experiments with stories only. These experiments directly manipulate the similarity between one target story and a set of decoy stories.

Design. We designed two treatments to study the role of story similarity. The incentives and basic setting are identical to our main experiment. Participants in both conditions learn about three products: a cafe, a restaurant, and a bar. Unlike in our main experiment, respondents receive a story in each of the three scenarios. The target story in both conditions that our analysis focuses on is a positive review about the bar. The stories about the restaurant and the cafe are decoy stories and both featured a negative review. In the *Baseline* condition, the three stories are distinct and specific to each cue. The bar story describes the interior of the bar, the restaurant story focuses on food quality, while the cafe story is concerned with the service quality. In the *Story Similarity* condition, we keep the target story about the bar identical to *Baseline*, but increase the similarity of the two decoy stories to the target story by modifying both the text struc-

[30]The story-statistic gap in belief impact is close to zero for the 1-product scenario.

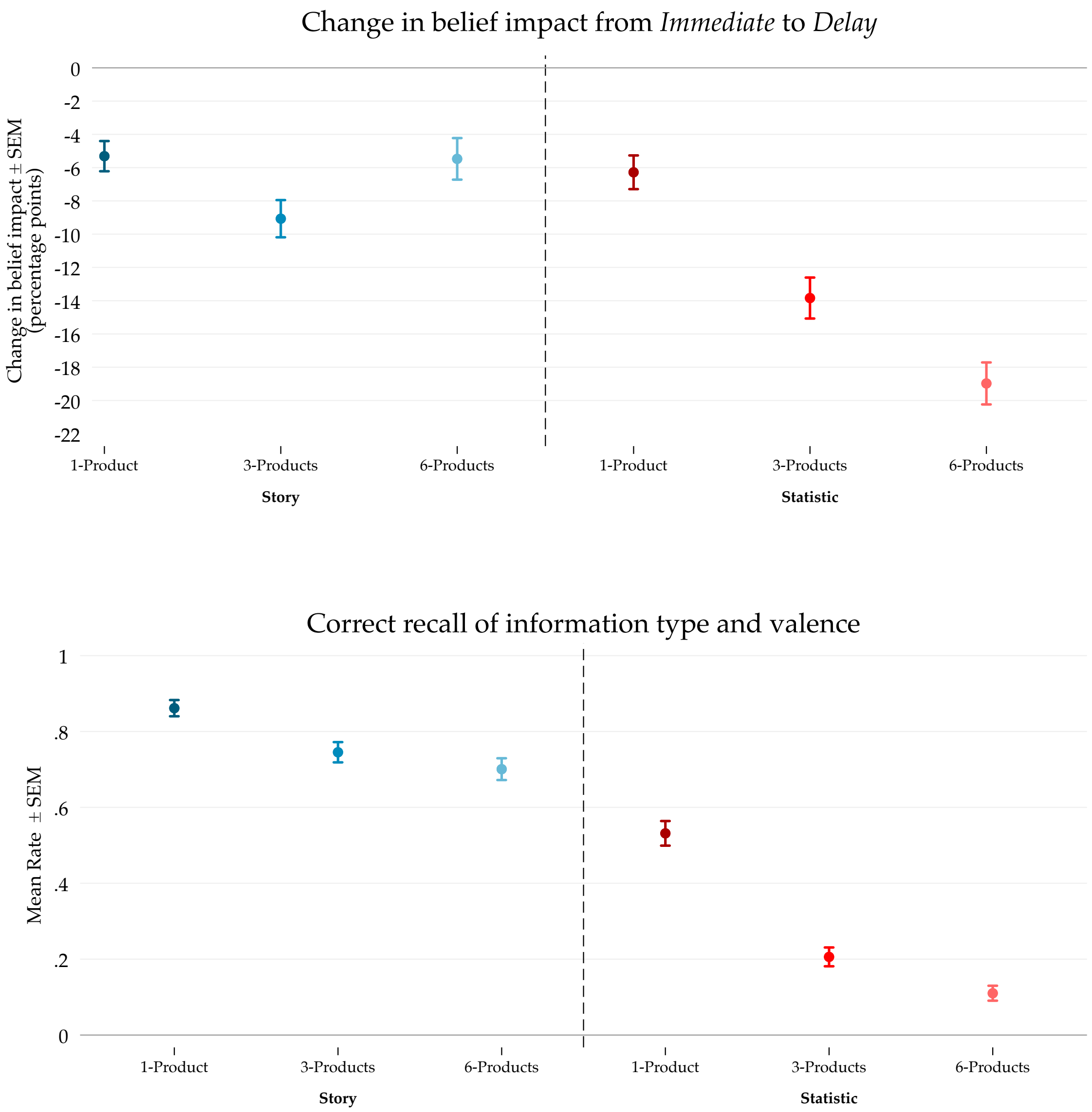


Figure 5: Change in belief impact and recall in Mechanism Experiment 2 (1,018 respondents). The top panel displays the change in belief impact in percentage points, defined as the difference in belief impact between *Delay* and *Immediate*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The bottom panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The dark blue markers illustrate change in belief impact and recall for the *1-product* condition, the blue markers illustrate the change in belief impact and recall for the *3-product* condition, while the light blue markers display the change in belief impact and recall for the *6-product* condition. Whiskers indicate one standard error of the mean.

ture and content. Specifically, in *Story Similarity*, the three products are still a cafe, a restaurant, and a bar, but all stories revolve around the interior design of the respective venues. Thus, our treatments fix the target story and only manipulate the similarity between the two decoy stories and the target story. All other design aspects are identical between the conditions. Appendix D.2 reproduces all stories that we used.

Sample and pre-registration. We recruited 1,150 respondents, of which 1,069 qualified for the follow-up. Respondents were randomized into the two conditions described above and a third condition described in Section 4.3.2. 879 respondents completed the follow-up survey. After the pre-specified sample restrictions, we have a sample size of 872, corresponding to a completion rate of 79 percent.[31] The pre-registration is available on AsPredicted, see `https://aspredicted.org/v7hh6.pdf`. The plan contains the two conditions described in this section as well as a third condition described in Section 4.3.2.

Prediction. The decay of belief impact for the target story is more pronounced in *Story Similarity* compared to the *Baseline* condition.

Results. The top panel of Figure 6 shows data on the belief impact of the target story in *Immediate* and *Delay*, separately for *Story Similarity* and *Baseline*. In line with the model prediction, the slope in belief impact is steeper in *Story Similarity* compared to *Baseline*. Delayed belief impact is significantly lower in *Story Similarity* than in *Baseline*, even though immediate belief impact is larger in the former condition.[32] While average delayed belief impact in *Story Similarity* is 1.25 p.p. (s.e. 1.17), it is 4.43 p.p. (s.e. 1.09) in *Baseline*. Table 2 confirms this visual pattern and shows that the difference-in-differences in belief impact (difference in slopes) is statistically significant ($p < 0.01$).

The bottom panel illustrates similar patterns for recall: Among respondents in *Baseline*, 47.04 p.p. (s.e. 0.03) correctly recall the information, compared to only 37.37 p.p. (s.e. 0.03) in *Story Similarity*. This difference of 10 p.p. is statistically significant ($p < 0.01$).

This finding has two implications. First, it provides strong evidence for the power of similarity relationships in determining the decay of belief impact and recall accuracy. Second, it delineates the limits of the stickiness of stories in memory. If the memory database contains many similar stories, retrieval of a target story gets crowded out and it becomes less likely that this story comes to mind.

[31]The completion rate to the follow-up survey does not differ significantly across treatment groups ($p = 0.79$).

[32]Immediate belief impact might be slightly larger in *Story Similarity* due to a more pronounced contrast effect when the target story is more similar to the decoy stories. Naturally, higher immediate belief impact works against us finding differences in the *Delay* condition and also does not affect our evidence on recall, which we discuss below.

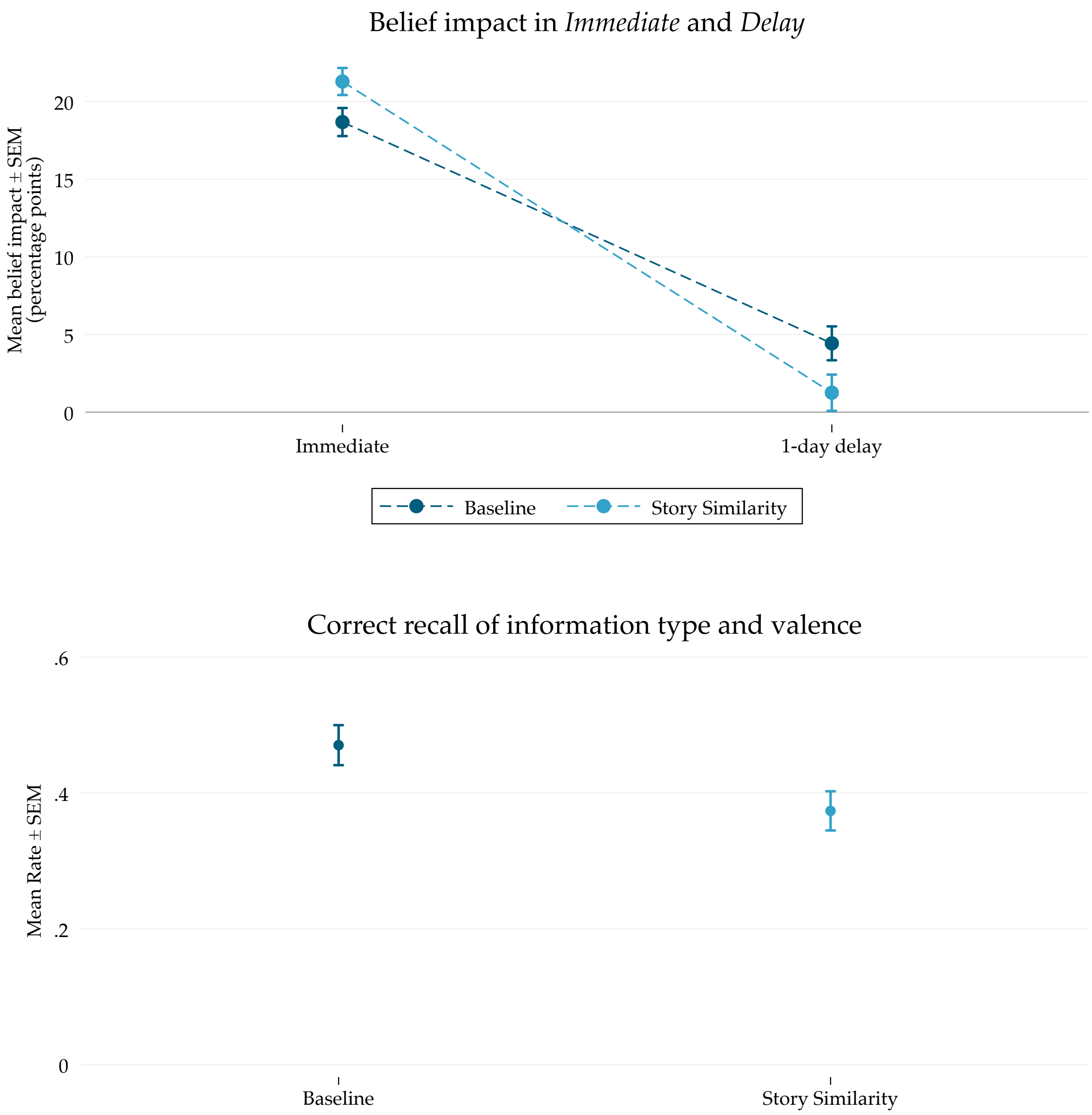


Figure 6: Belief impact and recall in Mechanism Experiment 3 (872 respondents). The top panel displays belief impact in percentage points, separately for conditions *Immediate* and *Delay*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The bottom panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The dark blue markers illustrate belief impact and recall for *Baseline*, while the light blue markers illustrate belief impact and recall for *Story Similarity*. Whiskers indicate one standard error of the mean.

4.2.3 Cue similarity

To obtain a complete picture of cross-similarity, we also conduct treatments that manipulate the similarity of product cues. Our model posits that more similar cues should decrease delayed belief impact by increasing cross-similarity. Yet, our model makes no clear-cut predictions about differential effects for stories versus statistics. The central insight emerging from an additional set of experimental analyses is that higher cue similarity indeed leads to a decrease in delayed belief impact, both for stories and statistics. We relegate details to Appendix A.2.

Our third main result summarizes the combined evidence on manipulations of cross-similarity:

Result 3. *We report experiments highlighting that cross-similarity significantly shapes delayed belief impact and recall. The story-statistic gap increases in the number of product scenarios. Moreover, delayed belief impact and recall of a story is impaired by higher similarity to stories in other scenarios.*

The practical implication of Result 3 is that policy communication may be particularly effective when the disseminated information is unique and distinct from other circulating information. Yet, stories as a communication device lose their edge in environments where similar stories circulate.

4.3 Cue-target Similarity

As outlined in Section 3, the inherent similarity between a cue and the target information content conceivably might shape recall and delayed belief impact. Appendix H.1 provides an extension to our baseline model that accommodates and formalizes this result on cue-target similarity. To test for the importance of cue-target similarity, we conduct experiments that manipulate (i) the similarity between a cue and the corresponding statistic and (ii) the similarity between a cue and the corresponding story.

4.3.1 Cue-Statistic Similarity

The similarity between a statistic and the cue might play a role to the extent that the *format* in which the statistic is provided resembles the format of the question that people are asked. For example, both might be presented in the similar format of a fraction, but can also be presented in less similar ways, as is the case in our main experiment, where one is an absolute number and one a percentage. Put differently, cue-target similarity between the statistic and the cue might be driven by the question part of the cue.

