

A Simple Model of Memory-Based Belief Formation

Benson Tsz Kin Leung^{*}.

September 8, 2023

Abstract

This paper presents a simple model of memory-based belief formation. An individual has a memory of signals, and upon receiving a prompt, recalls signals from his memory and forms belief using a naïve Bayes' formula. The recall probability of each signal depends on the cues of the signals and the prompt. The recall probability function can capture different stylized results of memory research, including how similarity, interference, salience, previous retrieval, etc. affect memory retrieval. I analyze three adaptations of the model and show that, first, varying recall probabilities w.r.t. likelihood ratio can lead to asymptotic erroneous belief. Second, the persuader prefers to send a less informative but more associative signal when the receiver has a large memory. Third, individuals overreact to new signals when they have a large memory, and will strengthen their strong prior beliefs even after receiving a belief-challenging signal.

Keywords: Memory, Belief, Learning, Persuasion

JEL codes: D83, D91

^{*}Hong Kong Baptist University. Email: btkleung@hkbu.edu.hk, ltkbenson@gmail.com.

1 Introduction

The role of memory in belief formation has been discussed and studied in many disciplines, including but not limited to Cognitive Science, Economics, and Psychology.¹ Recently, there has been abundant evidence, in both experimental and empirical studies, supporting the cognitive process in which individuals form belief by retrieving information from their memory (Chew et al. (2020); Enke et al. (2020); D’Acunto and Weber (2021); Graeber et al. (2022); Bordalo et al. (2023a,b)). This paper aims to propose a simple but flexible model of memory-based belief formation, that allows tractable analysis and could incorporate the many documented characteristics of memory retrieval in the literature.

In the baseline model of this paper, an individual has a memory of signals about some unknown state of the world, e.g., articles that she read, conversations that she had with friends and family, personal experiences, etc. Upon receiving a prompt, e.g., a question like “Do you believe that global warming is real?”, a suggestion like “Think about the winter that we just had!”, or new articles, etc., she recalls signals from her memory and forms her belief using a naïve Bayes’ formula.

The crucial feature of the model is that signals and the prompt are not only associated with a likelihood ratio but also a set of cues. The cues could represent their linguistic elements, the sender of the signals/prompts, their levels of salience, and other qualitative features. First, the role of the likelihood ratio is the same as in Bayesian updating. If a signal is recalled, the individual updates her belief using her prior and the likelihood ratio of the signal via the Bayes’ formula. On the other hand, the cues affect the recall probability of the signals. More specifically, the recall probability of a signal is a function of the cues of the prompt, the signal, and other signals in the individual’s memory. The recall probability function could capture different characteristics of memory retrieval documented across different disciplines. For example, the recall probability of a signal may increase in the similarity of its cues and the prompt: A prompt like “Think about the winter that we just had!” will likely induce retrieval of experiences that happened in previous winters. The individuals may recall signals that are more salient or surprising with a higher probability; signals that have been recalled recently may be more accessible due to the recency effect; and signals may interfere with each other in the retrieval process.

Given the flexibility of the recall probability function, the model encompasses existing models of memory-based belief formation in the literature, for example, Bordalo et al. (2020, 2022a,b, 2023a,b). Nonetheless, the baseline model is tractable and can be easily generalized to study different applied problems including persuasion and advertisement. In this paper, I analyze three slight variants of the baseline model.

¹Citing all relevant literature is impossible, For a few examples, see Wyer and Albarracín (2005); Li (2013); Connors and Halligan (2015).

In the first variant, I analyze how unbalanced recall probabilities with respect to likelihood ratio could lead to asymptotic wrong beliefs. A recall probability function is unbalanced with respect to the likelihood ratio if the individual recalls two signals with different likelihood ratios with different probabilities. For example, she may recall favorable signals with higher probabilities (Li (2013)), or she may recall more informative signals with higher probabilities (Leung (2020)). I show that when the recall probabilities are balanced, asymptotic belief converges to a mass point on the true state, as in standard Bayesian updating. In contrast, when the recall probabilities are sufficiently unbalanced, asymptotic belief could converge to a mass point on the wrong state, i.e., the individual believes that for sure the wrong state is true. Given that unbalanced recall probabilities are arguably natural, e.g., individuals pay more attention to very informative signals than some almost pure noise, this result shows that erroneous belief would be commonplace.

