# Stories, Statistics, and Memory*

**Thomas Graeber** **Christopher Roth** **Florian Zimmermann**

October 14, 2022

*Preliminary and incomplete*

**Abstract**

For most decisions, people rely on a myriad of relevant information encountered over the course of days, months or years. Such information comes in various forms, including abstract summaries of multiple data points – statistics – and contextualized anecdotes about individual instances – stories. We propose that people form beliefs based on information they recall on the spot, that they do not always retrieve the full wealth of their accumulated information, and that the information type – story versus statistic – is a central determinant of selective memory. In controlled experiments we show that the effect of information on beliefs decays rapidly and exhibits a pronounced story-statistic gap: the average impact of stories on beliefs fades by 33% over the course of a day, but by 73% for statistics. This pattern is driven by the role of context in memory: prompting contextual associations with statistics slows their temporal decay. Guided by a model of similarity and interference in memory, we experimentally examine the explanatory power of a broad swath of different dimensions of interference. Consistent with the model, similarity relationships – rather than, e.g., memory load per se – are the key driving force behind the story-statistic gap.

* We thank Peter Andre, Pedro Bordalo, Armin Falk, Nicola Gennaioli, Joshua Schwartzstein, Andrei Shleifer as well as seminar audiences at the Belief-Based Utility Workshop in Amsterdam, the Berlin Behavioral Economics Seminar, CERGE-EI Prague, Cornell, the briq Belief Workshop, Harvard Business School, MiddExLab and Sciences Po for helpful comments. We thank Andrea Amelio, Simon Cordes, Paul Grass, Tsahi Halyo, Constantin Schesch, and Malin Siemers for excellent research assistance. The full set of experimental instructions can be found on the following link: `https://raw.githubusercontent.com/cproth/papers/master/SSM_instructions.pdf`. 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 and ECONtribute, `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, people attend to and process a myriad of information over time so that beliefs should eventually gravitate toward the truth. Yet, empirical evidence documents persistent belief biases and misperceptions on a wide range of topics. The information we consume come in various forms, ranging from abstract summaries of large quantities of data – statistics – to vivid descriptions of single events – stories, or anecdotes –, the latter of which are more likely to be unrepresentative of reality at large. In this paper, we entertain the hypothesis that beliefs are not continuously updated every time a piece of information is received. Instead, beliefs are formed on the spot when needed as an input to an action. In light of the wealth of information we process over time, beliefs formed in any given instance then crucially hinge on *which* pieces of information come to mind. A central question for the evolution of beliefs – absent the arrival of new information – is therefore whether some types of information are recalled more easily than others. If stories – even if unrepresentative – are retrieved more easily than statistics, beliefs about important topics can remain persistently biased.

In this paper, using a series of pre-registered experiments, we study the role of memory for the effects of statistical information versus stories on beliefs. We make three key contributions: First, we document that stories shape beliefs more persistently than statistics, even when people initially update more strongly in response to a statistic. This pattern is supported by the recall accuracy of these different types of information: after some delay, people are better at retrieving the information contained in a story than that in a statistic. Second, we show that the rich contextual features of stories are the key reason for why they stick better than statistics. Once people are instructed to imagine typical associations for the information contained in a statistic, recall accuracy substantially improves. Third, guided by a model of similarity and interference in memory, we experimentally examine the determinants of the story-statistic gap and identify different features of interference.

We conceptualize statistics as quantitative information about multiple observations, whereas a story provides quantitative information about a single observation coupled with contextualized qualitative information. This simple taxonomy is relevant in practice: the news we consume are routinely composed of statistical information about a phenomenon and/or anecdotes, and there is a natural link between sample size and contextualized qualitative information: first, anecdotes tend to be specific, and the larger a sample, the fewer features tend to be shared between sample instances. Second, anecdotes are typically not intrinsically linked to the specific magnitude of a statistic, i.e., it's hard to imagine a story that is directly connected to a fraction of 78% (but not, say, 82%).[1]

In our baseline experiment, we aim to study the effects of stories and statistics on beliefs over time. Subjects are informed that a hypothetical product received a specific number of reviews, and are asked to state an incentivized guess whether a randomly selected review is positive. Before stating a guess, subjects are either exposed to a story, statistical information or no additional information. Closely following our conceptualization above, a statistic corresponds to quantitative information about multiple reviews, whereas a story provides quantitative information about a single review plus contextualized qualitative information. We experimentally induce a prior about the fraction of positive reviews. Each subject faces a sequence of three independent product scenarios. We use two key sources of variation. First, subjects need to state each belief twice, once directly when the information is received, and once again following a one-day delay. We inform subjects that the information we provide in the baseline survey will be payoff-relevant one day later. This temporal structure is central to our design: there are many differences between information in the form of stories and statistics that might lead to differential belief updating. However, all such differences that are *not* related to memory will already be borne out in the immediate update. Therefore, since no new information is received in between, any *change* in stated beliefs over time must be, by construction, rooted in memory. Second, for each product, we vary whether subjects are exposed to a story, a statistic or no information.[2]

[1]We do not posit, however, that statistics and stories are generally incompatible. In fact, we design treatments that aim to connect both. Instead, we argue that anecdotes in practice tend to about one or few instances rather than large samples. If a statistic is coupled with an anecdote, the anecdote is typically picking out an illustrative instance from the larger sample.

[2]Each subject receives one story, one statistic and once no additional information. This keeps the total

We document a story-statistic gap in the evolution of beliefs: the effect of stories on beliefs decays at a significantly lower speed than the effect of stories on beliefs. Pooling all statistics and stories presented in our baseline study, we find that, on average, the belief impact of statistics decays by more than twice as much as that of stories over the course of a day. Using a free recall task, we find that participants are more accurate at recalling the correct type and valence of the information for scenarios in which they received a story than those in which they received a statistic. We establish the robustness of our baseline result to the following experimental features: (i) the extremity and valence of statistical information; (ii) the valence of stories; (iii) the number and type of decoy information; (iv) and the question format of the belief elicitation.

Delving into underlying mechanisms, we begin by acknowledging that there are many dimensions along which stories and statistics differ, several of which might affect updating, such as the time spent engaging with the information, their emotionality, their vividness etc. In fact, there is comparably little research in experimental economics on how people learn from qualitative as opposed to quantitative information. We aim to make progress by focusing on a specific channel, the cue-dependent nature of memory. Intuitively, cue dependence means that memories are not retrieved by simply "thinking of them". Instead, retrieval typically requires thinking of something *associated* with a specific memory. Cue-dependent recall is one of the most well-documented and robust insights from research on human memory (Kahana, 2012) and we argue that the story-statistic gap is a natural application: the rich and highly contextualized elements of stories create associations that make them easier to retrieve given a relevant scenario cue than abstract and context-less statistical information. Consistent with this basic conjecture, an additional mechanism experiment reveals that prompting respondents to imagine a typical review when provided with statistical information increases delayed belief impact, even though immediate updating is unaffected. Put differently, asking subjects to add entirely fictional contextual features to a statistic on their own improves recall and slows the time decay of information in beliefs.

information load constant across subjects.

We develop a simple formal framework that can explain the story-statistic gap and guides our systematic analysis of underlying mechanisms. The main building blocks of the model are the principles of similarity and interference (Kahana, 2012). The more similar a target memory trace is to the cue, the higher the chances of retrieval. Interference occurs when similar non-target memory traces interfere with the retrieval of a target trace. The model builds on the theoretical frameworks of Bordalo et al. (2021b,a), but adapts those to accommodate both stories and statistics. It clarifies that both similarity and interference favor the recall of stories compared to statistics. The richness of stories makes relevant memory traces (i) more similar to themselves and to the cue, and (ii) more specific, which, in turn, makes them less susceptible to interference from non-relevant memory traces. Similarly, viewed through the lens of the model, the reason why being prompted to add contextual features to a statistic effectively boosts recall is that this mitigates the strong interference statistics tend to suffer from.

A key insight from the model is that recall of both stories and statistics hinges on (i) self-similarity, i.e., the similarity between memory traces created for a piece of information as well as those traces and the cue, and (ii) cross-similarity, the similarity between target and decoy information. Guided by this simple distinction, we conduct a series of mechanism experiments where we systematically manipulate self- and cross-similarity in different ways. We report the following findings, which are consistent with the predictions of our framework. First, as we move from one to three and then to six product scenarios, the story-statistic gap increases. Intuitively, a higher number of product scenarios increases cross-similarity, thereby creating a higher risk of memory interference. The richness of stories at baseline, however, makes them less vulnerable to this type of interference, hence widening the story-statistic gap. Second, focusing on the retrieval of stories, we show that higher similarity between the contextual features presented in different stories (higher cross-similarity) has a negative effect on the persistence of belief impact. Third, we probe self-similarity and show that the similarity between the product cue and the corresponding story somewhat eases the recall of the information, though these effects do not quite reach statistical significance.

Finally, a higher (cross-)similarity between cues has a negative effect on the delayed belief impact for both stories and statistics. Taken together, this evidence suggests that similarity relationships are the key driving force of the story-statistic gap. We find that alternative factors, such as overall memory load, appear to play a subordinate role.

Our work is a related to a recent literature on stories and narratives in economics (Shiller, 2017, 2020; Michalopoulos and Xue, 2021; Andre et al., 2022; Kendall and Charles, 2022; Morag and Loewenstein, 2021). This literature has mostly focused on the persuasive effects of narratives in the moral or political domains (Bénabou et al., 2018; Eliaz and Spiegler, 2020; Bursztyn et al., 2022, 2021; Alesina et al., 2022). Relatedly, a literature in psychology and management has focused on the power of stories in influencing people (Fryer, 2003; Monarth, 2014; Bruner, 1987). We add to these literatures by (i) focusing on the comparison of stories versus statistics for belief formation and (ii) studying the dynamics of belief formation after being exposed to stories versus statistics and (iii) providing a rich set of evidence on mechanisms with a focus on the role of contextual information and interference.

Our work contributes to a growing literature on the role of memory in economics (Bordalo et al., 2021d,c; Gennaioli and Shleifer, 2010). Our model heavily relies on Bordalo et al. (2021a,b), who provide theoretical frameworks in which people estimate probabilities by retrieving experiences from memory based on similarity and interference. On the empirical side, Enke et al. (2020) study the role of associative memory for belief formation and show that it can give rise to overreaction to news. Kwon and Tang (2020) and Charles (2021), using observational data, argue that associative memory may be a driver of investment behavior. Afrouzi et al. (2020) experimentally highlight the role of working memory in forecasting experiments.[3] Our paper differs from these papers in its focus on how different types of information, statistical versus anecdotal information, shape belief formation over time.[4] More broadly, our work builds on an extensive psychology literature on memory, see Schacter (2008) and Kahana (2012) for overviews.

[3]Other work in this area has explored the interaction between memory and motivated reasoning (Zimmermann, 2020).

[4]We also contribute to a large literature on biases in belief formation (Enke and Zimmermann, 2019; Graeber, 2022; Enke, 2020; Martínez-Marquina et al., 2019).

This paper proceeds as follows: Section 2 presents baseline experiments which demonstrate the existence and robustness of a story-statistic gap in memory. Section 3 provides evidence on the role of associations and contextual information in driving the recall of information. In Section 4, we outline a simple theoretical framework that formalizes the mechanisms underlying the story-statistic gap in memory. Section 5 provides evidence on different features of interference which shape forgetting. Section 6 discusses the implications of our findings.

# 2 Evidence for a Story-Statistic Gap in Memory

## 2.1 Design

Our baseline experiment is motivated by the following design objectives: (i) a dynamic structure with binding memory constraints; (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 where information both in the form of statistics and stories would be typical; and (iv) an incentive-compatible belief elicitation. Table A.8 provides an overview of all experimental designs.

**Task structure.** Subjects were informed that there are three different hypothetical products. Each of the products has received a number of reviews, with each review in turn being either positive or negative. For every product, subjects' task was to guess whether a randomly selected review is positive. To fix prior beliefs, we truthfully informed subjects that the actual number of positive reviews would be randomly and independently determined for each product, inducing a flat prior. For each product, participants then received either a piece of additional information or no additional information and were subsequently asked to state a belief.