Design. The experiment features one key treatment variation. The *Dissimilar Format* treatment elicits beliefs as before – about the likelihood that a randomly chosen review is positive – and thus exactly corresponds to our main experiment. In the *Similar Format* condition, by contrast, we elicit beliefs about the percentage of positive reviews in the overall population of reviews of the product.[33] The rationale of this manipulation was that in *Similar Format*, the question people answered is more similar to the type of information they were provided with in the statistic condition, which is about the count of positive reviews in a subsample of reviews about the product.

As an additional similarity manipulation, we randomize whether the statistical information itself is expressed in terms of an absolute number of positive reviews in a subsample (*Statistic Dissimilar*) – as in our main study – or in terms of a percentage of positive reviews in a subsample (*Statistic Similar*). This means we run a total of four between-subject conditions, reflecting a 2 (*Dissimilar Format / Similar Format*) × 2 (*Statistic Dissimilar / Statistic Similar*) factorial design.

The comparison between the *Similar Format* and the *Dissimilar Format* conditions allows us to examine how the similarity between the statistic and the cue affects recall. The *Statistic Similar* condition makes the additional information even more similar to the cue whenever the question format involves a fraction, creating a "high cue-statistic similarity" condition. This provides additional variation to study the role of statistic-cue similarity.

Sample and pre-registration. 1,532 respondents completed the baseline survey and also met the inclusion criteria. 922 respondents completed the follow-up survey, corresponding to a 60 percent completion rate.[34] The pre-registration for this experiment is available on AsPredicted, see `https://aspredicted.org/ZFF_88V`.

Prediction. *Similar Format* and the interaction effect between *Similar Format* and *Similar Statistic* decrease the decay of belief impact and forgetting.

Results. Appendix Figure A.3 and Table A.4 document that the *Similar Format* has a positive, yet small effect on delayed belief impact and recall. The decay of belief impact is somewhat smaller in *Similar Format* than *Dissimilar Format*. This effect is more pronounced in the recall data, and reaches significance for the case of statistical information (column (4) of Table A.4), in line with the notion that a higher similarity of the question format to the statistic slightly improves retrieval.

Moreover, we find that the effect of displaying statistical information as a percentage instead of an absolute number does not have significant effects on belief impact and recall. More specifically, we also do not observe a significant interaction effect between

[33]We accordingly adjust the description of incentives, which are framed in terms of guesses about the true percentage of positive reviews in this condition, but kept otherwise identical.

[34]The completion rate to the follow-up survey does not differ significantly across treatment groups ($p = 0.59$).

the question format and the display format of statistical information. A plausible interpretation is that these are already highly similar at baseline (in our main study), so that the manipulation of making them even more similar makes little difference. In practice, we might expect that the question format and the display format are much *less* similar to each other, and that variation in similarity across contexts plays a larger role.

Considering the insignificant effects on delayed belief impact and the mixed evidence on recall, we take this as, at best, suggestive evidence that the similarity of the question format to the piece of additional information shapes forgetting.

4.3.2 Cue-Story Similarity

The qualitative information contained in stories is often intrinsically related to the corresponding cues, e.g., a story for the cue "Restaurant" will typically feature restaurant-related content. Stories are typically associated with the part of the cue that encodes the scenario name. As a result, the cue-target similarity of anecdotal information may be higher for stories than for statistics. To examine the role of cue-target similarity, we conduct experiments that manipulate the extent of similarity between stories and cues.

Design. We employed the same *Baseline* condition as in Section 4.2.2, but compared it to a different treatment, *Cue-Story Similarity*.[35] This condition relied on the same decoy stories for the cafe and the restaurant as *Baseline*. However, the target story about a bar involved an experience that is entirely unrelated and unspecific to a bar. The objective was to exogenously reduce the similarity between the target story and the target cue, keeping all other design aspects fixed. Appendix D.2 contains the stories that we used.

Prediction. Recall accuracy is lower in *Cue-Story Similarity* than in baseline.

Results. As specified in the pre-analysis plan (`https://aspredicted.org/v7hh6.pdf`), we focus on the recall data, because the immediate belief impact was likely to be much stronger in *Baseline* than in *Cue-Story Similarity* (as was indeed the case in our data). Column (4) of Table 2 documents that, while correct recall in the *Baseline* was 47.04 percent (s.e. 0.03), recall in the *Cue-Story Similarity* condition was 40.21 percent (s.e. 0.03) percent. This difference is statistically insignificant ($p = 0.10$). Column (2) of Table 2 reports results on belief impact. The decay of belief impact points in the opposite direction, i.e., *Cue-Story Similarity* was associated with lower decay of belief impact over time. This result is hard to interpret given different baseline levels of belief

[35]Participants were randomized to either the *Cue-Story Similarity* condition, the *Story Similarity* condition or the *Baseline* condition

impact, but nevertheless highlights that the overall evidence of this manipulation for cue-target similarity is mixed. Our fourth main result is given as follows:

Result 4. *The effect of cue-target similarity for the story-statistic gap is, at best, of minor importance relative to the strong and consistent results of cross-similarity.*

Table 2: (Cue-)story similarity

	Dependent variable:			
	Belief Impact		Combined Recall	
Sample:	Story (1)	Cue-Story (2)	Story (3)	Cue-Story (4)
Story Similarity	2.61^{**}		-0.097^{**}	
	(1.25)		(0.04)	
Delay × Story Similarity	-5.79^{***}			
	(1.78)			
Cue-Story Similarity		-6.21^{***}		-0.068
		(1.21)		(0.04)
Delay × Cue-Story Similarity		4.27^{***}		
		(1.62)		
Delay	-14.2^{***}	-14.2^{***}		
	(1.16)	(1.16)		
Control Mean	18.68	18.68	0.47	0.47
Observations	1136	1136	568	568
R^2	0.21	0.15	0.01	0.00

Notes. This Table shows data from Mechanism Experiment 3 (872 respondents). *Delay* is an indicator taking value 1 for respondents in the follow-up survey, and value 0 for respondents in the baseline survey. *Story Similarity* takes value 1 for respondents who received similar decoy stories, and zero otherwise. *Cue-Story Similarity* takes value 1 for respondents who received a generic story that was less intrinsically related to the cue compared to the baseline condition. Columns (1) and (3) include respondents who were in the story similarity and baseline condition. Columns (2) and (4) include respondents who were in the cue-story similarity and baseline condition. Columns (1) and (2) display results on belief impact. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. Columns (3) and (4) display the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. OLS estimates, standard errors clustered at the respondent level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Decomposing the Story-Statistic Gap

The mechanism experiments reported in Section 4 provide causal evidence for the model of cue-dependent memory outlined in Section 3 as an explanation of the story-statistic gap. While pinpointing the roles of contextual associations and interference through different forms of cross-similarity, the body of evidence presented so far leaves some fundamental questions on the nature of memory unaddressed. One distinction with far-reaching consequences for both theorizing and empirical work on memory is that between the intensive margin – successfully retrieving a memory, but with partial information loss – and the extensive margin – retrieval failures. In the following, we provide a heuristic decomposition with the purpose of quantifying the importance of the intensive and extensive margins of memory in driving the story-statistic gap.[36]

To examine these ideas, note first that the basic memory retrieval process in models like ours can have three different possible outcomes: (a) retrieval failure, (b) successful retrieval without information loss, and (c) successful retrieval of a memory trace that is subject to information loss. Each of these retrieval outcomes is associated with a different signature in the decay of belief impact. First, the DM may not retrieve a target memory and therefore returns to the prior (class *FullDecay*). Such retrieval failure creates a clear benchmark of full belief decay. This is the extensive margin of forgetting. Second, the DM may correctly recall the full wealth of information contained in a story or statistic (class *ZeroDecay*). In that case, the DM would state his past posterior belief and we would observe no (or little) decay in beliefs over time. Third, the DM may successfully recall the target memory trace, but the retrieved information is subject to information loss (class *IntermediateDecay*). This is the intensive margin of forgetting, which would cause some (but not full) decay in belief movement. As we discussed in Section 3, while our baseline model focuses on the extensive margin of forgetting, it easily generalizes to adding an intensive margin (see Appendix H.3). There is no guidance from previous empirical or theoretical work on the precise nature of information loss in episodic memory, i.e., whether and statistics or stories are "summarized".[37]

The combination of recall and panel belief data in our experiments allow us to shed some light on the relative magnitudes of these recall channels. Specifically, we proceed as follows: We use recall data to identify retrieval failures and test whether this class is, in fact, associated with *FullDecay* in the corresponding stated beliefs. We then turn to the remaining data, which, by construction, only include observations where people correctly remember at least some of the information – information type and valence.

[36]The analyses in this section are exploratory in nature and were not pre-registered.

[37]The average signature of information loss in the *IntermediateDecay* class is plausibly bounded by the other two classes: information loss should lead to *some* belief decay. These bounds seem plausible irrespective of the exact mechanisms underlying the information loss. If the information loss is associated with the storage of a summary of the information, i.e., the DM treats statistical information based on the expected weight and extremity conditional on a statistic's valence, belief decay would fall between the two extremes of no and full decay established above (see Appendix H.3) If information loss is associated with pure noise, i.e., a statistic is randomly retrieved as more or less extreme than the truth, then there will be no belief decay on average (conditional on successful retrieval).

We focus on the corresponding belief data to assess the relative shares of *ZeroDecay* and *IntermediateDecay*. The following analysis focuses on our baseline experiment reported in Section 2.

First, the bottom panel of Figure 1 identifies the fraction of beliefs associated with retrieval failure, specifically, a failure to retrieve information (type and valence) across conditions. Following this metric, 38 percent of observations in *Story*, but 74 percent in *Statistic* fail to retrieve relevant information about the target trace. According to the model, these observations should be associated with beliefs that fully return to the prior of 50%, implying a belief impact of zero (class *FullDecay*). Figure 7 displays the story-statistic gap in belief impact separately for the sample of observations associated with correct and incorrect recall (following the definition of the bottom panel of Figure 1). The average belief impact for observations classified as *FullDecay* indeed reverts to close to zero in *Delay*, as predicted by the model.

Second, now turning to observations associated with correct recall of type and valence, we can identify the class of *ZeroDecay* as those that state identical beliefs in *Immediate* and *Delay*. This comprises 37.67 percent (47.66 percent of correct recall observations) of all observations in *Story* and 30.65 percent (56.05 percent of correct recall observations) of all observations in *Statistic*. Note that these figures likely identify a lower bound, because they do not take into account potential noise in beliefs. If people in *ZeroDecay* answer the belief questions with some added random noise, there would be no average belief decay, yet many would state beliefs that differ between the two periods.[38]

Finally, we turn to the remaining class, *IntermediateDecay*, which is associated with correct recall of type and valence, but at the same time features beliefs with *some* intensive-margin information loss by virtue of neither being part of *ZeroDecay* nor *FullDecay*. Above we already classified a substantial lower bound for the class *ZeroDecay*. Figure 7 displays average belief decay among observations associated with correct recall of type and valence. Strikingly, it reveals that there is zero average belief decay in the *Story* condition and a quantitatively minor, only marginally significant decay in the *Statistic* condition. Put differently, conditional on correct recall, we see close to no evidence for belief decay, strongly suggesting that the intensive margin of selective memory plays a minor role here.

Taken together, the following main insights emerge: The lion's share of the story-statistic gap appears to be driven by the extensive margin of memory, i.e., differential retrieval failures for stories and statistics. This underscores our mechanism evidence on the central role of cross-similarity that drives interference. In addition, and perhaps

[38]We can instead apply a more lenient benchmark than precisely zero decay, but, as will be clear below, this will, if anything, only strengthen our conclusion.

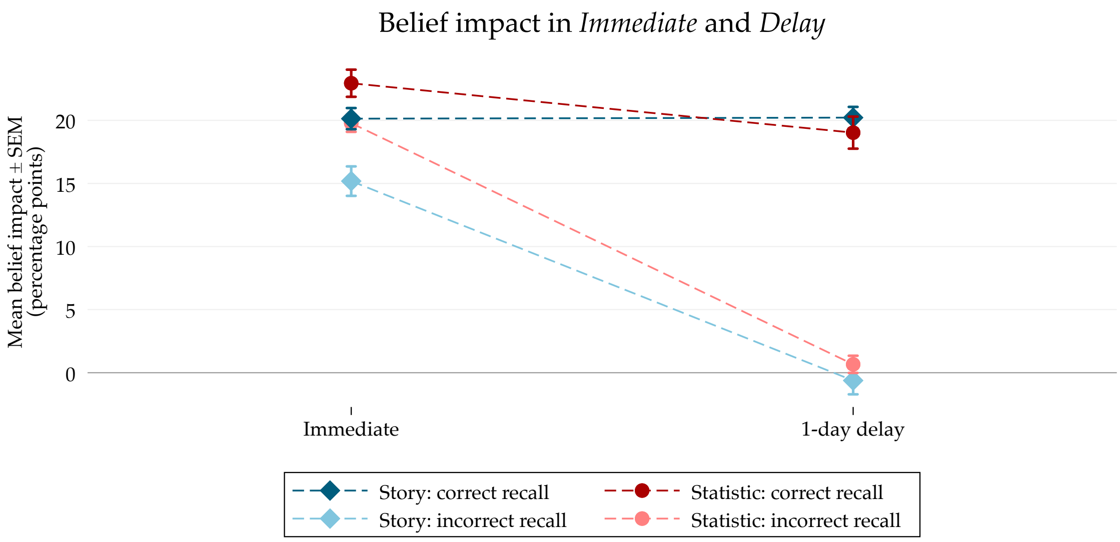


Figure 7: The decay of belief impact by recall accuracy in the baseline experiment (984 respondents). The figure displays belief impact in percentage points, separately for conditions *Immediate* and *Delay*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The dark blue markers illustrate belief impact for stories with correct recall, while the light blue markers illustrate belief impact for stories with incorrect recall. The dark red markers illustrate belief impact for statistics with correct recall, while the light red markers illustrate belief impact for statistics with incorrect recall. Whiskers indicate one standard error of the mean.

most strikingly, this exercise highlights that the intensive margin of selective recall has little to no significant quantitative impact.