Next, I analyze an extension of the baseline model in which a politician decides how to convince a voter. The politician decides to send one of two prompts to the voter, afterwards the voter recalls signals from her memory and forms a belief on whether the politician is the better candidate. One of the prompts is more informative, such that a Bayesian voter would change her belief more in favor of the politician. The other prompt is more associative, such that it invokes more retrieval of signals that are in favor of the politician. For example, the politician could present data and plan to the voter to show how good a candidate she can be, or ask the voter to think about the mess the current government has put us in and think about the improvement she can bring. I show that the politician prefers the associative prompt if and only if the voter’s memory is large. The result shows how narrative could be used for persuasion even when it is uninformative (Shiller (2017)), and suggests a limit in social learning: as the voter collects more information over time, the politician would prefer to send uninformative but associative prompts and therefore slows down learning.

The last variant of the model looks into heterogeneous reactions to new information. As different individuals have different memories in different contexts, new information triggers the retrieval of different information, and heterogeneous reactions arise naturally. As an example, after a natural disaster, individuals recall their experience/stories of previous disasters, which leads to overreaction especially when the individuals have a large memory (Bordalo et al. (2020); Huang et al. (2020)). In contrast, in laboratory experimental settings, signals and the nature of uncertainty are often abstract. Therefore, new information in the lab would unlikely trigger memory retrieval and not lead to overreaction (Benjamin (2019)). Similarly, after seeing a new belief challenging signal, the individual retrieves previous signals that she received which will likely support her prior belief. Therefore, the individual may strengthen their prior belief even after seeing counter-attitudinal information (Bail

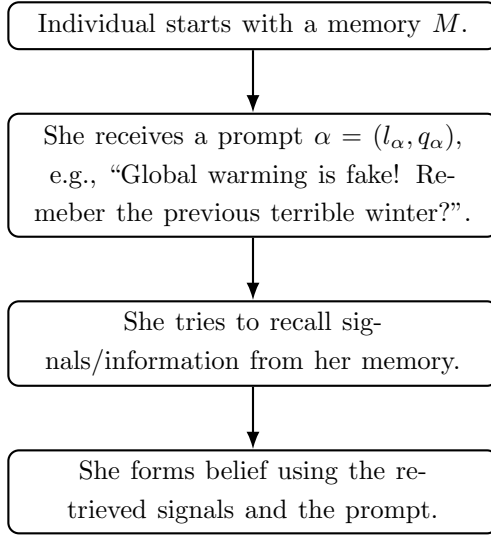


Figure 1: The belief formation process.

et al. (2018)).

2 Baseline Model

Consider a binary state space, $\omega = 1, 2$, and a uniform prior $p^1 = p^2 = \frac{1}{2}$. The model can be easily generalized to more than 2 states and with non-uniform prior. In this paper, I analyze a belief formation process described as follows and illustrated in Figure 1. First, an individual has a memory M that contains multiple signals $s \in M$. A signal s is a tuple (l, q) where $l = \frac{f^1(s)}{f^2(s)}$ is the likelihood ratio of receiving the signal in state 1 vis-a-vis state 2, f^ω is the p.d.f. in state ω , and $q \in Q$ is a set of cues associated with the signal. l is standard in models of Bayesian updating, while q is inspired by the memory literature. The set of cues could include the words used to convey the signal, the person who sends the signal, the level of salience, and other qualitative features.

Next, upon receiving a prompt denoted as $\alpha = (l_\alpha, q_\alpha)$, that $l_\alpha = \frac{f^1(\alpha)}{f^2(\alpha)}$ is the likelihood ratio of receiving the signal in state 1 vis-a-vis state 2 and $q_\alpha \in Q$ is a set of cues, the individual recalls signals from her memory probabilistically and together with the prompt, forms a belief over the state space using a naïve Bayes' formula. More specifically, if the individual successfully recalls signals $m \subseteq M$, her belief that state 1 is true, denoted by $\hat{p}^1$, follows:

$$\log \frac{\hat{p}^1}{1 - \hat{p}^1} = \log \frac{p^1}{1 - p^1} + K \log l_\alpha + \sum_{i: s_i \in m} \log l_i = K \log l_\alpha + \sum_{i: s_i \in m} \log l_i. \quad (1)$$

$K \in \mathbb{R}^+$ measures the initial underreaction or overreaction ($K < 1$ and $K > 1$ respectively) of belief to the prompt. Note that if $K = 1$, and that the individual successfully recalls all signals in her memory, i.e., $m = M$, Equation (1) is

equivalent to Bayesian updating. In contrast, I assume that the individual recalls memory probabilistically. Given the prompt α , the probability that the individual successfully recalls a signal $s_i \in M$ depends on q_α , q_i and q_{-i} , which I denote as $r(q_\alpha, q_i, q_{-i}) \in [0, 1]$. In the following, I provide some examples.