**Sources of variation.** We implemented two key sources of variation. First, we created variation in the timing of the belief elicitation in a within-subject design. In *Immediate*, beliefs were elicited directly on the screen on which they received additional information; in *Delay*, beliefs were elicited one day later.

Second, within-subject and across products, we varied the type of additional information subjects were exposed to. For each product, participants received either statistical information (condition *Statistic*), or anecdotal information (condition *Story*), or no further information. Randomization was blocked such that each individual received one story, one statistic and once no additional information. Moreover, the order of products was randomized and each individual received one positive signal and one negative signal. We thus leveraged within-subject variation in both the type and the direction of the information. Appendix D provides details on the implementation of the randomization.

Recall that 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 follows this taxonomy.

When subjects obtain statistical information, we provided them with information about the fraction 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 and precision of statistics. Below is an example of how statistics were communicated:

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

When subjects receive a story, we provided them with information about whether a randomly selected review was positive or negative, plus a qualitative description of that review. The description typically consisted of 6-7 sentences recounting the experience that led to the review. The valence of the text matched the overall review. Below is an example of a story accompanying a negative review about a restaurant:

> 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!"

Appendix C.1 reproduces all stories from the baseline experiment. To test the importance of the valence of story content, we cross-randomized whether the contextual information in the story was consistent with the overall rating of the reviews – i.e., included only positive sentences when the reviews was positive or negative ones when the reviews was negative –, mixed, or neutral. Section 2.3 summarizes our findings.

An important feature of stories is that they cannot be accommodated in a Bayesian framework as the informational content of qualitative features cannot be quantified in a fully objective way. For instance, in the above example, the qualitative description of the food arguably allows subjects to infer that other reviewers 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 subjects *perceive* each story to be. 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.

**Belief elicitation.** We elicit beliefs twice, once immediately after receiving the information and once one day later. Our main outcome of interest is respondents' beliefs about the likelihood that a randomly selected review was positive. The belief elicitation is incentivized using a binarized scoring rule where subjects can earn $30.[5]

**Recall elicitation.** To provide direct evidence on recall of the additional information received in the baseline survey, we asked our respondents in an unincentivized open-ended survey question:[6]

Please tell us anything you remember about this product scenario. Include as

[5]The precise payment formula was 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.

[6]We randomized the order of belief and recall elicitation.

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 where we replicate our baseline design, we include structured incentivized recall tasks and show that they yield very similar results (see Section 2.3).
To analyze this data, we designed and implemented a hand-coding scheme, which we describe in more detail in Appendix E. The hand-coding scheme records whether respondents mention the valence and type of information they encountered, and whether they correctly remember these characteristics. On top of this, the coding scheme measures 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.

**Procedures, payment and pre-registration.** All experiments were conducted online with participants recruited from the survey platform Prolific.

We pre-registered this study on AsPredicted, see `https://aspredicted.org/e5mw7.pdf`. The pre-registration includes the experimental design, predictions and 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 in advance that the information they receive will 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 respondents whose responses were then implemented to determine a bonus payment.[7] 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 and implemented.

We collect 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[8], our final sample consists of 985 participants, corresponding to a completion rate of 69 percent.[9] 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 considering stories with content that is consistent with the overall review rating. In Section 2.3, we examine the robustness of results when looking at mixed and neutral stories. Panel (a) of Figure 1 and Table A.7 show the average belief impact in *Immediate* and *Delay*, pooling the data across different 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 rational update.

In line with our hypothesis, the difference-in-difference estimate of belief impact between the *Immediate* and *Delay* conditions reveals a much slower temporal decay for stories than for statistics, which is highly significant ($p < 0.01$). Considering point estimates of belief impact in the *Delay* condition, Panel (a) of Figure 1 reveals that mean belief impact after

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

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

[9] Given that the key treatment variation is within-person, the attrition rate is not a threat to the internal validity of our findings.

1 day is substantially more pronounced for stories than for statistics. On average, beliefs moved from the 50% prior by 12.33 p.p. (s.e. 0.79) and 5.60 p.p. (s.e. 0.69) for stories and statistics, respectively. The divergence in belief impact is significantly different from zero ($p < 0.01$). This contrasts with belief impact in the *Immediate* condition. Average belief impact in *Immediate* is actually larger for the statistics treatment than for the story treatment. Beliefs moved on average by 20.63 p.p. (s.e. 0.59) and 18.26 p.p. (s.e. 0.69) for the statistics and stories treatment, respectively. Appendix Figure A.1 confirms the patterns with the cumulative distribution function of belief impact in *Immediate* and *Delay*, separately for stories and statistics.

This result on the relative decay of belief impact is corroborated by the hand-coded recall data. In what follows, we focus our presentation of results on the fraction of respondents who correctly recall both the type and the valence of the information they were provided.

Panel (b) of Figure 1 shows that correct recall is significantly higher for stories compared to statistics ($p < 0.01$). Average correct recall is 61.13 percent for stories and 26.71 percent for statistics. This suggests that the quantitative information in stories is more easily retrieved than statistical information. Moreover, the richness of the open-ended data reveals further striking features: (i) A large fraction of respondents (22.68%) mention qualitative features from the story without specifically being prompted to do so; (ii) a much smaller fraction of respondents (7.01%) mention the statistic they received; and (iii) only a negligible fraction (1.66%) mention the posterior belief they stated in the baseline wave.

Taken together, our first main result can be summarized as follows:

**Result 1.** *We document a story-statistic gap in memory: stories have a stronger effect on beliefs after a delay compared to statistics, even though statistics have stronger immediate effects. Recall accuracy is higher for stories than for statistics.*

## 2.3 Robustness

**Valence and extremity of statistics.** So far we presented results about average belief impact across different stories and statistics. This perspective masks potentially important heterogeneity related to the valence (positive vs. negative) and extremity of the information. Our finding of a story-statistic gap turns out to be highly robust. First, we find no difference in recall and delayed belief impact by whether the statistical information is mostly positive or mostly negative. Second, in Figure 2, we analyze heterogeneity by the extremity of immediate updating. In Panel (a) of Figure 2, we regress beliefs from *Delay* on beliefs in *Immediate* separately for respondents in the story and statistic condition. Perfect recall would imply a slope of one. The figure shows that the slope is significantly higher in the story condition than the statistic condition. This reveals that, for a given immediate reaction to a statistic and a story, the delayed reaction is more pronounced for the story. Panel (b) of Figure 2 corroborates this result, showing how correct recall of the type and valence of the provided information is related to immediate updating, again separately for stories and statistics. The Figure reveals that for all levels of immediate updating, recall accuracy is substantially higher for stories than for statistics. Third, and similarly, we find no heterogeneity in delayed belief impact by the randomly assigned objective extremity of statistics (results available upon request). This, in turn, suggests that neither the magnitude of the immediate subjective update nor the objective extremity of the provided information play a critical role in shaping recall in the case of statistical information.

**Valence of contextual information.** In our analysis so far we focused on stories that were consistent with the quantitative information contained in the review. To test the importance of the valence of story content, we cross-randomized whether the contextual information in the stories was (i) only consistently positive/negative, (ii) of mixed valence, or (iii) neutral.

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. Figure 3 and Column 5 of Table A.7 show that recall accuracy is relatively similar across all three conditions. Average correct recall is 61.13 percent in the consistent story condition compared to 54.92 and 48.79 percent in the mixed and neutral stories treatments, respectively (compared to 26.71 percent for statistics). These data suggest that recall of the quantitative information contained in the stories is moderately affected by the valence of the contextual information: as could be expected, recall is highest for consistent stories, where the content of the story can help infer the objective valence of the review. Mixed and neutral stories do not provide a similarly tight connection between the story content and the objective review rating.

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* (results available upon request).

### 2.4 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.

**Design.** How robust is the story statistics gap to the number type and valence of decoy information? Between-subjects, we compare the effect of receiving no decoy information to receiving either same-type information or information of a different type.

We adapt the baseline design but now focus on one "target scenario" that either provides a story or a statistic. For both types of target information, we then exogenously manipulate the type of information for the two other scenarios ("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. To obtain an incentivized measure of recall, instead of the open-ended measure, we implemented a structured recall task.[10] We ask respondents to indicate which type of information they

[10]We randomized the order of our measures of recall and the belief elicitation.

received about a given product. 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.[11]. Unless respondents indicated that they did not receive any information about this product, we additionally asked them to indicate whether the information they received was positive or negative.[12] Respondents were 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 bonus.

**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.[13]

We also pre-registered this experiment on AsPredicted, see `https://aspredicted.org/qy3wq.pdf`, including the experimental design, predictions and analysis, sample sizes, and exclusion criteria.

**Results.** Figure A.2 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. For each type of target, the 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

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

[12] We tailored the question wording for respondents according to whether they indicated having receive a single review, multiple reviews or "not knowing".

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

structured recall elicitation.[14] 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 story-statistic 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.

Appendix Figure A.3 shows how belief impact and recall vary depending on the valence of decoy information. We again document a robust and sizeable 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.

## 2.5 Robustness to Question Format

It is conceivable that the question format of the belief elicitation affects the size of the story-statistic gap: to the extent that the question format more closely resembles the display format of the story as opposed to the display format of the statistic, recall of stories might be favored. In our main experiment, we elicited beliefs about the likelihood of one randomly drawn review being positive. As a robustness exercise, we examined a different question format, which is plausibly more similar to the display format of the statistical information.

**Design.** The experiment featured one main treatment variation. The *Likelihood Format* treatment elicited beliefs as before – about the likelihood that a randomly chosen review is positive – and thus exactly corresponded to our main experiment. In the *Fraction Format* condition, by contrast, we elicited beliefs about the percentage of positive reviews in the overall population of reviews of the product.[15] The rationale of this manipulation was that

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

[15]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.

in *Fraction Format*, the question people answered might be closer in spirit 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 robustness manipulation that further zeroes in on the importance of the link between question and information display format, we randomized whether the statistical information itself was expressed in terms of an absolute number of positive reviews in a subsample – as in our main study – (*Statistic Number Display*) or in terms of a percentage of positive reviews in a subsample (*Statistic Percent Display*). This means we ran a total of four between-subject conditions, reflecting a 2 (*Likelihood / Fraction Format*) x 2 (*Statistic Number / Percent Display* ) treatment variation.

The comparison between the *Likelihood Format* and the *Fraction Format* conditions allows us to probe the robustness of the statistic-story gap to a different way of eliciting beliefs, the latter of which should increase the similarity between the statistic and the cue. The conditions involving the percentage display of statistical information allow to study two further ideas. First, the percentage display of information makes the additional information even more similar to the cue whenever the question format involves a fraction, creating a "high cue-statistic similarity" condition. And second, they provide an additional robustness check for whether the display of statistical information in terms of absolute numbers as opposed to a percentage affects the story-statistic gap.

**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.[16]

The pre-registration for this experiment can be found on AsPredicted, see `https://aspredicted.org/ZFF_88V`. The plan specifies the experimental design, predictions and analysis, sample sizes, and exclusion criteria.

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

**Results.** Appendix Figure A.4 and Table A.2 document that the fraction question format has a positive, albeit small effect on delayed belief impact and recall. The decay of belief impact is somewhat smaller in the *Fraction Format* than the *Likelihood Format*. This effect is somewhat more pronounced in the recall data, and reaches significance for the case of statistical information (column (4) of Table A.2), in line with the notion that a higher similarity of the question format to the statistic slightly improves retrieval. Taken together, we take this as evidence that the similarity of the question format to the piece of additional information likely affects the retrieval likelihood, but that this effect is small in magnitude.

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 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 has little impact. In practice, we might expect that question format and display format are much *less* similar to each other, and that variation in similarity across contexts plays a larger role.

# 3 The Role of Contextual Associations

Our robustness exercises indicate that the story-statistic gap may not merely be a result of excessive memory load. In the following, we aim to investigate the precise memory mechanism that is responsible for our main finding. We will focus on a specific channel, the *cue-dependent* nature of recall, which counts to the most robustly documented features of human memory (Kahana, 2012). Memory retrieval is determined by associations with other memories, which act as cues. We argue that this basic feature of memory naturally lends itself to our specific application: concretely, we explore the hypothesis that the qualitative, contextualized nature of stories creates many specific associations that make them easier to retrieve than abstract, context-free statistical information. Before outlining a formal framework along those lines in Section 4, we first provide a direct, "reduced-form" test of this intuition: that the advantage of stories in memory should vanish once the contextualized elements are removed from stories or added to statistics.