6 Discussion and Conclusion

This paper documents a story-statistic gap in memory. As time passes, the effect of information on beliefs generally decays, but this decay is much less pronounced for stories than for statistics. Using recall data, we show that stories are more accurately retrieved from memory than statistics. We causally show that this pattern is driven by the contextual features provided through stories: adding context to statistics increases delayed belief impact and recall accuracy. Guided by a simple model of cue-dependent memory, we experimentally examine the explanatory power of different features of cross-similarity as sources of interference. Consistent with the model, our evidence suggests that similarity relationships are the central force underlying the story-statistic gap. Stories tend to be distinct, whereas the abstract nature of statistics makes them similar to other, but irrelevant information.

A key insight from our analysis of underlying mechanisms is that the features of memory that favor the recall of stories are not unique to stories. It does not seem to be the case that the way we store and retrieve stories is fundamentally different from how we store and retrieve statistics. Rather, cross-similarity and interference account for the lion's share of the story-statistics gap.

We establish two novel findings that inform future work on memory. First, the role of interference as driven by cross-similarity appears to be far more powerful than the effects of cue-target similarity in the settings studied here. Second, our memory decomposition provides striking evidence that the extensive margin of retrieval failures appears to be the key driver of the story-statistic gap, rather than partial information loss through the storage of summaries of the information in memory.

Stories in the mass media. Our findings have implications for understanding several real-world phenomena. Mass media provides not only facts and statistics, but also relies on anecdotes about individual cases, which provide detailed qualitative information. Consider allegations about election fraud in the context of the 2020 presidential election, where some outlets reported stories about individual instances of election fraud, even though these were rare exceptions. Likewise, consider news reporting about welfare fraud where anecdotes about individual cases are abundant in the news media, but stand in stark contrast to official statistics on fraud incidence. For example, Ronald Reagan, beginning with his 1976 campaign, told extreme stories about "welfare queens":

> *She has 80 names, 30 addresses, 12 Social Security cards and is collecting veterans' benefits on four non-existing deceased husbands. And she's collecting Social Security on her cards. She's got Medicaid, getting food stamps, and she is collecting welfare under each of her names. Her tax-free cash income alone is over $150,000.*

Similarly, consider mass media coverage of immigration. While statistics about low crime rates among immigrants are widely reported by news outlets, extreme stories about immigrants committing severe crimes also regularly hit the headlines.

Our results indicate that stories disseminated in this way can have powerful effects on beliefs as they may come to mind more easily than more representative statistical information. This provides a potential explanation for the emergence of widely documented misperceptions about many real-world topics, and for the persistence of these distortions.

Policy communication. Our results also have implications for the communication of statistical information. If policymakers, marketers or leaders aim to convey statistical information effectively, they may wish to complement it with contextual, anecdotal associations to ensure that the information sticks with the audience. For instance, statistical information about economic quantities could be coupled with anecdotal information that is consistent and inherently reminiscent of the embedded statistical information. Moreover, our results highlight that persuaders should factor in the time structure when picking their mode of persuasion: if messaging occurs close in time to the audience's anticipated action, statistics and quantitative facts can be more powerful than stories; yet, as soon as a delay is involved, stories trump statistics.

Sharing of stories. A natural extension of our work is to examine *which* stories tend to be shared in practice. For example, the most extreme and surprising stories may be particularly likely to be told and re-told because they are "worth telling". If confirmed, this would point to the possibly harmful implications of the story-statistic gap in memory: the less representative the stories that are shared, the larger the final belief distortions, providing an explanation for the widely observed persistence of biased beliefs.

References

Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar, "Overreaction in Expectations: Evidence and Theory," *Quarterly Journal of Economics*, forthcoming.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva, "Immigration and Redistribution," *Review of Economic Studies*, 03 2022, *90* (1), 1–39.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective Models of the Macroeconomy: Evidence from Experts and Representative Samples," *Review of Economic Studies*, 02 2022, *89* (6), 2958–2991.

—, **Ingar Haaland, Christopher Roth, and Johannes Wohlfart**, "Narratives about the Macroeconomy," *Working Paper*, 2022.

Ba, Cuimin, Aislinn Bohren, and Alex Imas, "Over- and Underreaction to Information," *Working Paper*, 2023.

Baddeley, Alan, Michael W. Eysenck, and Michael C. Anderson, *Memory*, Routledge, 2020.

Bénabou, Roland, Armin Falk, and Jean Tirole, "Narratives, Imperatives, and Moral Reasoning," *Working Paper*, 2018.

Bordalo, Pedro, Giovanni Burro, Katie Coffman, Nicola Gennaioli, and Andrei Shleifer, "Imagining the Future: Memory, Simulation and Beliefs about Covid," *Working Paper*, 2022.

—, **John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer**, "Memory and Probability," *Quarterly Journal of Economics*, 2023, *138* (1), 265–311.

—, **Katherine Coffman, Nicola Gennaioli, Frederik Schwerter, and Andrei Shleifer**, "Memory and Representativeness.," *Psychological Review*, 2021, *128* (1), 71.

—, **Nicola Gennaioli, and Andrei Shleifer**, "Memory, Attention, and Choice," *Quarterly Journal of Economics*, 2020, *135* (3), 1399–1442.

—, —, **Yueran Ma, and Andrei Shleifer**, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–82.

Bower, Gordon H and Michal C Clark, "Narrative stories as mediators for serial learning," *Psychonomic Science*, 1969, *14* (4), 181–182.

Brewer, William F and James C Treyens, "Role of schemata in memory for places," *Cognitive Psychology*, 1981, *13* (2), 207–230.

Bruner, Jerome, "Actual Minds, Possible Worlds (Jerusalem-Harvard Lectures)," 1987.

Bursztyn, Leonardo, Aakaash Rao, Christopher P Roth, and David H Yanagizawa-Drott, "Opinions as Facts," *Review of Economic Studies*, 2022. rdac065.

__, Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth, "Justifying Dissent," *Quarterly Journal of Economics*, 2023. qjad007.

Charles, Constantin, "Memory and Trading," *Working Paper*, 2022.

Conlon, John J and Dev Patel, "What Jobs Come to Mind? Stereotypes about Fields of Study," *Working Paper*, 2022.

Danz, David, Lise Vesterlund, and Alistair Wilson, "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*, 2022, *112* (9), 2851–2883.

Eliaz, Kfir and Ran Spiegler, "A Model of Competing Narratives," *American Economic Review*, 2020, *110* (12), 3786–3816.

Enke, Benjamin, "What You See Is All There Is," *Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

__ and Florian Zimmermann, "Correlation Neglect in Belief Formation," *Review of Economic Studies*, 2019, *86* (1), 313–332.

__, Frederik Schwerter, and Florian Zimmermann, "Associative Memory and Belief Formation," *Working Paper*, 2022.

Foer, Joshua, *Moonwalking with Einstein: The Art and Science of Remembering Everything*, Penguin, 2012.

Fryer, Bronwyn, "Storytelling That Moves People," *Harvard Business Review*, 2003.

Gennaioli, Nicola and Andrei Shleifer, "What Comes to Mind," *Quarterly Journal of Economics*, 2010, *125* (4), 1399–1433.

Graeber, Thomas, "Inattentive Inference," *Journal of the European Economic Association*, 09 2022. jvac052.

Hartzmark, Samuel, Samuel Hirshman, and Alex Imas, "Ownership, Learning, and Beliefs," *Quarterly Journal of Economics*, 2021, *136* (3), 1665–1717.

Heath, Chip and Dan Heath, *Made to Stick: Why Some Ideas Survive and Others Die*, Random House, 2007.

Hossain, Tanjim and Ryo Okui, "The Binarized Scoring Rule," *Review of Economic Studies*, 2013, *80* (3), 984–1001.

Jiang, Zhengyang, Hongqi Liu, Cameron Peng, and Hongjun Yan, "Investor Memory and Biased Beliefs: Evidence from the Field," *Working Paper*, 2022.

Kahana, Michael Jacob, *Foundations of Human Memory*, OUP USA, 2012.

Kendall, Chad W and Constantin Charles, "Causal Narratives," *Working Paper*, 2022.

Kensinger, Elizabeth A and Daniel L Schacter, "Memory and Emotion," 2008.

Kwon, Spencer Yongwook and Johnny Tang, "Extreme Events and Overreaction to News," *Working Paper*, 2023.

Mandler, Jean M., *Stories, Scripts, and Scenes: Aspects of Schema Theory*, Erlbaum, 1984.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa, "Failures in Contingent Reasoning: The Role of Uncertainty," *American Economic Review*, 2019, *109* (10), 3437–74.

McAdams, Dan P, "Narrative Identity," *Handbook of Identity Theory and Research*, 2011, pp. 99–115.

Michalopoulos, Stelios and Melanie Meng Xue, "Folklore," *Quarterly Journal of Economics*, 2021, *136* (4), 1993–2046.

Monarth, Harrison, "The Irresistible Power of Storytelling as a Strategic Business Tool," *Harvard Business Review*, 2014.

Morag, Dor and George Loewenstein, "Narratives and Valuations," *Working Paper*, 2021.

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer, "Data quality of platforms and panels for online behavioral research," *Behavior Research Methods*, 2022, *54* (4), 1643–1662.

Schacter, Daniel L, *Searching for Memory: The Brain, The Mind, And The Past*, Basic books, 2008.

Schank, Roger C. and Robert P. Abelson, *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*, Erlbaum, 1977.

Shepard, Roger N and Lynn A Cooper, *Mental images and their transformations.*, The MIT Press, 1986.

Shiller, Robert J, "Narrative Economics," *American Economic Review*, 2017, *107* (4), 967–1004.

_ , *Narrative Economics*, Princeton University Press, 2020.

Zimmermann, Florian, "The Dynamics of Motivated Beliefs," *American Economic Review*, 2020, *110* (2), 337–61.

Online Appendix: Stories, Statistics, and Memory

Thomas Graeber   Christopher Roth   Florian Zimmermann

A Additional Results

A.1 Robustness to Decoy Information

To further probe into the robustness of the story-statistic gap, we examine the role of features of the decoy information using an additional experiment, in which we systematically manipulate the type and valence of decoy information.

We exogenously manipulate the type of information for the two decoy scenarios. Respondents either received two statistics for the decoys, two stories or twice no information. In addition, in contrast to the baseline design, we fully randomize the valence of the information provided for each scenario.

In the follow-up survey, we elicit beliefs exactly as in the baseline survey.

Sample and pre-registration. We recruited 2,250 respondents for the baseline survey. 2048 respondents qualified for the follow-up survey. 1,613 respondents completed the one-day follow-up survey. After the pre-specified sample restrictions, our final sample consists of 1,548 respondents, corresponding to a 76% completion rate.[1] The pre-registration for this experiment can be found on AsPredicted, see `https://aspredicted.org/qy3wq.pdf`.

Results. Figure A.1 summarizes our results. The left-hand panel shows the changes in belief impact between immediate and delay for the target story and target statistic across the three different decoy conditions. The right panel analogously displays the rate of correct recall across the three conditions separately for the story and statistic target.

We make three observations: First, there is a robust story-statistic gap across all conditions. The story-statistic gap has a similar magnitude irrespective of the number and type of decoy information. This is visible across both our beliefs data and the incentivized structured recall elicitation.[2] Second, we observe small effects at best of the number of decoy information. This suggests that memory load per se has muted effects on belief impact in this setting. Third, we do not observe significant effects of the type of decoy information on the size of the story-statistic gap. Jointly these results imply that the

[1] The completion rate to the follow-up survey does not differ significantly across treatment groups ($p = 0.60$).

[2] Results from our structured recall task are very similar to results from the free recall task, providing a validation of the latter.

story-statistic gap is robust to basic features of the decoys and that – in a setting with only three scenarios – the type and number of decoys is not a key driver of the decay of belief impact.

Figure A.2 shows how belief impact and recall of stories vary depending on the valence of decoy information. Compared to the statistics benchmark, we again find a robust and sizable story-statistic gap across decoys of different valence. We further find that decoy valence has a small but directionally plausible effect on the size of the gap: when decoy information has the same valence as the target information, both recall and delayed belief impact is larger than when the decoy information is mixed or of opposite sign.

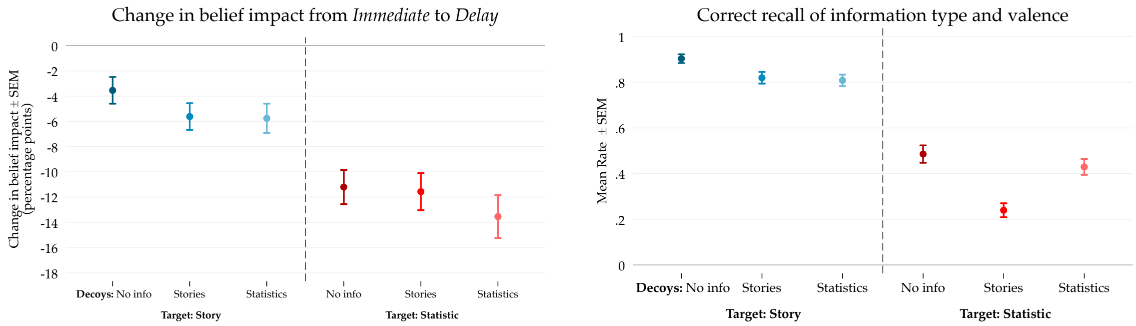


Figure A.1: Belief impact and recall in Robustness Experiment: The role of Decoy Information (1,513 respondents). The left panel displays the change in belief impact in percentage points, defined as the difference in belief impact between *Delay* and *Immediate*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The bottom panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The dark blue (dark red) markers illustrate change in belief impact and recall for stories (statistics) for the *Decoys: No Info* condition, the blue (red) markers illustrate the change in belief impact and recall for stories (statistics) for the *Decoys: Stories* condition, while the light blue (light red) markers display the change in belief impact and recall for stories (statistics) for the *Decoys: Statistics* condition. Whiskers indicate one standard error of the mean.