1. Similarity: the recall probability is higher if the signal and the prompt share more common cues, e.g., $r(q_\alpha, q_i, q_{-i}) = \frac{|q_\alpha \cap q_i|}{|q_\alpha|}$ where q_α, q_i are some subsets of all possible cues and $|q_\alpha| > 0$.
2. Salience: the recall probability is higher if the signal is more salient, e.g., $r(q_\alpha, q_i, q_{-i}) = q_i$ where $q_i \in [0, 1]$ measures the level of salience.
3. Competing attention and interference: for example, the recall probability of all signals in M summed to 1:

$$r(q_\alpha, q_i, q_{-i}) = \begin{cases} \frac{\tilde{r}(q_\alpha, q_i)}{\sum_{j: s_j \in M} \tilde{r}(q_\alpha, q_j)} & \text{if } \sum_{j: s_j \in M} \tilde{r}(q_\alpha, q_j) > 0; \\ \frac{1}{|M|} & \text{otherwise,} \end{cases}$$

where $\tilde{r}(q_\alpha, q_i) \geq 0$ for all q_α, q_i .

4. Thought-provoking prompt: if the prompt is more thought-provoking, e.g., when the sender is charismatic, or the prompt is more surprising or incentivizes individuals to exert more cognitive effort, e.g., $r(q_\alpha, q_i, q_{-i})$ increases in q_α where $q_\alpha \in \mathbb{R}^+$ measures how thought-provoking the prompt is.

As shown in the examples, this baseline model allows us to capture different characteristics of memory retrieval documented in the literature. In the following sections, I analyze variants of this baseline models to illustrate its applicability.

3 Varying/Invariant Recall Probability

There is abundant evidence suggesting that people pay more attention to, or recall with higher probability, information depending on their informativeness and the state it supports. For example, individuals recall favorable signals with higher probabilities than non-favorable signals, when they are self-related (Li (2013)). Individuals pay more attention to more informative signals or completely ignore relatively noisy signals due to limited attention (Wilson (2014); Schwartzstein (2014); Leung (2020)). And given that more attention is spent on informative signals, individuals subsequently recall them with a higher probability (Connors and Halligan (2015)). This section looks into how such varying recall probabilities impact asymptotic belief.

To focus on the impact of varying recall probabilities, I adapt the baseline model and assume that $q_i = l_i$ and that the recall probability of a signal depends only on its

likelihood ratio. With some abuse of notations, I use $r(l_i)$ to denote $r(q_\alpha, q_i, q_{-i})$. For example, individuals might recall more informative signals with higher probability, i.e., $r(l_i)$ increases in $\max\{l_i, 1/l_i\}$; or they might recall signals that support state 1 with higher probability, i.e., $r(l_i) > r(1/l_i)$ for $l_i > 1$.

Fix a function $r(l) : \mathbb{R}^{++} \rightarrow [0, 1]$, I say that r is balanced if $r(l)$ is constant across all $l \neq 1$, and is unbalanced if there exists $l, l' \in \mathbb{R}^{++} \setminus \{1\}$ such that $r(l) \neq r(l')$. The following Proposition analyzes how the balance of r affects asymptotic belief which I define as follows. Suppose that the individual's memory M contains T signals, i.e., $M = \{s_1, \dots, s_T\}$. Given the true state ω , s_t is drawn i.i.d. from the distribution $f^\omega(l)$ where $f^1(l) = lf^2(l)$. Now given $M = \{s_1, \dots, s_T\}$ and some prompt α , the individual recalls signals from M and form belief according to the naïve Bayes' formula, i.e., Equation (1). By the law of large number, as $T \rightarrow \infty$, under state ω , the individual's belief converges in probability to $\hat{p}_\infty^1$ where

$$\log \frac{\hat{p}_\infty^1}{1 - \hat{p}_\infty^1} = \log l_\alpha + \lim_{T \rightarrow \infty} \sum_{t=1}^T \int_l f^\omega(l) r(l) \log l \, dl,$$

which I call asymptotic belief under state ω .

Proposition 1. *If r is balance and strictly positive, for all signal distributions $f^1 \neq f^2$, $\hat{p}_\infty^1 = 1$ under state 1 and $\hat{p}_\infty^1 = 0$ under state 2. In contrast, if r is “sufficiently unbalance”, i.e., there exists some threshold $\epsilon > 0$, such that if*

$$r(l')/r(l) < \epsilon \text{ for some } l \neq l',$$

then there exists some signal distributions f^2 and $f^1 = lf^2$, $\hat{p}_\infty^1 = 0$ under state 1 or $\hat{p}_\infty^1 = 1$ under state 2.