**Design.** To causally examine the role of contextual information, we ask our subjects to add context to abstract information on their own. Specifically, we prompt respondents to imagine a typical review for the statistic or for a single review they learn about. Note that this intervention does not provide any objective information, qualitative or quantitative, allowing us to identify the distinct effect of associating obviously irrelevant contextual features with a piece of information in memory.[17]

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 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 receive information about a single review are now asked to imagine what the review might look like.

**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.[18]

[17]We note here that this is arguably a rather subtle intervention, hence one might not expect to see major effects.

[18]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.

The pre-registration for this experiment can be found on AsPredicted, see `https://aspredicted.org/v9gk7.pdf`, and again includes the experimental design, predictions and analysis, sample sizes, and exclusion criteria.

**Results.** The median (mean) number of words subjects wrote to describe an imaginary typical review was 22 (23). The text responses indicate that the vast majority of subject made an effort to describe a review, such as in the following example response from the *NoStoryPrompt* condition about a negative review about a videogame:

> 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 from the *NoStoryPrompt* and *StatisticPrompt*, as well as the *NoStory* and *baseline* condition.[19] Panel (a) of Figure 4 shows results on belief impact, while Panel (b) 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.1 reveals that the difference-in-difference (difference in slopes) is statistically significant ($p < 0.05$).

These patterns for *Delay* beliefs are underscored by results on recall. Panel (b) of Figure 4 shows that recall accuracy is 43.14 percent for respondents in *Prompt*, compared to 32.69 percent in the conditions without prompt. Table A.1 reveals that 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.

[19] Table A.1 shows results separately for all 4 conditions and confirms that the overall patterns are very similar.

Our second main result can be summarized as follows:

**Result 2.** *Contextual features cause more pronounced belief impact in delay and facilitate more accurate recall of information.*

This reduced-form evidence provides a first indication that contextual associations may play an important role for the story-statistic gap, motivating our theoretical framework as well as an in-depth empirical analysis of the potential mechanisms related to cue-dependent memory.

# 4 Outline of Conceptual Framework

In the following we outline a stylized model of cue-dependent memory. This framework is not meant as a general or exhaustive theory of human memory. Instead, it 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. All details and formal derivations are relegated to Appendix F. The basic mechanics of the model directly build on Bordalo et al. (2021b,a). The framework rests on two central assumptions: (i) people tend to forget previously formed beliefs and instead retrieve prior information from memory; (ii) retrieval is based on two key memory principles, similarity and interference.

## 4.1 Setup

Agent $i$ forms beliefs about the quality of different products $j \in \{1, ..., n\}$. For each product, there exists a number $r_j$ of reviews, each of which is positive or negative. The true fraction of positive reviews is drawn independently and uniformly for each product. There are two periods $t \in \{1, 2\}$ that we may think of as "past" and "present." In both periods, the agent needs to state a belief for each product $b^t_{i,j}$ about the likelihood that a randomly selected review for this product is positive.

In $t = 1$ (the "past"), the agent might receive information about a product. We distinguish two types of information, statistics and stories. A statistic for product $j$ consists of $n_j$

random draws (without replacement) from the total number $r_j$ of existing reviews. Let $m_j$ denote the number of positive random draws. A story for product $j$ consists of one randomly drawn review from the total number of reviews $r_j$, plus anecdotal information that describes the experience of the reviewer with the product. For now we assume that this anecdotal information has valence that is consistent with the outcome of the random draw (positive or negative).

## 4.2 Recall

Our key interest is in belief formation at $t = 2$. We assume that the agent has forgotten his period 1 posterior belief $b^1_{i,j}$ and hence relies on pieces of information retrieved from memory.[20] Retrieval is guided by cue-dependent memory that is subject to similarity and interference (Kahana, 2012; Bordalo et al., 2021b).

### 4.2.1 Memory Traces

Past information is encoded in the form of one or several *memory traces*. A memory trace is a binary vector of features. Let 1 denote that a certain feature is present and 0 that it is not. $M$ describes the set of all memory traces.

We distinguish between "context features" and "content features". Context features include the time and place a memory was made. It is useful to distinguish between information and cues from within the experiment and information from outside the experiment. For simplicity, we encode the entire context associated with the experiment as a single feature called 'Context-Experiment'. Moreover, we assume that the name of the product is a context feature of every memory trace. Content features capture all the additional information that is received and saved in a memory trace, in particular the quantitative and qualitative (anecdotal) information.

We assume that a story creates multiple memory traces, one for each element of the

[20] In the experiment, we remind subjects of the uniform prior. We hence assume here that the agent perfectly recalls his prior.

anecdotal information. All these traces include the quantitative information of the review and the content relates to the product itself. The valence of the anecdotal element is also part of the trace. A statistic instead is encoded as a single memory trace that contains the quantitative information but doesn't relate to the type of product via its content.

**Example:** Assume there are two product scenarios, 'Bicycle' and 'Restaurant'. Further assume that for the bicycle a negative story was provided. The anecdotal elements were: 'I had bicycle crash' and 'The bicycle is extremely heavy'. In the restaurant scenario, the agent obtained a positive statistic. So having only the two products 'Bicycle' and 'Restaurant', each memory traces starts with the following context features:

$$(1_{\textit{Context-Exp}}, 1_{\textit{Context-Bicycle}}, 1_{\textit{Context-Rest}}) \tag{1}$$

Next, moving to the content features, in the case of a statistic the indicator $1_{(m,n)}$ summarizes a hypergeometric statistic with $m$ successes out of $n$ random draws without replacement. In the case of a story, the indicators $1_{(1,1)}$ and $1_{(0,1)}$ stand for a positive and negative review, respectively. The story traces additionally feature entries for 'Crash', 'Weight' and 'Content-Bicycle'.[21] Memory traces in our example then comprise the following indicators:

$$(1_{\textit{Context-Exp}}, 1_{\textit{Context-Bicycle}}, 1_{\textit{Context-Rest}}, 1_{\textit{Content-Bicycle}}, 1_{\textit{Crash}}, 1_{\textit{Weight}}, 1_{(m,n)}, 1_{(1,1)}, 1_{(0,1)}, 1_{\textit{Pos}}, 1_{\textit{Neg}}) \tag{2}$$

The negative bicycle story hence leaves two memory traces, one for each anecdotal element:

- (1,1,0,1,1,0,0,0,1,0,1) - 'I had a crash with the bicycle'
- (1,1,0,1,0,1,0,0,1,0,1) - 'The bicycle is extremely heavy'

The statistic leaves a single memory trace given by:

- (1,0,1,0,0,0,1,0,0,1,0)

[21] We call the content feature indicating bicycle 'Content-Bicycle' to distinguish it from the context feature for bicycle, 'Context-Bicycle'.

### 4.2.2 The Cue

We think of the cue as the task description that an agent is presented with, i.e., the prompt to state a belief $b_{i,j}$. We hence conceptualize the cue as composed of the context of the task and the product name ('Context-Experiment'+'Name product'). Next, a cue invokes a set $C$, the *cued set*, that contains all memory traces sharing the features of the cue, i.e., all memory traces that were encoded during the experiment and are related to the product of interest. The cued set can be thought of as the target traces that the agent would like to retrieve upon being presented with the cue.

### 4.2.3 Similarity and Interference

Recall of a memory trace $e$ depends on the similarity between the cue $C$ and $e$. To begin, we formalize the similarity between two arbitrary memory traces, $e$ and $e'$. We propose a general formulation of similarity based on only the following assumptions:

(i) Adding the same feature to $e$ and $e'$, ceteris paribus, increases their similarity.

(ii) Adding a feature to $e$ but not to $e'$, ceteris paribus, decreases similarity.

We define the similarity between memory trace $e_0$ and $C$ as:

$$S(e_0, C) = \frac{1}{\|C\|} \sum_{e \in C} S(e_0, e) \tag{3}$$

Leveraging insights from basic research on similarity-based, cue-dependent memory(Kahana, 2012), we define the recall probability of memory trace $e_0$ given cue $C$ as:

$$r(e_0, C)) = \frac{S(e_0, C)}{\sum_{e \in M} S(e, C)} \tag{4}$$

Equation (4) illustrates that both similarity and interference shape the retrieval process. The similarity between the target trace $e_0$ and $C$ enhances recall. At the same time, the higher the similarity between $C$ and other memory traces, the more these traces interfere with the recall of $e_0$.

## 4.3 Beliefs

We model beliefs in $t = 2$.[22] Agents perform recall as introduced above to retrieve information from $t = 1$. Agents only sample once from their memory data base. If they retrieve a memory trace that is helpful (i.e., that contains information about the product of interest), they use this information to form a posterior following Bayes' rule. If they retrieve an irrelevant memory trace (i.e., that contains no information about the product of interest) they realize that the trace is not helpful and state their prior.

Our main interest is in the overall probability that one of the relevant traces (i.e., a trace in $C$) is retrieved. The probability for this is given as:

$$r(C, C) = \frac{\overbrace{S(C, C)}^{\text{Self-similarity}} \cdot \overbrace{\|C\|}^{\text{Target load}}}{S(C, C) \cdot \|C\| + \underbrace{S(\overline{C}, C)}_{\text{Cross-similarity}} \cdot \underbrace{\|\overline{C}\|}_{\text{Decoy load}}} \tag{5}$$

The intuition behind this expression is straightforward and structures our predictions. $S(C, C) = \frac{1}{\|C\|} \cdot \sum_{e \in C} S(e, C)$ is the self-similarity of the target set $C$. High self-similarity means that the target traces are highly similar to the cue (more precisely, the memory traces that the cue evokes). The higher self-similarity is, the higher is the likelihood that a relevant memory trace is retrieved. $S(\overline{C}, C) = \frac{1}{\|\overline{C}\|} \cdot \sum_{\overline{e} \in \overline{C}} S(e, C)$, with $\overline{C} = M \setminus C$, captures the non-cued set of memory traces, i.e., those traces that do not share the key features of the cue. The higher cross-similarity is, the higher is the likelihood that non-relevant memory traces interfere with relevant ones. $\|C\|$ and $\|\overline{C}\|$ are weights capturing the number of relevant (target) and irrelevant (decoy) memory traces.

[22] In $t = 1$, belief formation is trivial as agents perform Bayesian updating on their uniform prior given any possible additional information.

### 4.4 The Story-Statistic Gap and the Role of Contextual Features

We here summarize the key intuition following equation (5) for how our model captures the story-statistic gap as well as the role of context for retrieval. Appendix F contains formal proofs.

Equation (5) clarifies that there are three forces pushing for a higher retrieval rate of stories compared to statistics: (i) Stories create multiple memory traces hence generating a higher target load $\|C\|$; (ii) the rich contextual features cause a higher similarity between the different memory traces of a story, i.e., a higher self-similarity $S(C, C)$; and (iii) contextual features tend to be distinct and hence less susceptible to interference, causing lower cross-similarity $S(\overline{C}, C)$.

Equation (5) also offers a direct insight into why adding a prompt to imagine context related to a statistic improves recall, analogous to the model forces underlying the story-statistic gap. Viewed through the lens of our model, the prompt (i) produces additional target load, i.e., a higher $\|C\|$), (ii) triggers contextual features that increase self-similarity $S(C, C)$ and (iii) makes the relevant memory traces more distinct, thereby mitigating interference, leading to lower cross-similarity $S(\overline{C}, C)$.

## 5 Features of Interference

The model outlined in Section 4 is a productive guiding framework for our analysis of the mechanisms underlying the story-statistic gap. Equation (5) implies that there are two central comparative static forces that determine the recall of a piece of information: self-similarity and cross-similarity. Such similarity relationships are determined across the different elements of memory traces, moderating interference and hence forgetting. In the following, we will systematically explore similarity relationships in our experiment with the goal to (i) shed further light on the story-statistic gap and (ii) provide a deeper understanding of what makes stories (not) stick. We investigate the role of (i) the number of product scenarios presented within the experiment, (ii) the similarity of different pieces of information, (iii) the similarity between a product cue and its corresponding story (iv) the similarity between a cue and a corresponding statistic and (v) the similarity of product cues.

## 5.1 The Number of Product Scenarios

In a first step, we investigate how the number of product scenarios impacts the story-statistic gap. As discussed in detail below, the overall number of scenarios affects both cross-similarity and the decoy load.