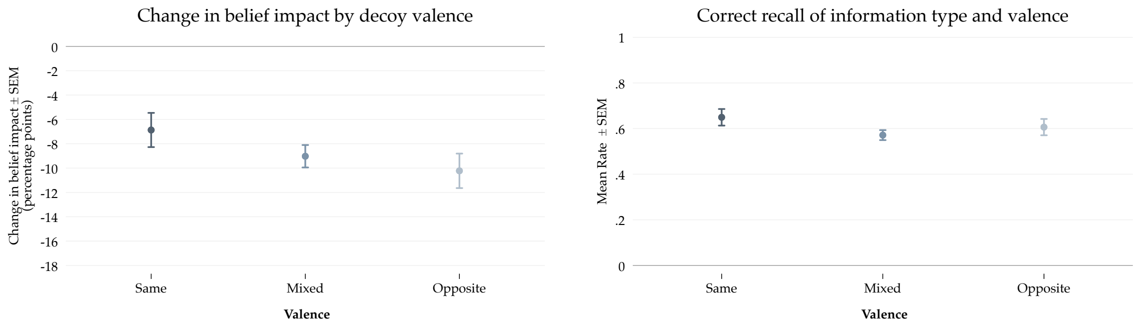


Figure A.2: Belief impact and recall in Robustness Experiment: The role of Decoy Information (1,513 respondents). The top panel displays the change in belief impact in percentage points, defined as the difference in belief impact between *Delay* and *Immediate*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The right panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The dark gray markers illustrate change in belief impact and recall for targets when decoys have the target's valence, the gray markers illustrate change in belief impact and recall for targets when decoys have mixed valence, while the light gray markers display the change in belief impact and recall for targets when decoys have the target's opposite valence. Whiskers indicate one standard error of the mean.

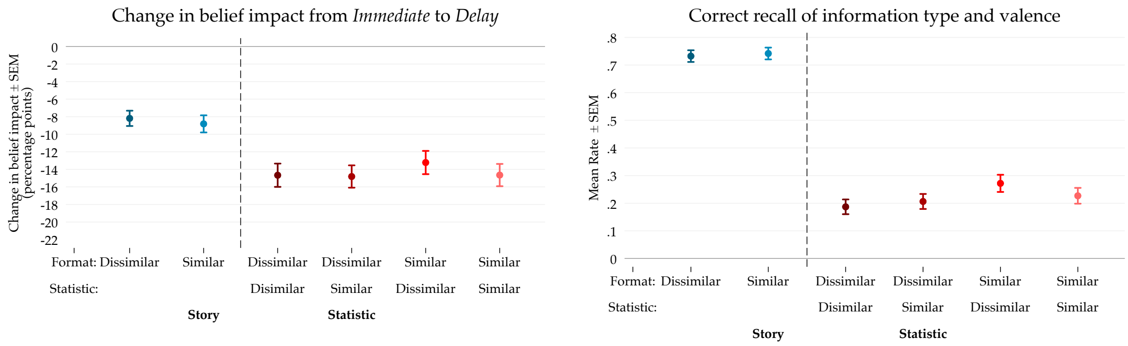


Figure A.3: Belief impact and recall in Mechanism Experiment 5: Question Format and statistic display (959 respondents). The left panel displays the change in belief impact in percentage points, defined as the difference in belief impact between *Delay* and *Immediate*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The right panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The dark blue markers illustrate change in belief impact and recall for the *Dissimilar Format* condition for stories, the light blue markers illustrate change in belief impact and recall for the *Similar Format* condition for stories, while the most dark red for the *Dissimilar Format / Statistic Dissimilar* condition, the dark red for the *Dissimilar Format / Statistic Similar* condition, the red for the *Similar Format / Statistic Dissimilar* condition and the light red for the *Similar Format / Statistic Similar* condition. Whiskers indicate one standard error of the mean.

A.2 Cue Similarity - Details

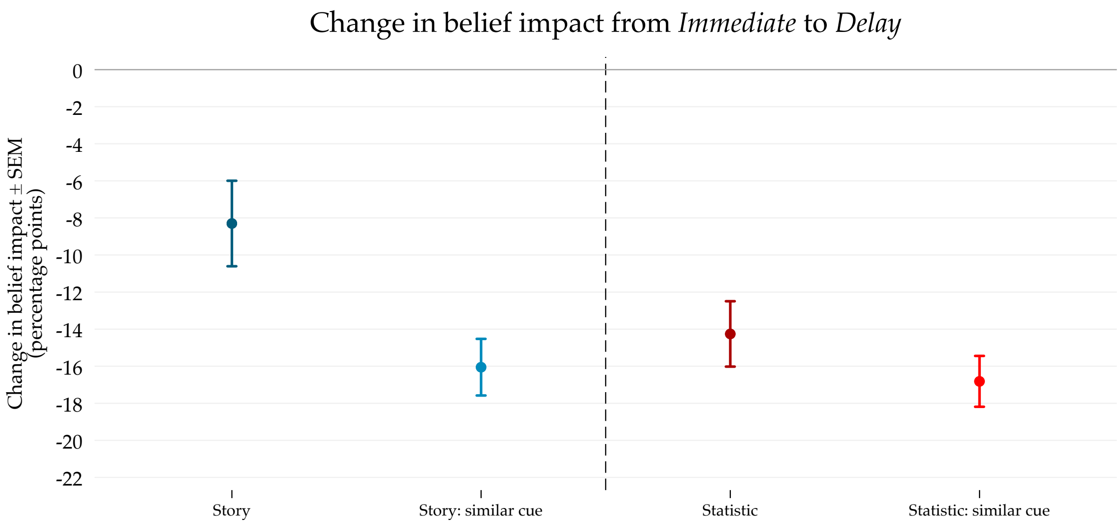
Design. Our design varied the similarity of cues, holding everything else constant. The basic set-up follows our main experiment. In *Baseline*, the three cues were Restaurant A, Bicycle and Videogame, with Restaurant always being the target cue in our analysis. Participants either received a story or a statistic in the restaurant scenario. In *Cue Similarity*, we kept everything identical to *Baseline*, including the target cue Restaurant A, but changed the labels of the decoy cues to Restaurant B and Restaurant C. In our analysis, as pre-registered, we compare belief impact and recall between the *Baseline* and *Cue Similarity*, separately for respondents who received a story and a statistic.

Sample and pre-registration. We recruited 1,150 respondents, of which 999 were eligible for the followup. Out of those, 599 respondents completed the follow-up survey. After the pre-specified sample restrictions, our final sample consists of 583 respondents, corresponding to a completion rate of 59 percent.[3] The pre-registration for this experiment is available at `https://aspredicted.org/h2fr3.pdf`.

Prediction. The decay of belief impact and forgetting of both stories and statistics are more pronounced in *Cue Similarity* than *Baseline*.

Results. Panel A of Figure A.4 displays changes in belief impact between *Immediate* and *Delay* for both treatments. The figure reveals that the change in belief impact is substantially larger in the cue similarity condition. This holds true both when the target is a story and when the target is a statistic (though the effect is less pronounced for statistics, possibly due to already very low levels of delayed belief impact and recall). Panel B of Figure A.4 largely displays the same pattern using our recall data. Table A.6 confirms this result.

[3]The completion rate to the follow-up survey does not differ significantly across treatment groups ($p = 0.53$).



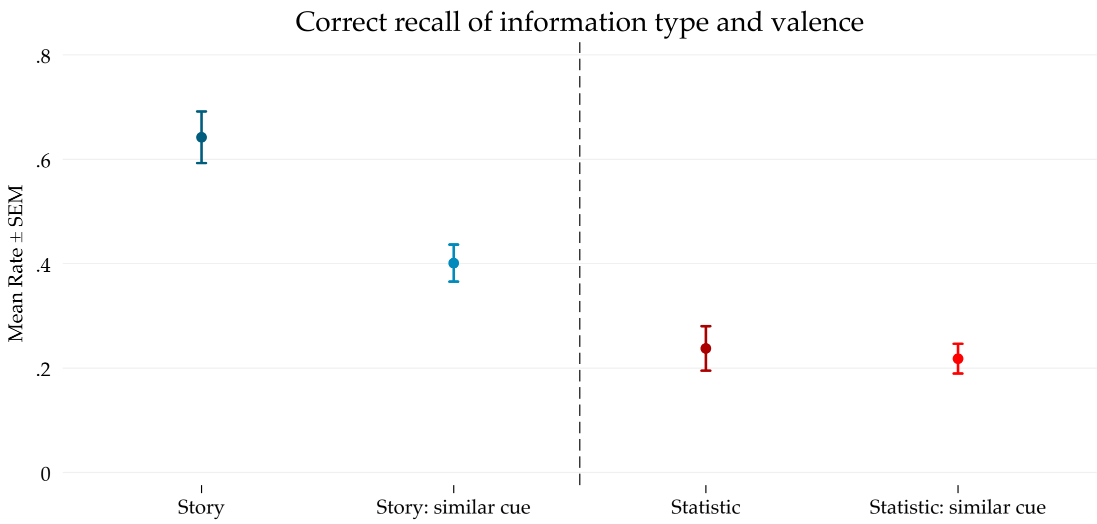


Figure A.4: Change in belief impact and recall in Mechanism Experiment 4 (1,018 respondents). The left panel displays the change in belief impact in percentage points, defined as the difference in belief impact between *Delay* and *Immediate*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. The right panel displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. The dark blue markers illustrate change in belief impact and recall for the *Story* condition, while the light blue markers illustrate the change in belief impact and recall for the *Story with Cue Similarity* condition. The dark red markers illustrate change in belief impact and recall for the *Statistic* condition, while the light red markers illustrate the change in belief impact and recall for the *Statistic with Cue Similarity* condition. Whiskers indicate one standard error of the mean.

B Additional Tables

Table A.1: Overview of data collections

Collection	Sample	Baseline Treatments	Additional Treatments	Main outcomes	Link to pre-analysis plan
Baseline experiments					
Baseline Experiment	Prolific (984 respondents)	3 products: story, statistic, no information	For story treatment 3 different types of contextual features: consistent, neutral, mixed.	Beliefs in immediate and delay; Open-ended recall in delay	https://aspredicted.org/e5mw7.pdf
Robustness Experiment: The role of Decoy Information	Prolific (1,513 respondents)	3 products (1 target and 2 decoy products): Target: Either Story or Statistic	Decoys: Either 2 stories, 2 statistics or 2 times no information	Beliefs in immediate and delay; Structured recall task	https://aspredicted.org/qy3wq.pdf
Mechanisms					
Mechanism Experiment 1: The role of associations	Prolific (666 respondents)	3 products. Decoys: Story and no information; Target varies across treatments	**Baseline condition**: statistic without prompt; **Prompt condition**: statistic with prompt; **No story condition**: Info on a single review without prompt; **No story prompt condition**: Info on a single review with prompt	Beliefs in immediate and delay; Structured recall task	https://aspredicted.org/v9gk7.pdf
Mechanism Experiment 2: Number of product scenarios	Prolific (1,018 respondents)	1 product: Statistic or story; 3 products (statistic, story, no info; 6 products: statistic, story and 4 times no info	None	Beliefs in immediate and delay; Structured recall task	https://aspredicted.org/as7i7.pdf
Mechanism Experiment 3: Story similarity and Cue-story similarity	Prolific (872 respondents)	3 products (bar, cafe and restaurant) with 3 stories	**Baseline**: 3 distinct stories about a bar, a restaurant and a cafe. **Story similarity**: same story about bar as in baseline, but now similar stories about a restaurant and bar. **Cue-story similarity**: As baseline, but the story about the bar is about an experience entirely unrelated and unspecific to a bar.	Beliefs in immediate and delay; Structured recall task	https://aspredicted.org/v7hh6.pdf
Mechanism Experiment 4: Cue similarity	Prolific (583 respondents)	3 products: story, statistic, no information	**Baseline condition**: Restaurant A, Bicycle, Videogame; **Cue similarity condition**: Restaurant A, Restaurant B and Restaurant C	Beliefs in immediate and delay; Structured recall task	https://aspredicted.org/h2fr3.pdf
Mechanism Experiment 5: Question Format and statistic display	Prolific (959 respondents)	3 products: story, statistic, no information	**Likelihood format**: same cue as in the baseline experiment. **Fraction format**: belief elicitation about the percentage of positive reviews **Statistic number display:** Statistical information is provided like in the baseline experiment, i.e. number of positive reviews. **Statistic percent display**: Statistical information is provided in terms of percentages.	Beliefs in immediate and delay; Structured recall task	https://aspredicted.org/ZFF_88V

This Table provides an overview of the different data collections. The sample sizes refer to the final sample of respondents that completed both waves and satisfied the pre-specified inclusion criteria for each of our collections.