**Design.** The design broadly follows the structure of the main baseline design. The key difference was that we varied, between-subjects, whether there were one, three or six product scenarios. In the *1-product* treatment, there was 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 saw three scenarios and received two pieces of information, one story, one statistic and once no information. In the *6-product* treatment, subjects saw six scenarios overall and also received 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 allowed us to cleanly study the effects of the number of product scenarios, while holding the total pieces of information constant.[23]

To keep incentives exactly constant between the different conditions, subjects in all treatments completed 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 completed 5 dot estimation tasks, while respondents 3-product treatment arm completed 3 dot estimation tasks, and respondents in the 6-product treatment only faced product-related tasks. The experimental instructions for the dot estimation task, in which subjects had 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`.

[23] 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.

**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 42 percent.[24]

The pre-registration for this experiment can be found on AsPredicted, see `https://aspredicted.org/as7i7.pdf`, and again includes the experimental design, predictions and analysis, sample sizes, and exclusion criteria.

**Prediction.** We again focus on the intuition and relegate the formal proof to Appendix F. Viewed through the lens of equation (5), more product scenarios imply a higher decoy load $\|\overline{C}\|$. Hence, both a target story and a target statistic are more difficult to recall as we increase the number of product scenarios. However, equation (5) also reveals that the strength of this effect depends on cross-similarity $S(\overline{C}, C)$. Since the rich contextual features of stories tend to make them less similar to additional product scenarios compared to statistics, cross-similarity is lower for stories. As a consequence, recall of statistics will be more adversely affected by an increase in product scenarios compared to the recall of stories. We predict that the story-statistic gap widens as we move from 1 to 3 to 6 scenarios.

**Results.** Figure 5 and Table A.3 illustrate changes in belief impact between immediate and delay as well as recall for stories and statistics across the different number of product scenarios. Panel (a) depicts the change in belief impact between *Immediate* and *Delay* across the three treatment arms, separately for stories and statistics. We find that the change in belief impact tends to become more pronounced as we increase the number of product scenarios. While this effect is small in the case of stories, it is sizeable in the case of statistics. In line with our model, the story-statistic gap widens with the number of products scenarios.

This pattern is also evident from the recall data, see Panel (b) of Figure 5. Recall accuracy of statistics drastically decreases as we move from 1 to 3 to 6 scenarios, while recall

[24]The completion rate to the follow-up survey does not differ significantly across treatment groups ($p = 0.85$).

accuracy of stories remains comparably stable.

**Result 3.** *The story-statistic gap increases in the number of product scenarios.*

Viewed through the lens of the model, Result 3 suggests that the differential effect of the number of product scenarios on stories versus statistics derives 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.

## 5.2 The Similarity of Story Content

Next, we focus on the specificity of stories, further zooming in on the role of cross-similarity. While we studied the notion of cross-similarity somewhat indirectly via the number of non-target scenarios in Section 5.1, we here directly manipulate cross-similarity via the similarity between a target story and decoy stories.

**Design.** We designed two treatments to study the role of story similarity. The incentives and basic setting were identical to our main experiment. Participants in both conditions learned about three products: a cafe, a restaurant, and a bar. Unlike in our main experiment, respondents received a story in each of the three scenarios. The target story in both conditions was a positive review about the bar. The stories about the restaurant and the cafe were decoy stories and both featured a negative review. In the *Baseline* condition, the three stories were distinct and specific to each cue. The bar story described the interior of the bar, the restaurant story focused on food quality, while the cafe story was concerned with the service quality. In the *Story Similarity* condition, we kept the target story about the bar identical to *Baseline*, but increased the similarity of the two decoy stories to the target story by adjusting both the text structure and content. Specifically, in *Story Similarity*, the three products were still a cafe, a restaurant, and a bar, but all stories were concerned with the interior design of the respective places. Thus, our treatments held the target story exactly constant, but we increased the similarity between the two decoy stories and this target story in *Story Similarity* relative to *Baseline*. We kept everything else fixed. Appendix C.2 provides an overview of all stories 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 5.3. 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.[25]

The pre-registration for this experiment can be found on AsPredicted, see `https://aspredicted.org/v7hh6.pdf`. The plan contains the two conditions described in this section and a third condition described in Section 5.3 and again specifies the experimental design, predictions and analysis, sample sizes, and exclusion criteria.

**Prediction.** Higher similarity between target and decoy stories directly causes higher cross-similarity $S(\overline{C}, C)$. As a consequence, per equation (5), the recall probability of the target story will decrease as decoy stories become more similar. We provide a formal proof in Appendix F.

**Results.** Panel (a) 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 delayed impact is significantly lower in *Story Similarity* than in *Baseline*, even though immediate belief impact is larger in the former condition. 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 A.4 confirms the visual pattern and shows that the difference-in-difference (difference in slopes) is statistically significant ($p < 0.01$).

Panel (b) illustrates similar patterns for the recall data: Among respondents in the *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 the *Story Similarity*.

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

**Result 4.** *Delayed belief impact and recall of a story is impaired by higher similarity to stories in other scenarios.*

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. Second, it delineates the limits of the stickiness of stories in memory. If the memory data base contains many similar stories, retrieval of a target story gets crowded out and it becomes less likely that this story comes to mind.

## 5.3 Cue-Story Similarity

We maintain our focus on the recall of stories, but move beyond variations of cross-similarity to zoom in on the role of self-similarity. Specifically, in the context of stories, self-similarity is naturally generated by the intrinsic link between the content of a story and the cue: 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. In the following, we attempt to exogenously manipulate the similarity between target cue and target story.

**Design.** To study the role of cue-story similarity, we employed the same *Baseline* condition as in the preceding subsection, but compared it to a different treatment *Cue-Story Similarity*.[26] This condition relied on the same decoy stories for the cafe and the restaurant as *Baseline*, but the bar story was about 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.

**Prediction.** Lower similarity between target cue and target story decreases self-similarity $S(C, C)$. As illustrated in equation (5), the recall probability of the target story decreases as it becomes less similar to the cue (see Appendix F for a formal proof).

[26]Subjects were randomized within-session to either the *Cue-Story Similarity* condition, the *Story Similarity* condition or the *Baseline* condition

**Results.** As specified in the pre-analysis plan (`https://aspredicted.org/v7hh6.pdf`), we focus on the recall data as immediate belief impact was expectably much stronger in *Baseline* than in *Cue-Story Similarity* (as was indeed the case in our data). Figure 7 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 only marginally significant ($p = 0.1$) and smaller in magnitude than the effects of story similarity or the number of product scenarios. Table A.4 confirms this pattern using regression analysis, and for completeness, also reports results on belief impact, which are harder to interpret in light of the baseline differences in the immediate belief impact.

## 5.4 Cue similarity

In a final step, we study the role of the similarity of cues from different scenarios. While we manipulated cross-similarity via the overall number of scenarios and the similarity of different stories above, we will complete our analysis of different dimensions of similarity by studying the similarity of cues.

**Design.** We conducted an additional experiment in which we 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. Subjects either received a story or a statistic in the restaurant scenario. In *Cue Similarity*, we keep everything identical to *Baseline*, including the target cue Restaurant A, but change the labels of the decoy cues to Restaurant B and Restaurant C. In our analysis, as pre-registered, we compare belief impact 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.[27]

The pre-registration for this experiment can be found here: `https://aspredicted.org/h2fr3.pdf`, and a before includes the experimental design, predictions and analysis, sample sizes, and exclusion criteria.

**Prediction.** Increasing the similarity between target cue and decoy cues increases cross-similarity $S(\overline{C}, C)$. Following equation (5), this holds similarly for stories and statistics. As a consequence, the recall probability of both a target story and target statistic should decrease as we boost cue similarity (see Appendix F for a formal proof).

**Results.** Panel A of Figure 8 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). Panel B of Figure 8 largely displays the same pattern using our recall data. Formal regression analysis in Table A.5 confirms this result.

**Result 5.** *An increase in the similarity of cues decreases belief impact in delay, both for stories and (less pronounced) for statistics.*

# 6 Discussion and Conclusion

Although ubiquitous in practice, little is known about how people learn from qualitative information such as stories or anecdotes. In this paper, using controlled experiments, we document a story-statistic gap in memory. As time passes, the effect of stories on beliefs is much more persistent than that of statistics. Put differently, the belief impact of stories decays at a much higher speed. Using recall data, we show that stories are more easily retrieved from memory that statistics. We causally show that this pattern is driven by

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

the rich contextual features of stories: adding context to statistics increases delayed belief impact and recall accuracy of statistics. Guided by a simple memory model of similarity and interference, we experimentally examine the explanatory power of different features of interference. Consistent with the model, our evidence suggests that similarity relationships are the key driving force behind the the story-statistic gap, while the effect of memory load per se is – in comparison – more second order. Stories tend to be distinct and tightly associated to cues, whereas the abstract nature of statistics makes them similar to other unrelated statistics and not inherently similar to cues.

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

Our findings have immediate implications for the communication of statistical information. Our results suggest that if policymakers, marketers or leaders wish to convey statistical information, they should complement it with contextual, anecdotal associations to ensure that the information sticks with the audience, be that households, consumers or employees. For instance, statistical information about economic quantities should be coupled with anecdotal information that is consistent and inherently reminiscent of the embedded statistical information.

## References

**Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar**, "Overreaction in expectations: Evidence and theory," *Available at SSRN*, 2020.

**Alesina, Alberto, Armando Miano, and Stefanie Stantcheva**, "Immigration and redistribution," *Review of Economic Studies*, 2022.

**Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart**, "Narratives about the Macroeconomy," 2022.

**Bénabou, Roland, Armin Falk, and Jean Tirole**, "Narratives, Imperatives, and Moral Reasoning," Working Paper 24798, National Bureau of Economic Research, July 2018.

**Bordalo, Pedro, Giovanni Burro, Katie Coffman, Nicola Gennaioli, and Andrei Shleifer**, "Imagining the Future: Memory, Simulation and Beliefs about Covid," *Working Paper*, 2021.

**_ , John J Conlon, Nicola Gennaioli, Spencer Yongwook Kwon, and Andrei Shleifer**, "Memory and probability," *Working Paper*, 2021.

**_ , 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," *The Quarterly journal of economics*, 2021.

**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," *Working Paper*, 2021.

**_ , Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth**, "Justifying Dissent," NBER Working Paper 27288 May 2022.

**Charles, Constantin**, "Memory and Trading," *Available at SSRN 3759444*, 2021.

**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," *The Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

_ **and Florian Zimmermann**, "Correlation neglect in belief formation," *The Review of Economic Studies*, 2019, *86* (1), 313–332.

_ **, Frederik Schwerter, and Florian Zimmermann**, "Associative memory and belief formation," Technical Report, National Bureau of Economic Research 2020.

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

**Gennaioli, Nicola and Andrei Shleifer**, "What comes to mind," *The Quarterly journal of economics*, 2010, *125* (4), 1399–1433.

**Graeber, Thomas**, "Inattentive inference," *Journal of the European Economic Association*, 2022.

**Kahana, Michael Jacob**, *Foundations of human memory*, OUP USA, 2012.

**Kendall, Chad W and Constantin Charles**, "Causal narratives," Technical Report, National Bureau of Economic Research 2022.

**Kwon, Spencer Yongwook and Johnny Tang**, "Extreme Events and Overreaction to News," *Available at SSRN 3724420*, 2020.

**Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa**, "Failures in contingent reasoning: The role of uncertainty," *American Economic Review*, 2019, *109* (10), 3437–74.

**Michalopoulos, Stelios and Melanie Meng Xue**, "Folklore," *The 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," *Available at SSRN 3919471*, 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.

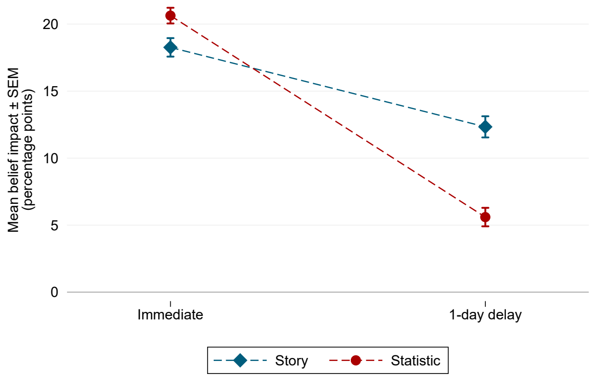
**Shiller, Robert J**, "Narrative economics," *American Economic Review*, 2017, *107* (4), 967–1004.