Table A.2: Associations and contextual information: belief impact and recall

	Dependent variable:					
	Belief Impact			Combined Recall		
Sample:	Pooled	Stat	NoStory	Pooled	Stat	NoStory
	(1)	(2)	(3)	(4)	(5)	(6)
Delay	-11.5***	-14.7***	-7.95***			
	(0.97)	(1.31)	(1.39)			
Prompt	-0.97	-1.47	1.00	0.20***	0.14***	0.26***
	(1.19)	(1.54)	(1.50)	(0.03)	(0.05)	(0.05)
Delay × Prompt	3.35**	4.22**	1.90			
	(1.34)	(1.93)	(1.83)			
Control Mean	14.47	21.57	6.66	0.19	0.22	0.16
Observations	1332	662	670	1332	662	670
R^2	0.09	0.15	0.06	0.05	0.02	0.08

Notes. OLS estimates, standard errors clustered at the participant level in parentheses. *Delay* is an indicator taking value 1 for respondents in the follow-up survey, and value 0 for respondents in the baseline survey. *Prompt* is an indicator taking value 1 for respondents who were prompted to imagine a typical review when provided with statistical information. All columns pool *Immediate* and *Delay*. Columns (1) and (4) include all respondents. Column (2) and (4) include respondents who received statistics. Columns (3) and (6) include observations who received information on a single review. Columns (1) to (3) display results on belief impact. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. Columns (4) to (6) display the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: The story statistics gap: page time heterogeneity

	Dependent variable:			
	Belief Impact			Recall combined
Sample:	Immediate	Delay	Pooled	Consistent
	(1)	(2)	(3)	(4)
Story	-2.67**	4.78***	-2.67**	0.35***
	(1.32)	(1.51)	(1.32)	(0.04)
Delay			-14.7***	
			(1.11)	
Story × Delay			7.46***	
			(1.56)	
Slow	0.39	-0.34	0.39	0.088**
	(1.17)	(1.39)	(1.17)	(0.04)
Story × Slow	0.60	3.91*	0.60	-0.026
	(1.80)	(2.13)	(1.80)	(0.05)
Delay × Slow			-0.72	
			(1.55)	
Story × Delay × Slow			3.31	
			(2.23)	
Control Mean	20.44	5.76	20.44	0.23
Observations	1168	1168	2336	1168
R^2	0.01	0.04	0.11	0.13

Notes. OLS estimates, standard errors clustered at the participant level in parentheses. *Delay* is an indicator taking value 1 for respondents in the follow-up survey, and value 0 for respondents in the baseline survey. *Story* takes value 1 for respondents who received a story for a given product, and zero otherwise. *Slow* is an indicator taking value 1 for respondents whose response time was above the median in their condition. Columns (1), (2), (3) and (4) include respondents who received consistent stories. Column (3) pools *Immediate* and *Delay*. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. Columns (4) displays the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Question format: belief impact and recall

	Dependent variable:			
	Belief Impact		Combined Recall	
Sample:	Story (1)	Stat (2)	Story (3)	Stat (4)
Similar Format	1.95*	0.53	0.0094	0.085**
	(1.10)	(1.27)	(0.03)	(0.04)
Delay × Similar Format	-0.63	1.45		
	(1.31)	(1.88)		
Statistic Similar		1.98		0.019
		(1.30)		(0.04)
Delay × Statistic Similar		-0.15		
		(1.84)		
Statistic Similar × Similar Format		-1.68		-0.064
		(1.78)		(0.06)
Delay × Statistic Similar × Similar Format		-1.28		
		(2.60)		
Delay	-8.19***	-14.7***		
	(0.87)	(1.33)		
Control Mean	18.50	20.63	0.73	0.19
Observations	1718	1718	859	859
R^2	0.06	0.19	0.00	0.01

Notes. OLS estimates, standard errors clustered at the participant level in parentheses. *Delay* is an indicator taking value 1 for respondents in the follow-up survey, and value 0 for respondents in the baseline survey. *Similar Format* takes value 1 for respondents whose beliefs were elicited in percent. *Statistic Similar* is an indicator taking value 1 for respondents who received statistics in a percentage format. Columns (1) and (3) include respondents who received stories. Columns (2) and (4) include respondents who received statistics. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. Columns (3) and (4) display the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: The story-statistic gap by number of products

	Dependent variable:			
	Belief Impact		Combined Recall	
Sample:	Story (1)	Stat (2)	Story (3)	Stat (4)
1-Product	-1.02 (1.39)	2.26 (1.44)	0.12*** (0.03)	0.33*** (0.04)
Delay × 1-Product	3.76*** (1.44)	7.52*** (1.59)		
6-Products	-1.44 (1.49)	2.76** (1.38)	-0.045 (0.04)	-0.096*** (0.03)
Delay × 6-Products	3.60** (1.68)	-5.13*** (1.76)		
Delay	-9.07*** (1.12)	-13.8*** (1.23)		
Control Mean	18.48	18.51	0.75	0.21
Observations	1562	1515	781	758
R^2	0.04	0.19	0.03	0.16

Notes. OLS estimates, standard errors clustered at the participant level in parentheses. *Delay* is an indicator taking value 1 for respondents in the follow-up survey, and value 0 for respondents in the baseline survey. *1-Product* is an indicator taking value 1 if the respondent receives one product scenario and 0 else. *6-Products* is an indicator taking value 1 if the respondent receives six product scenarios and 0 else. Columns (1) and (3) include respondents who received stories, while column (2) and (4) include respondents who received statistics. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. Columns (3) and (4) display the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Cue similarity

	Dependent variable:			
	Belief Impact		Combined Recall	
Sample:	Story	Stat	Story	Stat
	(1)	(2)	(3)	(4)
Similar Cue	0.21	-0.77	-0.24***	-0.020
	(2.13)	(1.68)	(0.06)	(0.05)
Delay × Similar Cue	-7.75***	-2.56		
	(2.77)	(2.23)		
Delay	-8.30***	-14.3***		
	(2.30)	(1.76)		
Control Mean	18.80	21.62	0.64	0.24
Observations	574	624	287	312
R^2	0.14	0.21	0.05	0.00

Notes. OLS estimates, standard errors clustered at the participant level in parentheses. *Delay* is an indicator taking value 1 for respondents in the follow-up survey, and value 0 for respondents in the baseline survey. *Similar Cue* is an indicator taking value 1 for respondents who received three restaurant scenarios. Columns (1) and (3) include respondents who received stories, while column (2) and (4) include respondents who received statistics. Belief impact is the signed distance between a stated belief and the prior (50%). Belief impact is signed in the direction of the rational update. Columns (3) and (4) display the fraction of respondents correctly recalling the type and valence of information they received in the baseline survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Summary statistics

	Baseline Experiments		Mechanisms				
Experiment:	Baseline	Decoy	Association	Product	Story Sim	Cue Sim	Format
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	0.541	0.506	0.560	0.496	0.506	0.528	0.507
Age (years)	39.782	40.902	39.851	37.351	40.589	36.367	37.090
College	0.611	0.645	0.596	0.619	0.676	0.611	0.626
Employed	0.742	0.785	0.746	0.779	0.771	0.760	0.764
Observations	985	1,548	666	1,018	849	599	922

Notes. Summary statistics. We include all participants who completed both the baseline and the follow-up survey. *Male* is an indicator taking value 1 if the respondent identifies as male and 0 else. *Age* is the respondent's age in years. *College* is an indicator taking value 1 if the respondent holds at least a Bachelor's degree and 0 else. *Employed* is an indicator taking value 1 if the respondent is employed and zero for all other respondents. The columns contain observations from each of the following experiments. Column (1): *Baseline Experiment*. Column (2): *Robustness Experiment: The role of Decoy Information*. Column (3): *Mechanism Experiment 1: The role of associations*. Column (4): *Mechanism Experiment 2: Number of product scenarios*. Column (5): *Mechanism Experiment 3: Story Similarity and Cue-story similarity*. Column (6): *Mechanism Experiment 4: Cue Similarity*. Column (7): *Mechanism Experiment 5: Question Format and statistic display*.

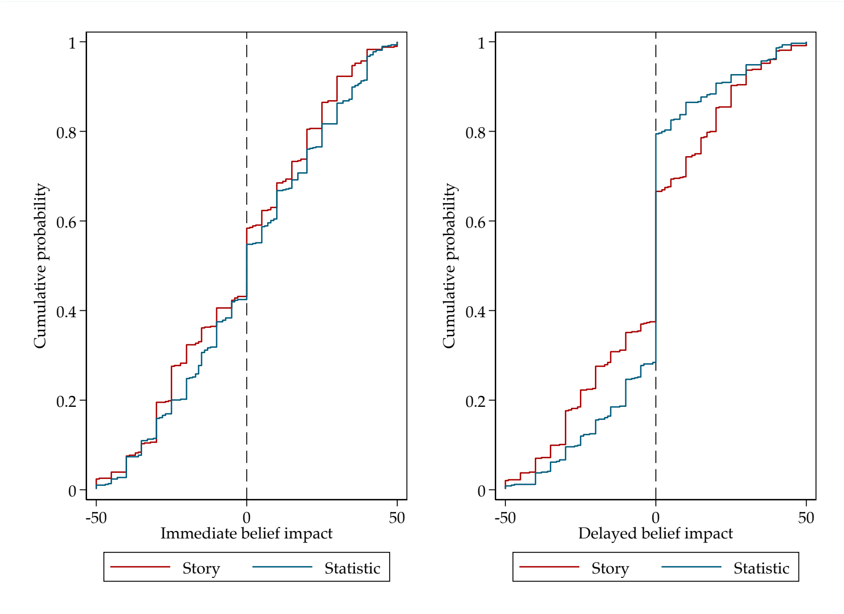
Table A.8: Attrition by conditions

	Dependent variable:						
	Wave 2 Completion						
Experiment:	Baseline	Decoy	Association	Product	Story Sim	Cue Sim	Format
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Neutral Story	0.012						
	(0.03)						
Mixed Story	0.020						
	(0.03)						
Decoy: Story		0.017					
		(0.02)					
Decoy: Statistic		-0.0054					
		(0.02)					
Prompt			0.0065				
			(0.05)				
1-Product				-0.014			
				(0.03)			
6-Products				-0.046			
				(0.03)			
Story Similarity					-0.017		
					(0.03)		
Cue Similarity					-0.019		
					(0.03)		
Similar Cue						0.020	
						(0.03)	
Belief: %							0.016
							(0.03)
Info: %							0.021
							(0.03)
Mean Completed	0.69	0.76	0.46	0.73	0.79	0.59	0.60
Observations	1437	2048	1442	1404	1069	1018	1532
p(Joint Null)	0.80	0.60	0.90	0.37	0.79	0.53	0.59
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes. OLS estimates, standard errors clustered at the participant level in parentheses. *Wave 2 Completion* is an indicator taking value 1 for respondents who completed the follow-up survey, and value 0 who completed the baseline survey only. The columns contain observations from each of the following experiments. Column (1): *Baseline Experiment*. Column (2): *Robustness Experiment: The role of Decoy Information*. Column (3): *Mechanism Experiment 1: The role of associations*. Column (4): *Mechanism Experiment 2: Number of product scenarios*. Column (5): *Mechanism Experiment 3: Story Similarity and Cue-story similarity*. Column (6): *Mechanism Experiment 4: Cue Similarity*. Column (7): *Mechanism Experiment 5: Question Format and statistic display*. The independent variables are indicators for each between-subject condition.

C Additional Figures

Figure A.5: CDFs: belief impact



Notes: Empirical cumulative distribution functions (CDFs) of belief impact in the *Immediate* (left) and *Delay* (right) conditions. Belief impact is the distance between a stated belief and the prior (50%). The data is from the baseline study. Red lines illustrate data from the Story condition, while blue lines illustrate data from the Statistic condition.

D Overview of stories

D.1 Baseline stories

Video games (positive) One of the reviews was randomly selected. The selected review is positive. It is written by 23-year-old Julia, who says she absolutely fell in love with the game. The game called "Planet of Conflict", is a novel concept of a multiplayer role-playing game based on World of Warcraft. Julia was blown away by the realistic graphics. This is the very first time she got totally hooked on a game. Julia mentions that she once played Planet of Conflict for 13 straight hours on a weekend because it was so entertaining. "I communicate with a lot of people online through this game, which I love", Julia says. "Planet of Conflict is just something else entirely. I think I'm a gamer now!"

Video games (negative) One of the reviews was randomly selected. The selected review is negative. It is written by 23-year-old Julia, who says she absolutely hates the game. The game called "Planet of Conflict" is an outdated concept of a multiplayer role-playing game based on World of Warcraft. Julia was disappointed by the pixelated graphics. This is the first time she ever got totally bored by a video game. Julia mentions that she almost fell asleep after the first 30 minutes of playing Planet of Conflict because nothing really happened. "I don't communicate at all with people through this game, which I hate", Julia says. "Planet of Conflict is just something else entirely. I don't think I like gaming anymore after this!"

Video games (mixed) One of the reviews was randomly selected. The selected review is [positive / negative]. It is written by 23-year-old Julia, who says she has mixed feelings about the game. The game called "Planet of Conflict" is a novel concept of a multiplayer role-playing game based on World of Warcraft. Julia was disappointed by the pixelated graphics. However, this is the very first time she got totally hooked on a game. Julia mentions that she once played Planet of Conflict for 13 straight hours on a weekend because it was so entertaining. "At the same time, I don't communicate at all with people through this game, which I hate", Julia says. "Planet of Conflict is just something else entirely. I disliked some parts of the game, but it got me excited about gaming!"

Video games (neutral) One of the reviews was randomly selected. The selected review is [positive / negative]. It is written by 23-year-old Julia. The game called "Planet of Conflict" is a multiplayer role-playing game based on World of Warcraft. Julia's review mentioned the graphics. Julia has played many other games before. Julia mentions that she played Planet of Conflict for a while last weekend. “I sometimes communicate with people through this game”, Julia says. She also stated “Planet of Conflict” is comparable to other video games she has played.

Bicycle (positive) One of the reviews was randomly selected. The selected review is positive. It was provided by Rufus, who is a passionate hobby cyclist. His experience with the bike, a large blue trekking model called “Suburban Racer”, could not have been any better. The bike was delivered after just 4 days. It didn’t require any assembly. The bike is extremely light; riding up his first little hill Rufus felt like he was flying. Rufus mentions that the bike is of exceptional quality. He wrote the report almost 5 years after purchasing it and still hasn’t experienced any problems that required repair. “If you want a worry-free cycling experience, this is the one”, Rufus states.

Bicycle (negative) One of the reviews was randomly selected. The selected review is negative. It was provided by Rufus, who is a passionate hobby cyclist. His experience with the bike, a large blue trekking model called “Suburban Racer”, could not have been any worse. The bike was delivered more than 7 months late. It required 13 hours of assembly work. The bike is extremely heavy; riding up his first little hill Rufus felt like he was crawling. Rufus mentions that the bike is of awful quality. He wrote the report no more than 3 months after purchasing it and has already experienced a number of problems that required expensive repair. “If you want a worry-free cycling experience, definitely go for something else”, Rufus states.

Bicycle (mixed) One of the reviews was randomly selected. The selected review is [positive / negative] . It was provided by Rufus, who is a passionate hobby cyclist. His experience with the bike, a large blue trekking model called “Suburban Racer”, was mixed. The bike was delivered after just 4 days. However, it required 13 hours of assembly work. The bike is extremely light; riding up his first little hill Rufus felt like he was flying. At the same time, Rufus mentions that the bike is of low quality. He wrote the report no more than 3 months after purchasing it and has already experienced a number of problems that required expensive repair. “If you want a worry-free cycling experience, not sure this is the right bike for you”, Rufus states.

Bicycle (neutral) One of the reviews was randomly selected. The selected review is [positive / negative] . It was provided by Rufus, who is a hobby cyclist. He describes his experience with the bike, a large blue trekking model called “Suburban Racer”. The bike was delivered around the time predicted by the manufacturer. It required some assembly work. The bike has a typical weight compared to other bikes. Rufus' review described the quality of the bike. He wrote the report a while after purchasing it and has made some repairs in the meantime.

Restaurant (positive) One of the reviews was randomly selected. The selected review is positive. It was provided by Justin. He and his friend had a wonderful experience at the Japanese restaurant called "Sushi4Ever". They ordered the sushi taster. The raw fish looked fresh and all sushi was expertly prepared. Justin was impressed by the authentic taste that reminded him of his holiday in Japan. The service was exquisite: his waiter was polite, highly attentive and the food was served promptly. After Justin had paid, the waiter served a traditional Japanese drink on the house that Justin had never heard of before and loved. As they left the restaurant, Justin was very happy and thought to himself "I'll be back!"

Restaurant (negative) One of the reviews was randomly selected. The selected review is negative. It was provided by Justin. He and his friend had an awful experience at the Japanese restaurant called "Sushi4Ever". They ordered the sushi taster. The raw fish looked stale and the sushi rolls were falling apart on the plate. Justin was disappointed by the Western taste that was very different from what he remembered from his holiday in Japan. The service was poor: his waiter was rude, not attentive and the food was served after a long wait. After Justin had paid, the waiter insisted on them leaving their table immediately. As they left the restaurant, Justin was very annoyed and thought to himself "I definitely won't be back!"

Restaurant (mixed) One of the reviews was randomly selected. The selected review is [positive / negative] . It was provided by Justin. He and his friend had a mixed experience at the Japanese restaurant called "Sushi4Ever". They ordered the sushi taster. The raw fish looked fresh and all sushi was expertly prepared. Justin was impressed by the authentic taste that reminded him of his holiday in Japan. The service, however, was poor: his waiter was rude, not attentive and the food was served after a long wait. After Justin had paid, the waiter insisted on them leaving their table immediately. As they left the restaurant, Justin was conflicted and thought to himself "Not sure whether I'll go again."

Restaurant (neutral) One of the reviews was randomly selected. The selected review is [positive / negative]. It was provided by Justin. Justin and his friend describe their experience at the Japanese restaurant called "Sushi4Ever". They ordered the sushi taster. The menu included raw fish and a variety of sushi rolls. Justins' review describes the taste of the sushi. He mentions the service, writes about how attentive the waiter was and how long they had to wait for the food. After Justin had paid, the waiter served a traditional Japanese drink. As they left the restaurant, Justin thought about whether he would come back to the restaurant or not.

D.2 Mechanism Experiment: Story similarity

Baseline condition

Bar One of the reviews was randomly selected. The selected review is positive. It was provided by David, who most of all cares about the interior. He mentions that the interior of the place was outstanding. He describes a luxurious, spacious layout with a modern feel yet cozy atmosphere. "Entering this place will improve your mood immediately!" The second thing David really cares about is the view. According to David, the cherry on the cake is a breath-taking view from this rooftop location on the 51st floor. A majestic look over the entire city completes this phenomenal place that David describes as offering the "best overall vibe of the city".

Restaurant One of the reviews was randomly selected. The selected review is negative. It was provided by Justin, who most of all cares about the quality of the food. He and his friend had an awful experience at the Japanese restaurant called "Sushi4Ever". They ordered the sushi taster. The raw fish looked stale and the sushi rolls were falling apart on the plate. The second thing Justin really cares about is how authentic the food is. Justin was disappointed by the Western taste that was very different from what he remembered from his holiday in Japan. As they left the restaurant, Justin was very annoyed and thought to himself "I definitely won't be back!"

Cafe One of the reviews was randomly selected. The selected review is negative. It was provided by Linda, who most of all cares about the service quality. She complained that the service quality was incredibly poor. Nobody initially showed her to a table so she stood in the entrance for a full 10 minutes. Even though there were few customers, the waiters all seemed stressed and were rude to her. The waiter spilled hot coffee over Linda's pants. The second thing Linda really cares about are waiting times. Because the waiter brought the wrong food, Linda had to wait another half hour. The waiter did not apologize. Linda describes the service in the cafe as the disappointment of a lifetime and was fuming with rage as she left the cafe.

Story similarity condition

Bar Same as in baseline condition

Restaurant One of the reviews was randomly selected. The selected review is negative. It was provided by Justin, who most of all cares about the interior. He mentions that the interior of the place was poor. He describes a worn-down, claustrophobic space with an outdated feel and depressing atmosphere. “Entering this place will kill your mood immediately!” The second thing Justin really cares about is the view. According to Justin, what adds insult to injury is the practically non-existent view from this basement location. The lack of daylight completes this disappointing place that Justin describes as the “worst vibe you can possibly get in this city”.

Cafe One of the reviews was randomly selected. The selected review is negative. It was provided by Linda, who most of all cares about the interior. She mentions that the interior of the place was disappointing. She mentions a time-worn, carelessly put together furnishing that did not look clean and was slightly smelly. “Coming here will make you want to leave immediately!” The second thing Linda really cares about is the view. According to Linda, what made matters worse is the absence of any windows and the glaring fluorescent lighting. The absence of natural light completes this frustrating venue that Linda describes as the “most dismal vibe in the area”.

Cue-story similarity condition

Bar One of the reviews was randomly selected. The selected review is positive. It is written by 34-year-old John. John had a fantastic experience going shopping for clothes on a Saturday a few weeks ago. He intended to buy only a new pair of shoes but ended up buying also a pair of pants and a sweater, all of which have since become his favorite pieces. The store he wanted to go to was closed so he went to a different store that he had not previously been to, and the clothes they had blew him away. He tried on a number of different styles and sizes because he directly fell in love with various outfits sold in the store. He spent about one hour in the store, but would have loved to stay even longer. Afterwards, he celebrated this wonderful shopping experience at the new store, wandering around in the area all afternoon.

Restaurant Same as in baseline condition.

Cafe Same as in baseline condition.

E Implementation Details on the Experiments

Randomization

In the baseline survey, the randomization is implemented by drawing true fractions of positive reviews for the videogame, the restaurant and the bicycle i.i.d. uniformly over [0,1]. The total number of reviews is always fixed at 14, 19 and 17 respectively. The lowest fraction is then assigned a "negative" signal valence, while the highest is given a “positive” valence. The product with the median fraction is assigned to the “no information” treatment, which doesn’t have a valence. Finally, the type of signal for the two other products is drawn by assigning “story” and “statistic” or “statistic” and “story” to the lowest and highest respectively, each with probability 1/2.

For the product with the “story” signal, the review is either “consistent”, “mixed” or “neutral” (cf. Section 2.3) with probabilities 0.6, 0.2 and 0.2. For the “statistic" signal, a signal fraction is drawn as $s \sim \mathscr{U}[0, 0.5]$ if the valence is negative and $s \sim \mathscr{U}[0.5, 1]$ if it is positive. Since the signal is indicated as “out of b randomly drawn reviews, a are positive”, we chose a and b to minimize $|a/b - s|$, with a integer and $b \in \{4, 5, 6, 7, 8, 9, 10, 11\}$. In case of ties, we favor lower denominators to increase variability. Moreover, we impose that $a/b < 0.5$ or $a/b > 0.5$ depending on the valence.

F Coding Manual for data on open-ended recall

Free-form responses are provided together with subject identifier and information on the product and the type of information received (story, statistic or no info, plus whether the info was positive or negative) in an Excel sheet. All of the below should be coded as binary variables, 1 for presence of a phenomenon in the text and blank for its absence. People may express uncertainty “maybe”, “could be”. Always count this as if people would be stating the same statement with certainty.

Table A.9: Coding Manual for data on open-ended recall

Category	Explanation	Examples
Lack of memory	Statement that participants do not recall whether and what information they received. This includes instances in which a participant remembers the product, but not whether and what information they received. This does not include statements like "I remember that I received no additional information" or "I don't think I received any additional information about the bicycle" when they actually received no info. Sometimes, it may be hard to distinguish between participants indicating "they donât remember" and âthey remember getting no additional informationâ, e.g., when just stating âNoneâ. It can help looking at the subjectâs two other responses.	"I do not have any recollection about this product/scenario." "I cannot remember anything"
Mention type of information	They mention whether they received a single review, multiple reviews or no information.	"For this product I received no additional information." "I received information on multiple reviews" "There was one review about the videogame. [Details about the review...]"
Misremember type of information	State that received a different type of information than they truly did.	"I received information on a number of reviews." [When in reality, they received a story about a single review]
Mention valence	Response indicates positive or negative tendency. This can be about the majority of reviews being pos/neg, a single review categorized as positive/negative, or about the implicit valence of qualitative features without saying positive/negative.	"The information was mostly positive." "The review was negative." "The bike was of high quality."
Misremember valence	State that information was positive (negative), when it was really negative (positive). This does not include misremembering the exact number of positive reviews of a statistic, as long as the remembered number points in the same direction (positive/negative) as the true one.	"The information was mostly positive." [When the actual information provided was a majority of negative reviews]
Confusion	Answer exclusively talks about things unrelated to the scenario in question, e.g., repeating general instructions, talking about the task in general terms, or talk about what they remember for a different scenario.	
Recall stat correctly	Statements of specific numbers of positive reviews, or total reviews received. Only indicate this if the remembered numbers are correct!	"Out of the 11 sampled reviews 2 were positive and 9 were negative."
Mention qual. factors	Mention specific qualitative elements from a story. This needs to be specific, i.e., does not include âI remember reading information about a personâs review which was really positive.â	"I think they took the bicycle out on hilly terrain, or on some sort of holiday or outing."
Mention first	This is only about a specific order: Mention specific qualitative factors before indicating anything else, such as the valence of the overall review (i.e. whether the review is positive or negative).	"The review selected was from a person that had the bike for 5 years and still thought it worked perfectly. The bike came already assembled. The review selected was a positive review."
Recall immediate belief	Mentions the belief that subject thinks they indicated on the prior day. Indicate independently of whether it is correct.	"In this one, I wrote 85% because it gave a positive review."
Full confusion	Answer exclusively talks about things unrelated to the scenario in question, e.g., repeating general instructions, talking about the task in general terms, or talking about what they remember for a different scenario.	
Misremembering across scenarios	Each participant gave three responses that are in adjacent rows in the Excel file. This category should be coded if the subjectâs response talks about information that is in line with what they received in a different scenario.	Assume the subject got no info for the bicycle, but a positive story for the restaurant, but states the following for the bicycle: "I remember reading about a positive review about the bicycle."
Flag for misc. or uncertain coding	Indicate this if the response includes something distinctive (meaningful) that is not covered by our criteria, or if you are uncertain about your coding I do remember that the first one didnât give much if any information, the second one gave a little more and the third I think gave a little more again.	

This Table provides an overview of the coding scheme. The examples are all taken from the baseline experiment.

G Conceptual Framework

In the following, we formally describe the conceptual framework outlined in Section 3 and derive the corresponding behavioral predictions. This model of episodic memory builds on Bordalo et al. (2023, 2022).

G.1 Notation

Memory traces. Episodic memory traces are encoded as vectors of $F > 1$ features with values in $V_1, \dots, V_F$. All sets of values V_f contain the null element 0, indicating the absence of a feature. For example, a statistic about a bicycle contains the product type and the value of the statistic, while a story about a restaurant contains the product type, the value of the review and additional anecdotal features. Then, these memory traces might be represented as:

$$
\begin{aligned}
m_{bicycle} &= \textit{(bicycle, 3 out of 7 reviews were positive, 0, 0, } \dots \textit{)} \\
m_{restaurant} &= \textit{(restaurant, 1 out of 1 review is positive, Justin and his friends had} \\
&\quad \textit{a wonderful experience, they ordered the sushi taster, } \dots \textit{)}
\end{aligned}
$$

Cued set. The cued set $C_p = \{m_p\}$ consists of a single memory trace which encodes the experience of product scenario p. In the extension in Section H.2, the cued set C_p contains several memories. In a further extension in Section 4.3, we allow the cue to be distinct from the target trace; this provides a way to account for the inherent similarity between cue and target.