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

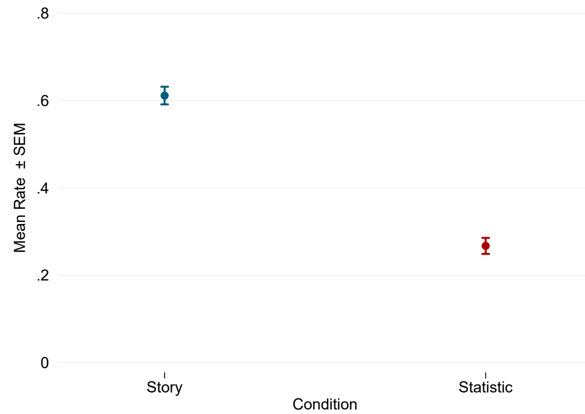
**Zimmermann, Florian**, "The dynamics of motivated beliefs," *American Economic Review*, 2020, *110* (2), 337–61.

# Figures

**Figure 1:** The story-statistic gap in memory

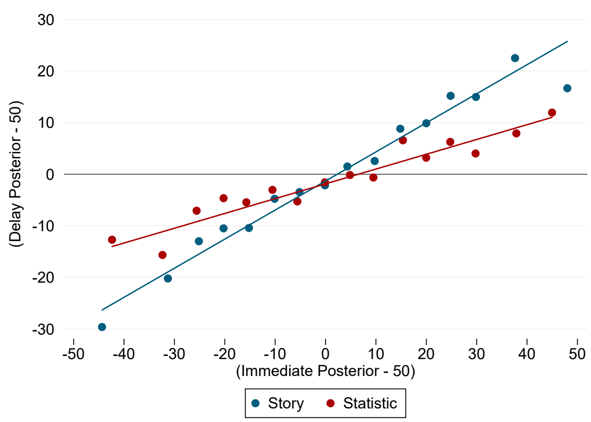



**(a)** Belief impact in *Immediate* and *Delay*

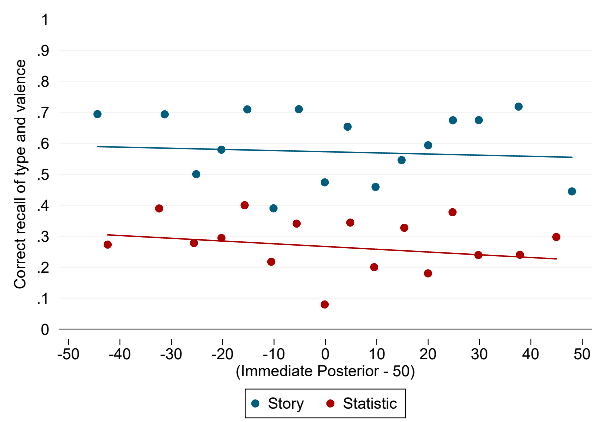



**(b)** Combined recall of type and valence

**Figure 2:** Heterogeneity by extremity of immediate updating

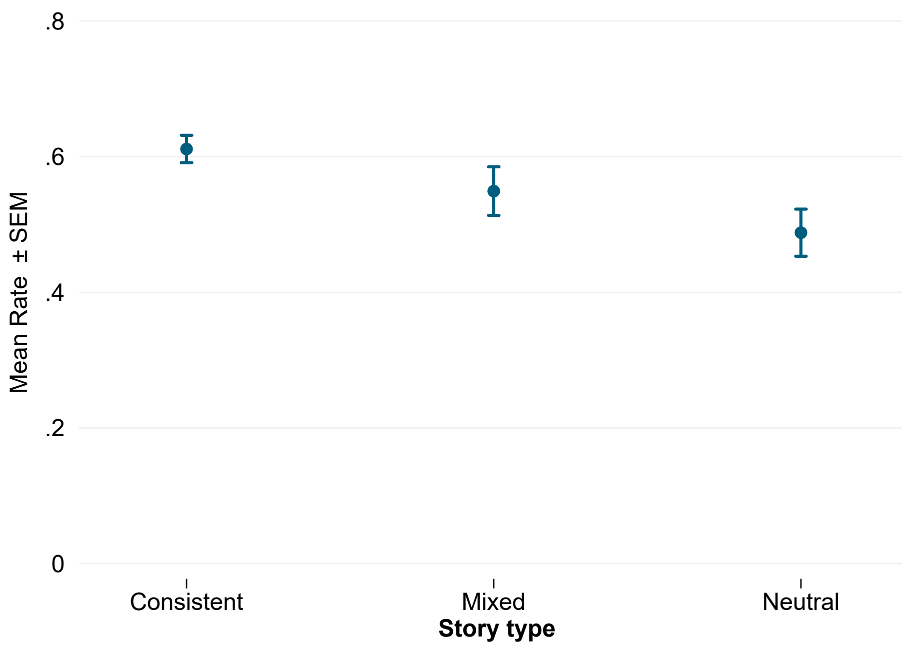



**(a)** Extremity of belief updating



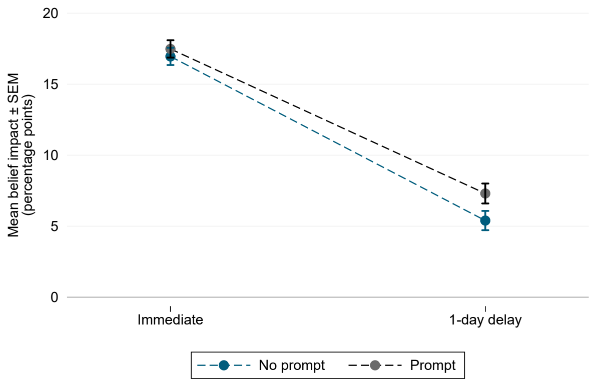

**(b)** Combined recall probability

**Figure 3:** Correct recall of type and valence by story type

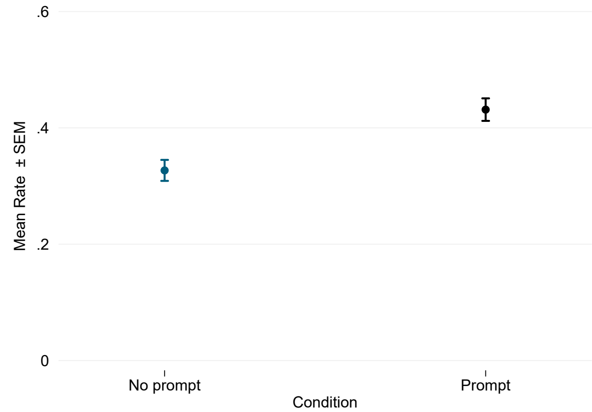



*Notes:* This Figure displays correct recall by story type.