Sets of memories. We call M the set of all memories in the memory database and E the set of memories created during the experiment. $\overline{E} := M \setminus E$ is the set of memories from outside the experiment, and memories of non-cued products are denoted by $C_{-p} := E \setminus C_p$. Note that C_p, C_{-p} and $\overline{E}$ form a partition of M.

Similarity. Following Bordalo et al. (2022), we introduce a similarity measure S that hinges on the presence of features, only requiring it to have the minimal properties:

Assumption 1. *Memory trace similarity* $S(m_1, m_2) : M \times M \to [0, 1]$ *is (i) symmetric, (ii) increasing in the number of shared features, i.e. features present in both, (iii) decreasing in the number of non-shared features, i.e. features present in one but not the other, (iv) maximal when* $m_1 = m_2$.

We will only rely on these axiomatic properties, but a simple illustration would be:

$$S(m_1, m_2) = \frac{\text{\# of shared features}}{\text{\# of total features}}$$

By shared features, we mean features that are non-null in both traces, not features that have the same value in both. By total features, we mean features that are present in either, i.e. the sum of non-null features in both minus the number of shared features. Although this illustration is discrete, we will often treat similarity as continuous. Indeed, our goal is to remain relatively agnostic on the precise form of trace similarity, and we will think about similarity in more general terms since it is more accessible to intuition.

We extend similarity to sets $M_1, M_2 \subset M$ as:

$$S(m_1, M_2) := \frac{1}{|M_2|} \sum_{m_2 \in M_2} S(m_1, m_2) \qquad S(M_1, M_2) := \frac{1}{|M_1|} \frac{1}{|M_2|} \sum_{m_1 \in M_1} \sum_{m_2 \in M_2} S(m_1, m_2)$$

Recall. We posit that the probability of recalling a memory trace m when being cued for the cued set C is:

$$r(m, C) := \frac{S(m, C)}{\sum_{m' \in M} S(m', C)}, \tag{3}$$

so that $r(\cdot, C)$ defines a probability measure over M. Since $S(C_p, C_p) = 1$, the probability of recalling the cued set C_p when being cued for exactly this set is:

$$r(C_p) := r(C_p, C_p) = \frac{1}{\sum_{m \in M} S(m, C_p)} \tag{4}$$

Using the partition of M, this can be rewritten more intelligibly as:

$$r(C_p) = \frac{1}{1 + |C_{-p}|\, S(C_{-p}, C_p) + |\overline{E}|\, S(\overline{E}, C_p)} \tag{5}$$

We refer to $S(C_{-p}, C_p)$ as cross-similarity, and to $S(\overline{E}, C_p)$ as out-similarity.

G.2 Assumptions

We next state two Assumptions and derive a Lemma that will justify our main results.

Assumption 2. *Each product-scenario creates a single memory trace $C_p = \{m_p\}$.*

This Assumption is mainly made for clarity of exposition, and is relaxed in Section H.2 which allows for multiple memory traces.

Assumption 3. *Stories are less similar to out-of-experiment memories than statistics:*

$$S(C_p^{story}, \overline{E}) < S(C_p^{stat}, \overline{E})$$

Since we want to remain agnostic on out-of-experiment traces $\overline{E}$, we cannot establish Assumption 3 formally in our framework, but it is highly intuitive: the greater number of features in story traces make them less similar to out-of-experiment memories. Put simply, in- and out-of-experiment stories are very different, while in- and out-of-experiment statistics are more easily confused.

The following Lemma establishes an analogous property for cross-similarity, which can be proven using the properties of similarity.

Lemma 1. *Stories are less similar to non-cued scenarios than statistics:*

$$S(C_p^{story}, C_{-p}^{story}) < S(C_p^{stat}, C_{-p}^{stat})$$

Proof. In the *Baseline* experiment, there is one statistic, one story and one no-info product. Under Assumption 2, they respectively leave memories m^{stat}, m^{story} and m^{noinfo}. Since m^{noinfo} only contains the product name, the number of shared features is always identical; but stories contain more non-shared features than statistics, so that by the properties of similarity in Assumption 1, $S(m^{story}, m^{noinfo}) < S(m^{stat}, m^{noinfo})$. Therefore:

$$\begin{aligned} S(C_p^{story}, C_{-p}^{story}) &= \frac{1}{2}\left(S(m^{story}, m^{noinfo}) + S(m^{story}, m^{stat})\right) \\ &< \frac{1}{2}\left(S(m^{stat}, m^{noinfo}) + S(m^{story}, m^{stat})\right) = S(C_p^{stat}, C_{-p}^{stat}) \end{aligned}$$

□

G.3 Proofs

Proposition 1. *Stories are more likely to be recalled than statistics:*

$$r(C_p^{stat}) < r(C_p^{story}) \tag{6}$$

Proof. Assumption 2 and Lemma 1 imply:

$$|C_{-p}^{story}|\, S(C_p^{story}, C_{-p}^{story}) < |C_{-p}^{stat}|\, S(C_p^{stat}, C_{-p}^{stat}) \tag{7}$$

Since $|\overline{E}|$ is unaffected by the information type, Assumption 3 implies:

$$|\overline{E}|\, S(C_p^{story}, \overline{E}) < |\overline{E}|\, S(C_p^{stat}, \overline{E}) \tag{8}$$

Together with (5), these yield:

$$
\begin{aligned}
r(C_p^{stat}) &= \frac{1}{1+|C_{-p}^{stat}|\, S(C_{-p}^{stat}, C_p^{stat}) + |\overline{E}|\, S(\overline{E}, C_p^{stat})} \\
&< \frac{1}{1+|C_{-p}^{story}|\, S(C_{-p}^{story}, C_p^{story}) + |\overline{E}|\, S(\overline{E}, C_p^{story})} = r(C_p^{story})
\end{aligned}
\tag{9}
$$

□

Corollary 1.1. *Adding features to the target scenario, e.g. by prompting respondents to imagine a potential review, increases recall.*

Proof. Increasing the number of non-null features in the cued set C_p decreases cross-similarity $S(C_{-p}, C_p)$ and out-similarity $S(\overline{E}, C_p)$. (5) then yields the conclusion. □

Corollary 1.2. *Keeping similarity to out-of-experiment memories constant, increasing the similarity of stories decreases recall.*

Proof. Increasing the similarity of stories means increasing $S(C_{-p}^{story}, C_p^{story})$. Since $S(\overline{E}, C_p^{story})$ stays constant, (5) implies the conclusion. □

We now turn to the effect of decoy scenarios.

Proposition 2. *Adding decoy scenarios decreases recall.*

Proof. Adding a decoy scenario to the non-cued set C_{-p} increases total non-cued similarity $|C_{-p}|\, S(C_{-p}, C_p)$, which per (5) decreases the probability of recall. □

Corollary 2.1. *Keeping the average similarity $S(C_{-p}, C_p)$ constant, adding decoys increases recall by:*

$$
\frac{\partial r(C_p)}{\partial |C_{-p}|} = -S(C_{-p}, C_p)\, r(C_p)^2 < 0 \tag{10}
$$

Note that here we tolerate an abuse of notation as $|C_{-p}|$ is discrete.

Proof.

$$
\frac{\partial r(C_p)}{\partial |C_{-p}|} = \frac{-S(C_{-p}, C_p)}{(1+|C_{-p}|\, S(C_{-p}, C_p) + |\overline{E}|\, S(\overline{E}, C_p))^2} = -S(C_{-p}, C_p)\, r(C_p)^2 < 0
$$

□

Corollary 2.2. *Keeping the number of decoys $|C_{-p}|$ constant, increasing the similarity of decoys increases recall by:*

$$
\frac{\partial r(C_p)}{\partial S(C_{-p}, C_p)} = -|C_{-p}|\, r(C_p)^2 < 0 \tag{11}
$$

Proof.

$$\frac{\partial r(C_p)}{\partial S(C_{-p}, C_p)} = \frac{-|C_{-p}|}{(1+|C_{-p}|\, S(C_{-p}, C_p) + |\overline{E}|\, S(\overline{E}, C_p))^2} = -|C_{-p}|\, r(C_p)^2 < 0$$

□

Corollary 2.3. *The effect of adding a decoy is highest for the information type with the highest similarity to the information type of the decoy scenario.*

Proof. Having higher similarity to the decoy means that the marginal effect from adding a decoy is larger, per Corollary 2.1. □

We would like to compare the effect of adding decoys on the recall gap between stories and statistics,

$$g := r(C_p^{story}) - r(C_p^{stat}). \tag{12}$$

Two countervailing forces are at play: statistics have a higher similarity to decoys, but stories have a higher initial recall. Even if the effect initially widens the story-statistic gap, eventually recall for statistics becomes so low that the effect on the gap reverses. This is intuitive: after an extreme accumulation of statistics and stories, overall recall becomes so low that the gap between the two can only narrow. Our experimental evidence suggests that, indeed, the effect on the gap is initially positive.

Proposition 3. *Keeping similarities constant, adding decoys potentially first increases the story-statistic recall gap g and always decreases it eventually.*

Proof. Corollary 2.1 implies

$$\frac{\partial g}{\partial |C_{-p}|} = -S(C_{-p}^{story}, C_p^{story}) r(C_p^{story})^2 + S(C_{-p}^{stat}, C_p^{stat}) r(C_p^{stat})^2, \tag{13}$$

which is positive when:

$$\frac{S(C_{-p}^{story}, C_p^{story})}{S(C_{-p}^{stat}, C_p^{stat})} < \frac{r(C_p^{stat})^2}{r(C_p^{story})^2} = \left(\frac{1+|C_{-p}^{story}|\, S(C_{-p}^{story}, C_p^{story}) + |\overline{E}|\, S(C_p^{story}, \overline{E})}{1+|C_{-p}^{stat}|\, S(C_{-p}^{stat}, C_p^{stat}) + |\overline{E}|\, S(C_p^{stat}, \overline{E})}\right)^2$$

To simplify, we use notations $x := |C_{-p}|, A^{stat} := S(C_{-p}^{stat}, C_p^{stat}), A^{story} := S(C_{-p}^{story}, C_p^{story})$, $B^{stat} := 1 + |\overline{E}|\, S(C_p^{stat}, \overline{E})$ and $B^{story} := 1 + |\overline{E}|\, S(C_p^{story}, \overline{E})$. Note that $0 < A^{story} < A^{stat}$ and $0 < B^{story} < B^{stat}$. Then, (13) has the same sign as:

$$A^{stat} A^{story} (A^{story} - A^{stat}) x^2 + 2 A^{stat} A^{story} (B^{story} - B^{stat}) x + A^{stat}\, (B^{story})^2 - A^{story}\, (B^{stat})^2$$

Since the quadratic and linear terms are negative, $\frac{\partial g}{\partial |C_{-p}|}$ is negative for x high enough, i.e. $|C_{-p}|$ high enough. Depending on the sign of $A^{stat}(B^{story})^2 - A^{story}(B^{stat})^2$, the effect on g might initially be positive before becoming negative. □

G.4 Beliefs

We model beliefs in two periods. Before the first period, respondents have uniform priors. In the first period, they (potentially) receive additional information on a product and form Bayesian beliefs. In the second period, participants are asked to recall the first-period information. With probability $r(C_p)$, they recall the correct memory trace and again form Bayesian beliefs. With probability $1-r(C_p)$, they recall an incorrect memory trace: we assume there is no confusion, i.e. they recognize it as incorrect and therefore state beliefs corresponding to the prior.

Notation. For a given product p, we call the total number of reviews N, the total number of positive reviews K, the number of observed reviews n and the number of observed positive reviews k. Participants are asked about the probability $\pi := K/N$ of a randomly drawn review being positive. The distribution of N and n has no effect as both are drawn before the experiment. Since we say that the qualitative elements of stories do not convey any inherent information, to a Bayesian a story is simply a statistic with $n = 1$.

Prior beliefs. Respondents are informed that the number of positive reviews is uniformly distributed, so that their prior is:

$$K \sim \mathcal{U}[\![0, N]\!] = \text{BetaBinomial}(N, 1, 1) \tag{14}$$

Indeed, a uniform distribution over $[\![0, N]\!]$ is identical to a Beta-Binomial distribution with parameters N and $\alpha = \beta = 1$.

Prior beliefs under no-recall. When they recall the wrong memory trace, respondents understand that it is the wrong trace and that they do not have any additional information. Payoff is then maximized by reporting the mean of the prior, which is:

$$\hat{\pi}_2^{\textit{no-recall}} = \mathbb{E}_{prior}[\pi] = \frac{1}{2}$$

Bayesian beliefs under recall. When they recall the memory trace, respondents remember that they saw k positive reviews out of n, drawn without replacement from N

total reviews, so that the signal follows the hypergeometric conditional distribution:

$$k|K \sim \text{HyperGeometric}(N, K, n) \tag{15}$$

As Beta-Binomial and hypergeometric distributions are conjugate priors, beliefs about the remaining reviews follow a Beta-Binomial distribution with parameters $N-n$, $\alpha' := \alpha + k = 1 + k$ and $\beta' := \beta + n - k = 1 + n - k$:

$$K - k|k \sim \text{BetaBinomial}(N - n, 1 + k, 1 + n - k) \tag{16}$$

Note that the average of this distribution is $(N-n)\frac{\alpha'}{\alpha'+\beta'} = (N-n)\frac{k+1}{n+2}$. The payoff is then maximized by reporting the mean of the belief distribution, which is:

$$\hat{\pi}_2^{recall} = \mathbb{E}_{posterior}[\pi] = \frac{k}{N} + \frac{N-n}{N}\frac{k+1}{n+2}.$$

Recall and belief decay. We are interested in belief decay, i.e. the difference in beliefs between the first and the second period, which we denote $\hat{\pi}_1$ and $\hat{\pi}_2$. Under our model, conditional period 1, $\mathbb{E}(\hat{\pi}_2) = r(C_p)\hat{\pi}_1 + \big(1 - r(C_p)\big)\frac{1}{2}$. The key observation is then that the behavioral belief impact is the rational belief impact scaled down by recall:

$$\mathbb{E}\left(\hat{\pi}_2 - \frac{1}{2}\right) = r(C_p)\hat{\pi}_1 + \big(1 - r(C_p)\big)\frac{1}{2} - \frac{1}{2} = r(C_p)\left(\hat{\pi}_1 - \frac{1}{2}\right) \tag{17}$$

Therefore, belief decay is exactly recall. Our predictions on recall map straightforwardly onto predictions on beliefs.