**Figure 4:** Associations and contextual information: belief impact and recall

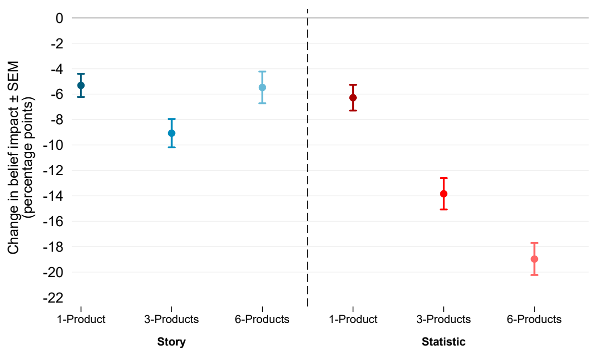



**(a)** Change in belief impact

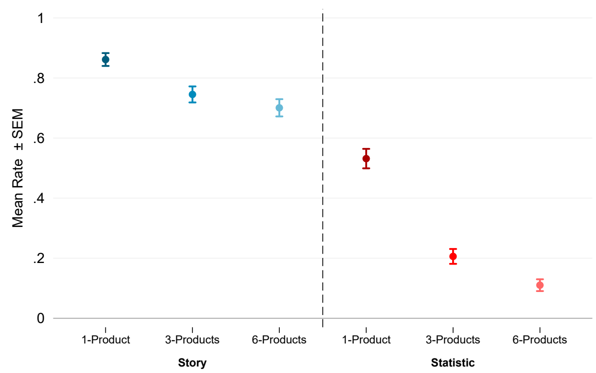



**(b)** Combined recall of type and valence

**Figure 5:** The story-statistics gap by number of products

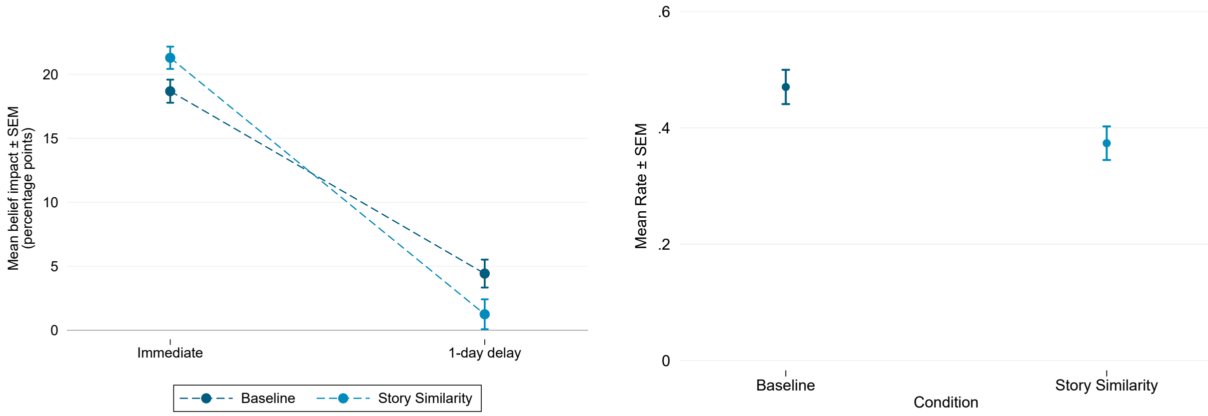



**(a)** Change in belief impact




**(b)** Combined recall of type and valence

**Figure 6:** Story similarity and belief impact




**(a)** Change in belief impact

**(b)** Combined recall of type and valence

**Figure 7:** Combined recall of type and Valence: cue-story similarity

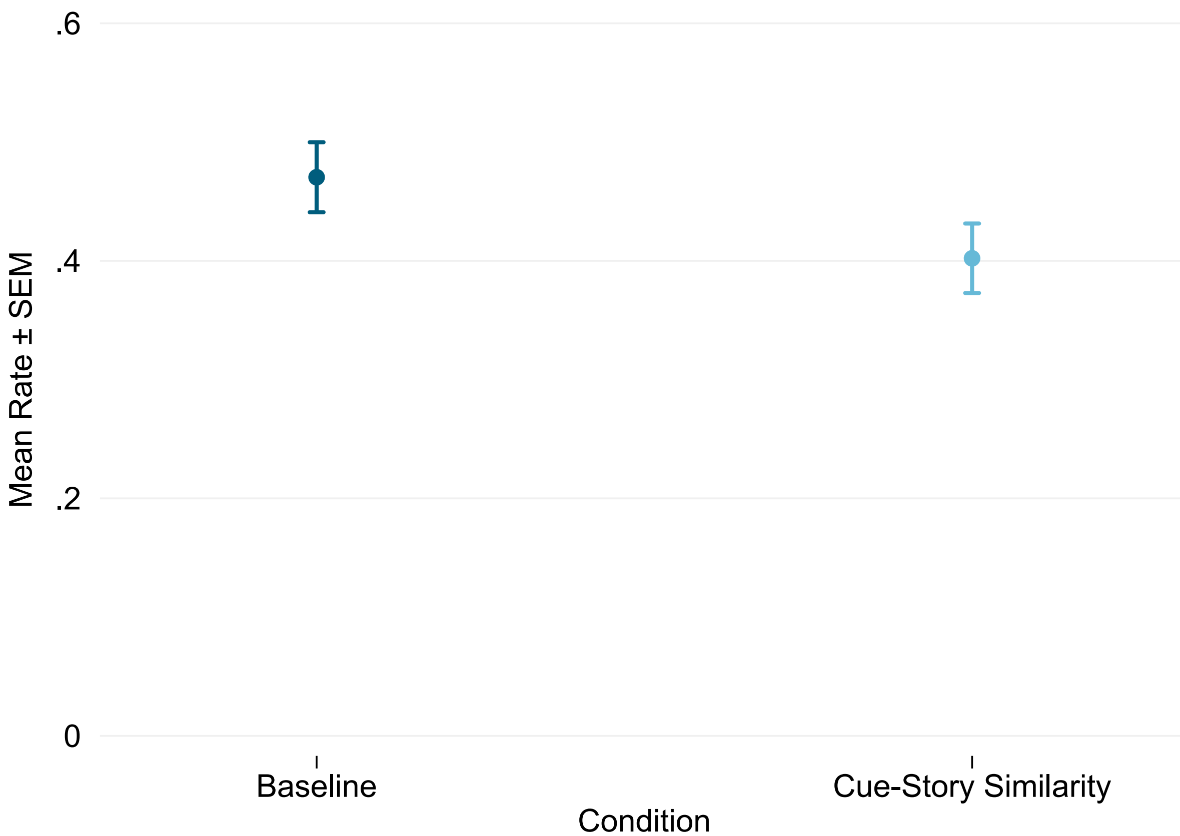
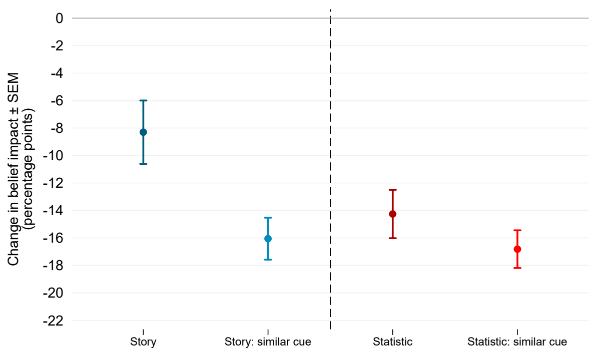
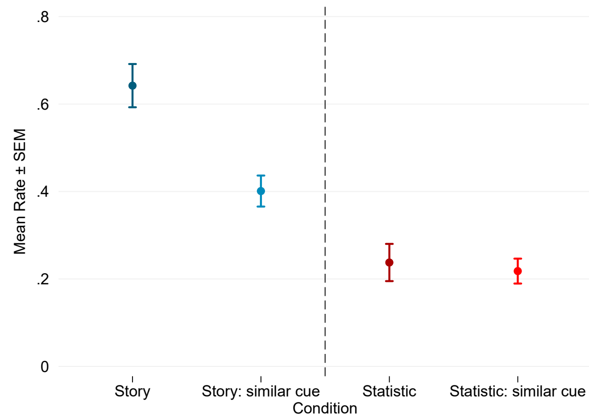


**Figure 8:** Cue similarity




**(a)** Change in belief impact




**(b)** Combined recall of type and valence

# Tables

**Table 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 1: The role of Decoy Information | Prolific (n=1,513) | 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` |
| Robustness Experiment 2: 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` |
| **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 bar story is about 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` |

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 2:** The story-statistics gap in memory

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | Belief Impact | | | Recall combined | |
| *Sample:* | Immediate<br>(1) | Delay<br>(2) | Pooled<br>(3) | Consistent<br>(4) | Story<br>(5) |
| Story | -2.37***<br>(0.87) | 6.73***<br>(1.05) | -2.37***<br>(0.84) | 0.34***<br>(0.02) | |
| Delay | | | -15.0***<br>(0.90) | | |
| Story × Delay | | | 9.10***<br>(1.28) | | |
| Neutral Story | | | | | -0.12***<br>(0.04) |
| Mixed Story | | | | | -0.062<br>(0.04) |
| Control Mean | 20.63 | 5.60 | 20.63 | 0.27 | 0.61 |
| Observations | 1168 | 1168 | 2336 | 1168 | 984 |
| $R^2$ | 0.54 | 0.52 | 0.43 | 0.64 | 0.01 |

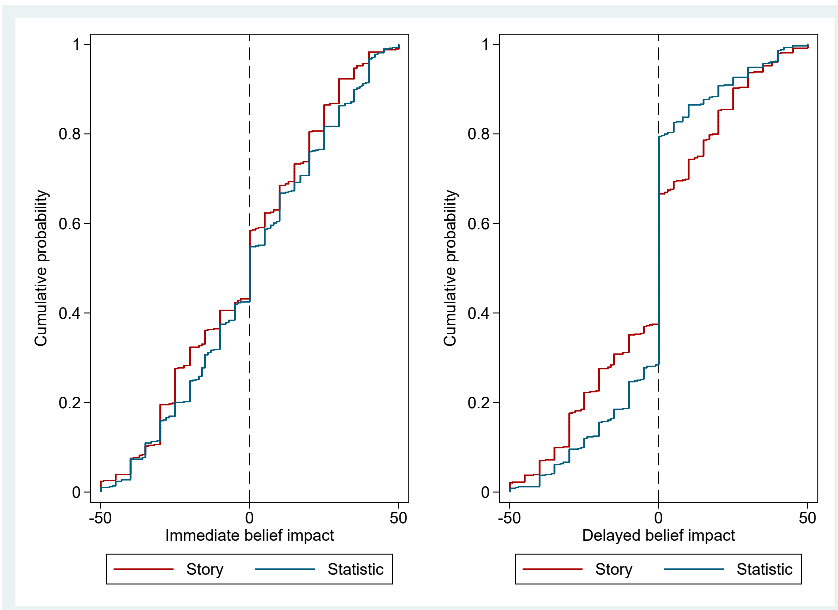
*Notes.* OLS estimates, robust standard errors in parentheses. Columns (1), (2) and (4) include respondents who received consistent stories. Column (3) pools *Immediate* and *Delay*. Column (5) includes observations who received stories. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix: Stories, Statistics and Memory

Thomas Graeber   Christopher Roth   Florian Zimmermann

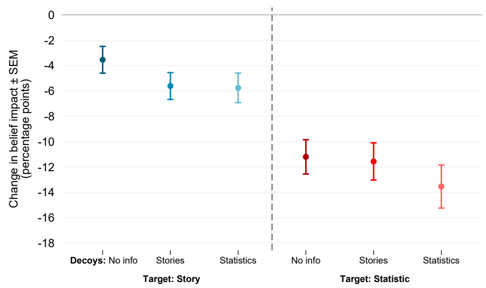
# A Additional Figures

**Figure A.1:** CDFs: belief impact

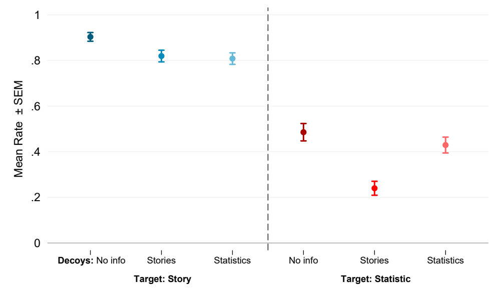



*Notes:* Empirical cumulative distribution functions (CDFs) of belief impact in the *Immediate* (left) and *Delay* (right) conditions. The data is from the baseline study.

**Figure A.2:** The story-statistic gap by decoy information

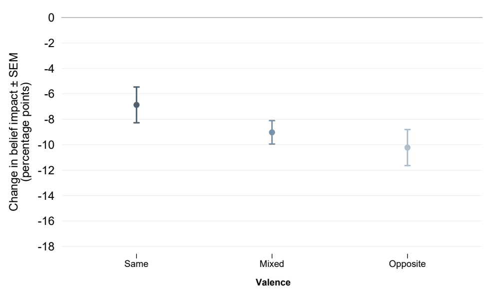



**(a)** Change in belief impact by decoy types

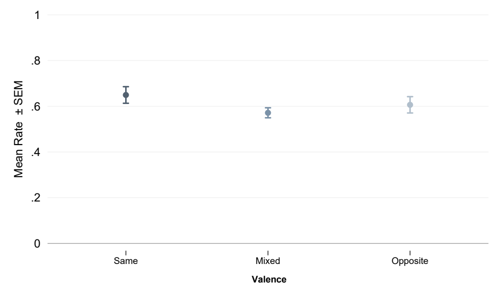



**(b)** Combined recall of type and valence

**Figure A.3:** The story-statistic gap by decoy Valence

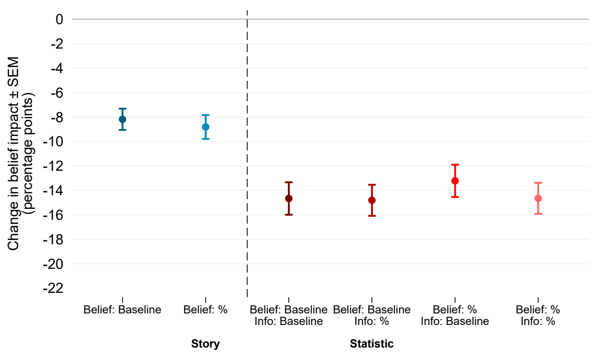



**(a)** Belief impact by decoy valence

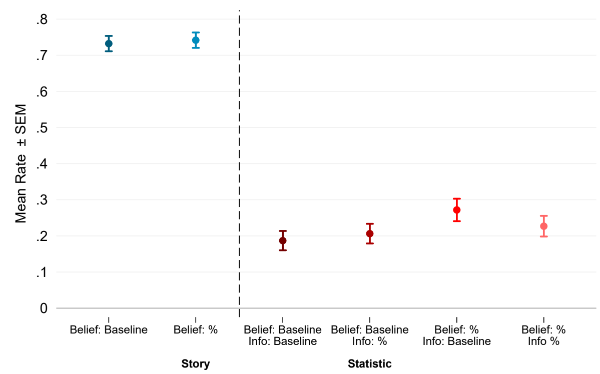



**(b)** Combined recall of type and valence

**Figure A.4:** The story-statistic gap by question format




**(a)** Change in belief impact by format




**(b)** Combined recall of type and valence

# B Additional Tables

**Table A.1:** Associations and contextual information: belief impact and recall

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Belief Impact | | | Combined Recall | | |
| *Sample:* | Pooled (1) | Stat (2) | NoStory (3) | Pooled (4) | Stat (5) | NoStory (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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.2:** Question format: belief impact and recall

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Belief Impact | | Combined Recall | |
| *Sample:* | Story (1) | Stat (2) | Story (3) | Stat (4) |
| Belief: % | 1.95* | 0.53 | 0.0094 | 0.085** |
| | (1.10) | (1.27) | (0.03) | (0.04) |
| Delay × Belief: % | -0.63 | 1.45 | | |
| | (1.31) | (1.88) | | |
| Info: % | | 1.98 | | 0.019 |
| | | (1.30) | | (0.04) |
| Delay × Info: % | | -0.15 | | |
| | | (1.84) | | |
| Info: % × Belief: % | | -1.68 | | -0.064 |
| | | (1.78) | | (0.06) |
| Delay × Info: % × Belief: % | | -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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.3:** 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 | 2.26 | 0.12*** | 0.33*** |
| | (1.39) | (1.44) | (0.03) | (0.04) |
| Delay × 1 Product | 3.76*** | 7.52*** | | |
| | (1.44) | (1.59) | | |
| 6 Products | -1.44 | 2.76** | -0.045 | -0.096*** |
| | (1.49) | (1.38) | (0.04) | (0.03) |
| Delay × 6 Products | 3.60** | -5.13*** | | |
| | (1.68) | (1.76) | | |
| Delay | -9.07*** | -13.8*** | | |
| | (1.12) | (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 respondent level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.4:** (Cue-)story similarity

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Belief Impact | | Combined Recall | |
| *Sample:* | Story | Cue-Story | Story | Cue-Story |
| | (1) | (2) | (3) | (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.* OLS estimates, standard errors clustered at the respondent level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.5:** 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 respondent level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.6:** Summary statistics

| | Baseline Experiments | | | Mechanisms | | | |
|---|---|---|---|---|---|---|---|
| *Experiment:* | Baseline (1) | Decoy (2) | Format (3) | Association (4) | Product (5) | Story Sim (6) | Cue Sim (7) |
| Male | 0.541 | 0.506 | 0.507 | 0.560 | 0.496 | 0.506 | 0.528 |
| Age (years) | 39.782 | 40.902 | 37.090 | 39.851 | 37.351 | 40.589 | 36.367 |
| College | 0.611 | 0.645 | 0.626 | 0.596 | 0.619 | 0.676 | 0.611 |
| Employed | 0.258 | 0.215 | 0.236 | 0.254 | 0.221 | 0.229 | 0.240 |
| Observations | 985 | 1,548 | 922 | 666 | 1,018 | 849 | 599 |

*Notes.* Summary statistics. We include all subjects who completed study 2.

**Table A.7:** Attrition by conditions

| | *Dependent variable:* | | | | | | |
|---|---|---|---|---|---|---|---|
| | Wave 2 Completion | | | | | | |
| *Experiment:* | Baseline | Decoy | Format | Association | Product | Story Sim | Cue Sim |
| | (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) | | | | | |
| Belief: % | | | 0.016 | | | | |
| | | | (0.03) | | | | |
| Info: % | | | 0.021 | | | | |
| | | | (0.03) | | | | |
| Prompt | | | | 0.0065 | | | |
| | | | | (0.05) | | | |
| 1 Product | | | | | -0.0046 | | |
| | | | | | (0.02) | | |
| 6 Products | | | | | -0.016 | | |
| | | | | | (0.03) | | |
| Story Similarity | | | | | | -0.017 | |
| | | | | | | (0.03) | |
| Cue Similarity | | | | | | -0.019 | |
| | | | | | | (0.03) | |
| Similar Cue | | | | | | | 0.020 |
| | | | | | | | (0.03) |
| Mean Completed | 0.69 | 0.76 | 0.60 | 0.46 | 0.42 | 0.79 | 0.59 |
| Observations | 1437 | 2048 | 1532 | 1442 | 2422 | 1069 | 1018 |
| p(Joint Null) | 0.80 | 0.60 | 0.59 | 0.90 | 0.85 | 0.79 | 0.53 |
| $R^2$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

*Notes.* OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# C Overview of stories

## C.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.

## C.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.

# D Implementation Details on the Experiments

## D.1 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 $s \sim \mathcal{U}[0, 0.5]$ if the valence is negative and $s \sim \mathcal{U}[0.5, 1]$ 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.

# E 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.8:** Coding Manual for data on open-ended recall

| Category | Explanation | Examples |
|---|---|---|
| **Lack of memory** | Statement that subjects doesn't recall whether and what information they received. This includes instances in which subject 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 subjects indicating "they don't remember" and "they remember getting no additional information", e.g., when just stating "None". It can help looking at the subject's 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 "I remember reading information about a person's review which was really positive." | "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 independent 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 talk about what they remember for a different scenario. | |
| **Misremembering across scenarios** | Each subjects gave three responses that are in adjacent rows in the Excel file. This category should be coded if the subject's 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 didn't 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 different data collections.

# F Formal Memory Framework

In the following we rely on our framework from Section 4. We first introduce parts of this framework more formally and then proof our key predictions.

## F.1 Notation

**Sets of memories** Let $M$ be the set of all memory traces. The set of memories can be split into cued memories $C$ and non-cued memories $M \setminus C = \overline{C}$. The set of memory traces can also be split into memories made during the experiment $E$ and memories made outside the experiment $M \setminus E = \overline{E}$. Memories made during the experiment can be split into memory traces being part of the cued set $C$ and memory traces made during the experiment, but not being part of the cued set $E \setminus C$.

Participants are asked to assess the probability of a randomly drawn review being positive for several products. We introduce $C'$ as the memories being cued, when having a different product as cue.

We introduce $C_{Story}$ as the cued set of memories, when having a story as additional information and $C_{Statistic}$, when having a statistic as additional information.

**Recall** Now that we introduced the different sets of memories, we can rewrite the probability of recalling a memory $e_0$ when given a certain cue in several ways. We can either split memories into cued and non-cued memories and rewrite the recall probability as:

$$
\begin{aligned}
r(e_0, C) &= \frac{S(e_0, C)}{\sum_{e \in C} S(e, C) + \sum_{e \in \overline{C}} S(e, C)} && (6) \\
&= \frac{S(e_0, C)}{S(C, C) \cdot \|C\| + S(\overline{C}, C) \cdot \|\overline{C}\|} && (7)
\end{aligned}
$$

We can also further split the non-cued memories into memories of other product scenarios and memories from outside the experiment.

$$r(e_0, C) = \frac{S(e_0, C)}{\sum_{e \in C} S(e, C) + \sum_{e \in E \setminus C} S(e, C) + \sum_{e \in \overline{E}} S(e, C)} \tag{8}$$

$$= \frac{S(e_0, C)}{S(C, C) \cdot \|C\| + S(E \setminus C, C) \cdot \|E \setminus C\| + S(\overline{E}, C) \cdot \|\overline{E}\|} \tag{9}$$

The probability to retrieve a cued memory is given as the sum over the probabilities of recalling a memory belonging to the cued set of memories:

$$r(C, C) = \frac{\sum_{e \in C} S(e, C)}{S(C, C) \cdot \|C\| + S(E \setminus C, C) \cdot \|E \setminus C\| + S(\overline{E}, C) \cdot \|\overline{E}\|} \tag{10}$$

$$= \frac{S(C, C) \cdot \|C\|}{S(C, C) \cdot \|C\| + S(E \setminus C, C) \cdot \|E \setminus C\| + S(\overline{E}, C) \cdot \|\overline{E}\|} \tag{11}$$

$$= \frac{S(C, C) \cdot \|C\|}{S(C, C) \cdot \|C\| + S(\overline{C}, C) \cdot \|\overline{C}\|} \tag{12}$$

We introduce a short notation for the recall probability:

We define the self-similarity $A(C, C)$ of the cued set as the average similarity between two memories in the cued set, multiplied with the size of the cued set:

$$A(C, C) = \sum_{e \in C} S(e, C) = S(C, C) \cdot \|C\| \tag{13}$$

Increasing the number of cued memories made during a product scenario or changing two memories to be more similar to each other increases the expression.

We define the cross-similarity $B(C, \overline{C})$ as the average similarity of cued memories to non-cued memories multiplied with the number of non-cued memories:

$$B(C,\overline{C}) = S(\overline{C},C) \cdot \|\overline{C}\| \tag{14}$$
$$= S(E \setminus C, C) \cdot \|E \setminus C\| + S(\overline{E},C) \cdot \|\overline{E}\| \tag{15}$$

With these two expressions we have a short notation for the recall probability:

$$r(C,C) = \frac{A(C,C)}{A(C,C) + B(C,\overline{C})} \tag{16}$$

## F.2 Proofs

In the following we provide formal proofs for all our predictions. For simplicity, we do so focusing on our secondary outcome measure, recall. In Section F.3 we clarify the relation between recall and $t = 2$ beliefs and show that beliefs follow in a straightforward way from what is being recalled.

### F.2.1 Story vs. Statistic

Changing the additional information from a statistic to a story, is the same as adding cued memories with more specific content. So memories created in a scenario with a story can be split in a statistical part $C_{\text{Statistic}}$, i.e. the set of memory traces that are the same for statistics and stories (Title, Hypergeometric statistic) and a qualitative part $C_{\text{Qualitative}}$, i.e. additional target traces with more detailed content, only present when having a story. This can be summarized by $C_{\text{Story}} = C_{\text{Statistic}} \dot\cup C_{\text{Qualitative}}$. The statistical part is almost identical to the memories made when having a statistic, the only difference being the numbers inserted into the statistic $1_{(m,n)}$. The second part adds the content-features.

**Assumption 1**:

(i) $S(C_{\text{Qualitative}}, C_{\text{Qualitative}}) > S(C_{\text{Statistic}}, C_{\text{Statistic}})$

(ii) $S(C_{\text{Qualitative}}, C_{\text{Statistic}}) > S(C_{\text{Statistic}}, C_{\text{Statistic}})$

The similarity of memories within a scenario with a statistic is only based on the two features 'Context-Experiment' and 'Name Product'.

Assumption a) can be justified, because the qualitative part is really self-similar. All memories in a story share content as well as context features. The memories have the same valence, are part of the same review, and share features of the specific experience.

Assumption b) can be justified, because the qualitative part matches the context as well as the quantitative information of the statistical part. Additionally the qualitative part is related to the product, not only by context but also by content, making it more similar to the memory trace encoding the title.

**Proposition 1** (Average similarity story vs. statistic)**.**

$$S(C_{Story}, C_{Story}) > S(C_{Statistic}, C_{Statistic}) \tag{17}$$

This just means that the average similarity of two cued traces in a product scenario with a story is higher than in a scenario with a statistic.

*Proof.* Let $C = C_0 \dot{\cup} C_1$ and $S(C_1, C_1) > S(C_0, C_0)$ and $S(C_1, C_0) > S(C_0, C_0)$ than we get the following inequality

$$\begin{aligned} S(C, C) \cdot |C| \cdot |C| &= S(C_0, C_0) \cdot |C_0| \cdot |C_0| + 2 \cdot S(C_0, C_1) \cdot |C_0| \cdot |C_1| + S(C_1, C_1) \cdot |C_1| \cdot |C_1| \\ &> S(C_0, C_0) \cdot (|C_0| \cdot |C_0| + 2 \cdot |C_0| \cdot |C_1| + |C_1| \cdot |C_1|) \\ &= S(C_0, C_0) \cdot |C| \cdot |C| \end{aligned}$$

It directly follows that $S(C, C) > S(C_0, C_0)$. ∎

Since a story consists of a statistical part as well as a qualitative part ,i.e. $C_{\text{Story}} = C_{\text{Statistic}} \dot{\cup} C_{\text{Qualitative}}$, stories create more memories than statistics:

$$\|C_{\text{Story}}\| > \|C_{\text{Statistic}}\| \tag{18}$$

Proposition 1 together with (18) lead to the following:

**Corollary 1** (Self-similarity story vs. statistic)**.**

$$A(C_{Story}, C_{Story}) = S(C_{Story}, C_{Story}) \cdot \|C_{Story}\| \tag{19}$$

$$> S(C_{Statistic}, C_{Statistic}) \cdot \|C_{Statistic}\| = A(C_{Statistic}, C_{Statistic}) \tag{20}$$

When changing the additional information from a statistic to a story, the set $\overline{C}$ of non-cued memories doesn't change. So we only have to pay attention to $S(C, \overline{C})$, the average similarity between cued and non-cued memories.

Statistics are way more generic than stories. Adding more specific features to memory traces makes them more similar to some memories (the ones sharing these specific features) and less similar to others (the ones not sharing these features). It seems very plausible to assume, that there are way more memories not sharing these specific features and therefore becoming less similar. This means that cross-similarity for stories is lower than for statistics:

**Assumption 3**:

$$B(C_{\text{Story}}, \overline{C}_{\text{Story}}) \leq B(C_{\text{Statistic}}, \overline{C}_{\text{Statistic}}) \tag{21}$$

Now we are ready to proof, that recall is higher for stories than statistics:

**Proposition 2** (Recall story vs. statistic)**.**

$$r(C_{Story}, C_{Story}) > r(C_{Statistic}, C_{Statistic}) \tag{22}$$

*Proof.* The probability to recall cued memories is higher for stories than for statistics if and only if:

$$
\begin{aligned}
r(C_{\text{Story}}, C_{\text{Story}}) = & \frac{A(C_{\text{Story}}, C_{\text{Story}})}{A(C_{\text{Story}}, C_{\text{Story}}) + B(C_{\text{Story}}, \overline{C}_{\text{Story}})} \\
> & \frac{A(C_{\text{Statistic}}, C_{\text{Statistic}})}{A(C_{\text{Statistic}}, C_{\text{Statistic}}) + B(C_{\text{Statistic}}, \overline{C}_{\text{Statistic}})} = r(C_{\text{Statistic}}, C_{\text{Statistic}})
\end{aligned}
$$

This is equivalent to the following inequality:

$$
A(C_{\text{Story}}, C_{\text{Story}}) \cdot B(C_{\text{Statistic}}, \overline{C}_{\text{Statistic}}) > A(C_{\text{Statistic}}, C_{\text{Statistic}}) \cdot B(C_{\text{Story}}, \overline{C}_{\text{Story}}) \tag{23}
$$

Corollary 1 and Assumption 3 together now complete the proof. ■

#### F.2.2 Number, type and valence of decoy information

**Proposition 3.** *Having more decoy scenarios with additional information decreases recall*

*Proof.* More decoy information leads to more memories created in other product scenarios. This means it increases $\|E \setminus C\|$. Like we have seen before

$$
\frac{\delta r(C, C)}{\delta \|E \setminus C\|} < 0 \tag{24}
$$

This means it decreases recall. ■

**Proposition 4.** *Recall is lowest for the type of additional information with the highest similarity to the additional information of the target scenario.*

*Proof.* Changing the type of decoy information has an effect on recall through the similarity of cued memories and memories belonging to decoys, i.e. through $S(C', C)$. This has an impact on recall through $S(E \setminus C, C)$:

$$
S(E \setminus C, C) \cdot \|E \setminus C\| = S(E \setminus (C \cup C'), C) \cdot \|E \setminus (C \cup C')\| + S(C', C) \cdot \|C'\| \tag{25}
$$

The recall depends on $S(C', C)$ in the following way:

$$\frac{\delta r(C,C)}{\delta S(C',C)} = \frac{-A(C,C)\cdot \|C'\|}{(A(C,C)+B(C,C'))^2} < 0 \tag{26}$$

Recall decreases with the similarity of target information and decoy information and therefore is lowest for the type of information with the highest similarity.