H Model Extensions

H.1 Cue-target similarity

Recall based on an external cue. So far, for simplicity, we have assumed that respondents query their memory using the target set C_p as a cue. In this extension we consider in a more explicit way that the prompt in the second-wave survey plays the role of a cue. As a consequence, target and cue are no longer identical so that their (dis)similarity plays an important role.

We formalize this intuition by also representing the cue as a trace. More precisely, we now represent traces as a vector of $F+1$ features, with values in $V_1, \ldots, V_{F+1}$. The F first entries are the same as before, and the last one is simply the question being asked, e.g.

$$c_{bicycle} = (\textit{bicycle, \ldots, What is the probability of a random review being positive?})$$

Similarity between the cue and the target will then hinge on all the intermediary features. More precisely, we imagine that, beyond simply containing features *directly* contained in a story, a statistic or a cue, traces will encode *higher-order* associations as features. This relates to the broader concept of semantic memory (Kahana, 2012), i.e. general knowledge about concepts and categorical relationships between concepts.

For instance, a story about a *restaurant* might contain the anecdote "*Justin and his friends ordered the sushi taster.*", which might trigger the presence of higher-level features like *food*, *friends* and *sushi*. Conversely, a cue about a *restaurant* might elicit associations with *food* and *friends*. The similarity between the cue and the target memory of a product episode would then be determined by the similarity of their higher-order associations.[4]

In this more general framework, denoting c_p the cue and C_p the cued i.e. target set, recall depends on similarity of the two:

$$r(C_p, c_p) = \frac{|C_p|\, S(C_p, c_p)}{|C_p|\, S(C_p, c_p) + |C_{-p}|\, S(C_{-p}, c_p) + |\overline{E}|\, S(\overline{E}, c_p)} \tag{18}$$

Cue-statistic similarity. In the *Cue-Statistic Similarity* experiment, statistics were presented as counts or as percentages, while beliefs were elicited as probabilities or as percentages. We argue that recall is higher when the initial information and the elicitation are of the same type, because they will tend to trigger more similar higher-level associations in the memory trace.

Assumption 4. *Statistics presented and elicited in the same format are more similar than when they are in a different format, i.e.* $S(C_p^{pct}, c_p^{pct})$ *and* $S(C_p^{prob}, c_p^{prob})$ *are greater than* $S(C_p^{pct}, c_p^{prob})$ *and* $S(C_p^{pct}, c_p^{prob})$*.*

Assumption 5. *The elicitation format has no effect on cross-similarity* $S(C_{-p}, c_p)$ *and out-similarity* $S(\overline{E}, c_p)$*.*

Proposition 4. *Eliciting beliefs in a format closer to the initial presentation of statistics increases recall.*

[4]An alternative modeling approach might be iterative sampling. In this model, when a participant is given a cue, they undergo a repetitive process of accessing their memory database. At each iteration, retrieval probabilities are determined by the degree of similarity to the sample from the previous step, continuing until the target is retrieved or after a predetermined number of steps. For instance, when provided with the cue *restaurant*, a participant might initially access an irrelevant experience associated with *food*, then another associated with *sushi*, before arriving at the relevant experience. The set of all memory traces and similarities between them will then drive cue-target similarity. We chose to use semantic memory because it is closer to our baseline model, and therefore more comparable.

Proof. Assumption 4 implies that $S(C_p^{pct}, c_p^{pct})$ and $S(C_p^{prob}, c_p^{prob})$ are greater than $S(C_p^{pct}, c_p^{prob})$ and $S(C_p^{pct}, c_p^{prob})$. Changing the elicitation has no effect on $S(C_{-p}, c_p)$ and $S(\overline{E}, c_p)$, so:

$$\begin{aligned} r(C_p^{pct}, c_p^{prob}) &= \frac{|C_p^{pct}|\ S(C_p^{pct}, c_p^{prob})}{|C_p^{pct}|\ S(C_p^{pct}, c_p^{prob}) + |C_{-p}^{pct}|\ S(C_{-p}^{pct}, c_p^{prob}) + |\overline{E}|\ S(\overline{E}, c_p^{prob})} \\ &< \frac{|C_p^{pct}|\ S(C_p^{pct}, c_p^{pct})}{|C_p^{pct}|\ S(C_p^{pct}, c_p^{pct}) + |C_{-p}^{pct}|\ S(C_{-p}^{pct}, c_p^{pct}) + |\overline{E}|\ S(\overline{E}, c_p^{pct})} = r(C_p^{pct}, c_p^{pct}), \end{aligned}$$

along with the analogous results for the other combinations. □

Cue-story similarity. Similarity between stories and the cue is driven by higher-order associations from the anecdotes contained in the story. The *Cue-Story Similarity* experiment replaces the story with an entirely unrelated story, thus lowering the similarity of the cue with the story memory, in turn lowering recall.

Assumption 6. *Telling a story that is unrelated to the target product has no effect on cross-similarity $S(C_{-p}, c_p)$ and out-similarity $S(\overline{E}, c_p)$.*

Proposition 5. *Telling a story that is unrelated to the target product lowers recall.*

Proof. Since $S(C_p^{unrelated}, c_p)$ is lower than $S(C_p^{related}, c_p)$ by construction, while similarity of the cue with non-target memories and out-of-experiment memories remains the same:

$$\begin{aligned} r(C_p^{unrelated}, c_p) &= \frac{|C_p^{unrelated}|\ S(C_p^{unrelated}, c_p)}{|C_p^{unrelated}|\ S(C_p^{unrelated}, c_p) + |C_{-p}|\ S(C_{-p}^{unrelated}, c_p) + |\overline{E}|\ S(\overline{E}, c_p)} \\ &< \frac{|C_p^{related}|\ S(C_p^{related}, c_p)}{|C_p^{related}|\ S(C_p^{related}, c_p) + |C_{-p}|\ S(C_{-p}^{related}, c_p) + |\overline{E}|\ S(\overline{E}, c_p)} = r(C_p^{related}, c_p) \end{aligned}$$

□

These derivations also give a hint as to why the effect of cue-target similarity uncovered in the experiments is not very large: it seems that changes in cue-target similarity, e.g. from $S(C_p^{prob}, c_p^{prob})$ to $S(C_p^{pct}, c_p^{prob})$ for statistics or from $S(C_p^{related}, c_p)$ to $S(C_p^{unrelated}, c_p)$ for stories, are not very large compared to the magnitude of cue-non-target similarity $|C_{-p}|S(C_{-p}, c_p)$, which drives effects in the other experiments.

H.2 Multiple target traces

Multiple episodic memories per product. We extend the model by allowing each product scenario to leave multiple memory traces, so that the cued set becomes e.g.

$C_p = \{m_p^1, m_p^2, m_p^3, \dots\}$. Recall then involves a self-similarity term:

$$r(C_p) = \frac{|C_p|\, S(C_p, C_p)}{|C_p|\, S(C_p, C_p) + |C_{-p}|\, S(C_{-p}, C_p) + |\overline{E}|\, S(\overline{E}, C_p)} \tag{19}$$

We maintain Assumption 3. Since the size of the sets C_p and C_{-p} are now unknown, we need to restate Lemma 1 in terms of total similarity, along with an Assumption for self-similarity.

Assumption 7. *Stories have lower total cross-similarity than statistics:*

$$|C_{-p}^{story}|\, S(C_{-p}^{story}, C_p^{story}) < |C_{-p}^{stat}|\, S(C_{-p}^{stat}, C_p^{stat}) \tag{20}$$

Assumption 8. *Stories have higher total self-similarity than statistics:*

$$|C_p^{stat}|\, S(C_p^{stat}, C_p^{stat}) < |C_p^{story}|\, S(C_p^{story}, C_p^{story}) \tag{21}$$

Both Assumptions seem intuitive: stories are more specific so will be more similar between themselves and less similar to other memories than statistics. For self-similarity, there are countervailing forces at play: stories seem likely to leave more memory traces, so that $|C_p^{stat}| < |C_p^{story}|$, but statistics are simpler and might therefore be more similar. Assumption 8 asserts the net effect favors stories.

Proposition 6. *With multiple memory traces, stories are more likely to be recalled than statistics:*

$$r(C_p^{stat}) < r(C_p^{story}). \tag{22}$$

Proof. Assumptions 3, 7 and 8, along with (19), yield:

$$\begin{aligned} r(C_p^{stat}) &= \frac{|C_p^{stat}|\, S(C_p^{stat}, C_p^{stat})}{|C_p^{stat}|\, S(C_p^{stat}, C_p^{stat}) + |C_{-p}^{stat}|\, S(C_{-p}^{stat}, C_p^{stat}) + |\overline{E}|\, S(\overline{E}, C_p^{stat})} \\ &< \frac{|C_p^{story}|\, S(C_p^{story}, C_p^{story})}{|C_p^{story}|\, S(C_p^{story}, C_p^{story}) + |C_{-p}^{story}|\, S(C_{-p}^{story}, C_p^{story}) + |\overline{E}|\, S(\overline{E}, C_p^{story})} = r(C_p^{story}) \end{aligned}$$

□

The next Proposition offers an additional explanation, besides Corollary 1.1, for why prompting respondents to imagine contextual features increases recall.

Proposition 7. *Adding memories to the target scenario increases recall.*

Proof. Adding memories to the target scenario means increasing total self-similarity of the cued set $|C_p|\, S(C_p, C_p)$, so that per (19) the probability of recall increases.

□

H.3 Partial information loss in episodic recall

So far, we assumed that, when respondents recall a product's memory trace, they remember it perfectly and can thus form their Bayesian posterior again. This Section relaxes this assumption and allows respondents to remember, but *imperfectly*. More specifically, we now assume that, when they recall the cued trace, participants only remember the valence of the signal, i.e. whether it was positive or negative. Since this creates information loss for statistics, but not for stories, it gives stories an additional edge over statistics. In the following, we formalize the belief formation process if statistics are recalled with information loss.

Stories. For stories, only remembering the valence is sufficient to update priors exactly. This means beliefs will be the same as in Section G.4.

Statistics. For statistics, only remembering the valence means that, conditional on recalling the cued trace, participants will only remember (i) that they saw a sample of random size and (ii) whether the majority of reviews was positive or negative.

We call $\hat{\pi}(N, n', k') = \frac{k'}{N} + \frac{N-n'}{N}\frac{k'+1}{n'+2}$ the rational posterior for the share of positive reviews $\pi := K/N$ after having observed k' positive reviews from a sample of n' reviews taken without replacement from N total reviews (see Section G.4). $\hat{\pi}_1 = \hat{\pi}(N, n, k)$ is the first-period, i.e. correct, posterior.

For now, we focus on positive valences. We only consider strict majorities, as our randomization prevents ties. Define PV as the set of statistics that have a positive valence:

$$PV := \left\{ (k', n') \,\middle|\, k' < \left\lfloor \frac{n'}{2} \right\rfloor,\ 1 \leq n' \leq N \right\} \tag{23}$$

Conditional on recall, second-period beliefs are given by:

$$\hat{\pi}_2^{infoloss,recall} = \frac{1}{|PV|} \sum_{n'=1}^{N} \sum_{k'=0}^{\lfloor n'/2 \rfloor} \hat{\pi}(N, n', k') \tag{24}$$

Moreover, turning to belief impact:

$$\hat{\pi}_2^{infoloss,recall} - \frac{1}{2} = \frac{1}{|PV|} \sum_{n'=1}^{N} \sum_{k'=0}^{\lfloor n'/2 \rfloor} \left(\hat{\pi}(N, n', k') - \frac{1}{2} \right), \tag{25}$$

All summands are of the same sign as $\hat{\pi}_1 - \frac{1}{2}$ since they have the same valence, so that $\hat{\pi}_2^{infoloss,recall}$ is also of the same sign as $\hat{\pi}_1 - \frac{1}{2}$. By the same reasoning, this also applies to negative valences, with the opposite sign. Conditional on the first period,

$$\mathbb{E}\left(\hat{\pi}_2^{infoloss} - \frac{1}{2}\right) = r(C_p)\mathbb{E}\left(\hat{\pi}_2^{infoloss,recall} - \frac{1}{2}\right). \tag{26}$$

Finally, the remark on signs implies $\text{Cov}\left(\hat{\pi}_2^{infoloss,recall} - \frac{1}{2}, \hat{\pi}_1 - \frac{1}{2}\right) > 0$, while:

$$\text{Cov}\left(\hat{\pi}_2^{infoloss} - \frac{1}{2}, \hat{\pi}_1 - \frac{1}{2}\right) = r(C_p)\text{Cov}\left(\hat{\pi}_2^{infoloss,recall} - \frac{1}{2}, \hat{\pi}_1 - \frac{1}{2}\right) \tag{27}$$

Without information loss, in Section G.4, belief decay was exactly recall. Here, it is a product of recall and information loss: since the latter is positive and independent from recall, the increasing relationship between belief decay and recall carries over. Our predictions on recall also map onto predictions on beliefs in this more general model.

Although we have presented a specific model of information loss, this only relies on the positive covariance of the rational posterior and the partial information posterior: arguably, all reasonable models of *partial* information loss should exhibit this feature.