■

**Assumption 4**:

$$S(C'_{Statistic}, C_{Statistic}) > S(C'_{Story}, C_{Statistic}) \tag{27}$$

**Corollary 2.** *Statistics as decoys lead to higher interference for statistics*

*Proof.* This directly follows from Assumption 4 and Proposition 4. ■

We cannot make an analogous assumption about similarity for stories. The ordering depends on how similar the two products/stories are. If they are not too similar Proposition 4 tells us that a statistic as a decoy will lead to higher interference, than the story does. If they are really similar it will be the other way around.

**Proposition 5.** *Decoys with the same valence as the target scenario decreases recall*

*Proof.* If decoys have the same valence as the target scenario, they share more features, leading to a higher similarity. This means it increases $S(E \setminus C, C)$. The recall depends on the similarity in the following way:

$$\frac{\delta r(C,C)}{\delta S(E\setminus C,C)} = \frac{-A(C,C)\cdot \|E\setminus C\|}{(A(C,C)+B(C,E\setminus C,\overline{E}))^2} < 0 \tag{28}$$

This means it decreases recall ■

### F.2.3 Prompting Contextual Features

**Proposition 6.** *Adding memories to the target scenario increases recall.*

*Proof.* If we add memories to the cued set, this means increasing $\|C\|$.

$$\frac{\delta r(C,C)}{\delta\|C\|} = \frac{S(C,C)\cdot B(C,\overline{C})}{(A(C,C)+B(C,\overline{C}))^2} > 0 \tag{29}$$

So the probability to recall a cued memory is higher if the number of cued memories increases. ■

### F.2.4 The Number of Product Scenarios

**Proposition 7.** *Adding decoy product scenarios decreases recall.*

*Proof.* Increasing the number of product scenarios translates into increasing $\|E \setminus C\|$, which has the following effect on recall:

$$\frac{\delta r(C,C)}{\delta\|E \setminus C\|} = \frac{-A(C,C)\cdot S(E\setminus C,C)}{(A(C,C)+B(C,C'))^2} < 0 \tag{30}$$

If we increase the number of product scenarios without changing their average similarity to the cued traces, then this reduces the probability of recall, more so if they are on average more similar to the cued traces. ■

The total effect might be different for stories and statistics.

The effect for stories is smaller if and only if:

$$\frac{A(C_{\text{Story}},C_{\text{Story}})\cdot S(E\setminus C_{\text{Story}},C_{\text{Story}})}{(A(C_{\text{Story}},C_{\text{Story}})+B(C_{\text{Story}},\overline{C}_{\text{Story}}))^2} < \frac{A(C_{\text{Statistic}},C_{\text{Statistic}})\cdot S(E\setminus C_{\text{Statistic}},C_{\text{Statistic}})}{(A(C_{\text{Statistic}},C_{\text{Statistic}})+B(C_{\text{Statistic}},\overline{C}_{\text{Statistic}}))^2} \tag{31}$$

It is not quite clear in what direction this goes. Stories have a higher self similarity i.e. $A(C_{\text{Story}},C_{\text{Story}}) > A(C_{\text{Statistic}},C_{\text{Statistic}})$. On the other hand, the interference is lower, i.e. $B(C_{\text{Story}},\overline{C}_{\text{Story}}) \leq B(C_{\text{Statistic}},\overline{C}_{\text{Statistic}})$. In most cases the similarity to decoys is higher for statistics, i.e. $S(E\setminus C_{\text{Statistic}},C_{\text{Statistic}}) > S(E\setminus C_{\text{Story}},C_{\text{Story}})$.

#### F.2.5 Story Similarity

**Proposition 8.** *Increasing the story similarity decreases recall*

*Proof.* Since stories are part of the cued sets, making the stories more similar leads to an increase of $S(E \setminus C, C)$. Like we've seen before:

$$\frac{\delta r(C,C)}{\delta S(E \setminus C, C)} < 0 \tag{32}$$

So this leads to a decrease in recall. ■

#### F.2.6 Cue-story Similarity

**Proposition 9.** *Increasing cue-story similarity increases recall*

*Proof.* Increasing Cue-story similarity means to make the story more related to the product itself. This means it increases the similarity of the memory traces belonging to the story to the memory trace encoding the title. This increases the average self-similarity $S(C, C)$. Since

$$\frac{\delta r(C,C)}{\delta S(C,C)} = \frac{B(C,\overline{C}) \cdot \|C\|}{(A(C,C) + B(C,\overline{C}))^2} > 0 \tag{33}$$

this increases recall. ■

#### F.2.7 Cue Similarity

**Proposition 10.** *More similar cues lead to lower recall*

*Proof.* The cue is composed of the context of the experiment and the product. A higher cue similarity means to make the two products more similar. This means to increase the average similarity of memory traces having one of the products as a feature. Let $P_1$ be the set of memories having product 1 as a feature and $P_2$ be the set of memories having product 2 as a feature. The similarity of the two products is then given by:

$$S(P_1, P_2) = \frac{1}{\|P_2\|} \cdot \sum_{e' \in P_2} S(e', P_1) = \frac{1}{\|P_1\| \cdot \|P_2\|} \cdot \sum_{e' \in P_2} \sum_{e \in P_1} S(e', e) \tag{34}$$

So similarity depends on background memories. Since every memory trace of the cued set contains the product as a feature, the memory traces of the two cued sets become more similar. This means it increases $S(C, C')$. We've already seen the impact on the recall:

$$\frac{\delta r(C, C)}{\delta S(C', C)} < 0 \tag{35}$$

Increasing the cue similarity decreases recall. $\blacksquare$

## F.3 Beliefs in time period $t = 2$

We assume, that participants sample once. There is no confusion, i.e. participants realize whether they retrieved a cued memory or not and once they retrieve a memory they remember every detail of it.

### F.3.1 Notation

Participants are asked to assess the probability of a randomly drawn review being positive. If the total number of reviews for a product $j$ is $r_j$ and the total number of positive reviews is $m_j$, the probability of sampling a positive review is $p_j = \frac{m_j}{r_j}$.

Let $b_{i,j}$ be the belief distribution over the total number of positive reviews $M_j$ and $\hat{b}_{i,j}$ the belief distribution over the probability to draw a positive review. If we fix the total number of reviews $r_j$, the two belief distributions have the following relationship:

$$\hat{b}_{i,j}(p) = b_{i,j}(p \cdot r_j) \tag{36}$$

This relationship allows us to focus on the belief distribution over the total number of positive reviews. We will denote the belief distribution over the total number of positive reviews of participant $i$ in time period $t$ by $b^t_{i,j}$. The time period can take on value 1 or 2. In both periods, participants can update their beliefs after potentially having received additional information. We will therefore use $b^0_{i,j}$ to denote the prior beliefs of participants, before having received any additional information.

### F.3.2 Prior belief

Participants are always informed about the total number of reviews $r_j$. The unknown variable is the total number of positive reviews $M_j$. Participants know that the number of positive reviews $M_j$ is randomly drawn. This means that participants prior beliefs follow a discrete uniform distribution with support $\{0, 1, \ldots, r_j\}$. This is equivalent to a beta-binomial distribution with parameters $\alpha = \beta = 1$.

$$M_j \sim BetaBin(r_j, 1, 1) \tag{37}$$

This translates into the density function:

$$b^0_{i,j}(m_j|r_j) = \frac{1}{r_j + 1} \text{ for } 0 \leq m_j \leq r_j \tag{38}$$

### F.3.3 Posterior belief

There are two possible cases. Either the participant retrieves a cued memory, which happens with probability $r(C, C)$. Or the participant retrieves a non-cued memory, which happens with probability $(1 - r(C, C))$.

We will now derive the posterior for both cases.

**First case** A participant retrieves a cued memory.

**Proposition 11.** *Let the total number of reviews for product $j$ be $r_j$. If a participant retrieves a cued memory, and the statistical information tells him/her, that $k_j$ out of $n_j$ reviews were* *positive, then he/she will state a belief of:*

$$\frac{(r_j - n_j)(k_j + 1)}{r_j \cdot (n_j + 2)} + \frac{k_j}{r_j} \tag{39}$$

*Proof.* The participant knows the total number of reviews $r_j$, as well as the sample size $n_j$ and the number of successes $k_j$. The total number of positive reviews $M_j$ is unknown, but participants form beliefs over the true value.

Like we mentioned before the prior is given by:

$$M_j \sim BetaBin(r_j, 1, 1) \tag{40}$$

Given that $m_j$ is the total number of positive reviews and $n_j$ reviews are drawn without replacement, the probability of having $k_j$ positive reviews follows a hypergeometric distribution:

$$\pi(k_j|n_j, m_j, r_j) = \frac{\binom{m_j}{k_j} \cdot \binom{(r_j - m_j)}{(n_j - k_j)}}{\binom{r_j}{n_j}} \tag{41}$$

When updating according to Bayes' rule, the posterior belief over $M_j$ is given by:

$$b^2_{i,j}(m_j|r_j, (n_j, k_j)) = \binom{r_j - n_j}{m_j - k_j} \cdot \frac{B(m_j + 1, r_j - m_j + 1)}{B(k_j + 1, n_j - k_j + 1)} \tag{42}$$

Here $B(a, b)$ denotes the beta function.

When defining $\alpha' = k_j + 1$, $\beta' = n_j - k_j + 1$, $r'_j = r_j - n_j$ and $m'_j = m_j - k_j$ one can see, that the posterior follows a beta-binomial distribution:

$$M'_j \sim BetaBin(r'_j, \alpha', \beta') \tag{43}$$

with support $\{0, \ldots r'_j\}$. This is equivalent to $M_j \sim BetaBin(r'_j, \alpha', \beta') + k$ with support $\{k_j, \ldots r_j - n_j + k_j\}$.

The payoff is maximized when reporting the mean of the belief distribution:

$$E[M_j] = E[M_j'] + k = r_j' \cdot \frac{\alpha'}{\alpha' + \beta'} + k \tag{44}$$

$$= \frac{(r_j - n_j) \cdot (k_j + 1)}{n_j + 2} + k_j \tag{45}$$

This leads to a reported probability of:

$$E[p] = \frac{E[M_j]}{r_j} = \frac{(r_j - n_j)(k_j + 1)}{r_j \cdot (n_j + 2)} + \frac{k_j}{r_j} \tag{46}$$

So the expected value lies in between $\frac{k_j}{r_j}$ and $\frac{(r_j - n_j + k_j)}{r_j}$, which are the only possible values after having observed $k_j$ out of $n_j$ successes. ■

**Second case:** A participant retrieves a non-cued memory.

In this case, the participant has no information, except the total number of reviews $r_j$. The participant therefore relies on the prior.

$$M_j \sim BetaBin(r_j, 1, 1) \tag{47}$$

Again the participant states the probability maximizing the payoff, which is given by the mean, i.e. $\frac{1}{2}$.