Note that an unbalanced recall probability function does not necessarily bias toward one state versus the other. For example, while $r(l) = 1 - \frac{1}{\max\{l, l^{-1}\}}$ is unbalanced (“extremely unbalanced” as it vanishes as l goes to 1), it assigns equal recall probability to equally informative signals that support different states. It is only “biased” against less informative signals, and different theoretical studies have shown that ignoring marginally informative signals is optimal under limited cognitive resources (Wilson (2014); Leung (2020)). Nonetheless, there exist signal distributions where signals supporting state 1 are less informative than signals supporting state 2, such that individuals recall excessively more signals supporting state 2 even when state 1 is true. As a result, individuals eventually become convinced that the wrong state is true.

4 Persuasion: Informativeness or Association

This section looks into persuasion with memory-based belief formation. Suppose a politician tries to convince a voter to vote for her among two candidates. Should the politician try to convince the voter by presenting informative information, or should she try to invoke retrieval of memory that flavors her, possibly by easily associative narratives? Similarly, when a climate scientist tries to convince someone that global warming is severe, should she present detailed data and academic research, or should she ask the person to recall the wildfires and the increasingly warm winter in the previous years?

To answer this question, consider a voter with a memory of T signals, i.e., $M = \{s_1, \dots, s_T\}$. Similar to the previous section, I assume that $q_i = l_i$ for all signal realization, and $s_t = (l_t, q_t)$ is drawn i.i.d. with distribution $f^\omega(l)$ where $f^1(l) = lf^2(l)$. Different from the previous section, a politician decides one of the two prompts to send to the voter, $\alpha_I = (l_{\alpha_I}, q_{\alpha_I})$ or $\alpha_A = (l_{\alpha_A}, q_{\alpha_A})$, to maximize $E(\hat{p}_T^1)$. The politician wants to convince the voter that state 1 is true and maximize the expected voter's belief of state 1. For simplicity, I also assume that the politician knows what the true state is. Generalizing this assumption does not affect our result but makes exposition more cumbersome.

α_I is more informative than α_A in the sense that $l_{\alpha_I} > l_{\alpha_A} \geq 1$. In contrast, α_A is more associative to signals that support state 1: it improves the retrieval of information that supports state 1. For simplicity, I assume that the recall probability of a signal only depends on the cues of the prompt and the likelihood ratio of the signal. Mathematically, given that l is the likelihood of the signal, for some $\beta \in [0, 1)$, the recall probability follows:

$$r(q_{\alpha_A}, l) = \begin{cases} r_H & \text{for } l > 1; \\ r_L & \text{for } l < 1. \end{cases}$$

$$r(q_{\alpha_I}, l) = \begin{cases} \beta r_H & \text{for } l > 1; \\ r_L & \text{for } l < 1. \end{cases}$$

Proposition 2. *The politician prefers the informative prompt α_I to the associative prompt α_A if and only if*

$$T < \frac{(1 - \beta)r_H \int_{l>1} f^\omega(l) \log l \, dl}{\log l_{\alpha_I} - \log l_{\alpha_A}}.$$

The Proposition shows that when the voter does not have much prior knowledge, i.e., T is small, the politician prefers to persuade by presenting an informative signal; as the voter accumulates enough knowledge, i.e., T is large, the politician instead prefers to persuade by sending prompt that “strikes a cord”, such that the voter is

more likely to retrieve memory that favors the politician. Note that this result holds regardless of whether the politician tries to convince the voter of a true or wrong state, and the result holds qualitatively when the politician does not know what the true state is.

The model illustrates a simple way to analyze the use of narrative in persuasion (Shiller (2017); Roos and Reccius (2021)), and the result provides a possible explanation for the increasingly polarized congressional speech in the US in recent years (Gentzkow et al. (2019)). More specifically, as information technology advances, more information is available, i.e., T increases. As a result, politicians prefer to use associative prompts, with more partisan languages in their speech. The result also implies a limit in social learning: as the voter collects more information, the politician turns to less informative but more associative prompts, thereby slowing down the learning of the voter.

5 Heterogeneous Reaction to New Information

There are numerous studies on how individuals update their beliefs with new information. The existing evidence suggests that reactions to new information are heterogeneous in different contexts, and across different individuals. For example, experimental studies generally find underreaction to new information (Benjamin (2019)), while empirical studies of the financial market typically document overreaction (e.g., Dessaint and Matray (2017); Huang et al. (2020); Bordalo et al. (2020, 2022b)). On the other hand, individuals with different prior beliefs could react differently to new information, depending on whether it supports or challenges their existing beliefs (Bail et al. (2018)). In particular, individuals might reinforce their prior beliefs after being exposed to belief-challenging information. Thus, public information can polarize the beliefs of two individuals with opposite priors. This section shows how the reaction to new information depends on the existing memory of individuals and sheds light on the aforementioned phenomena.

Similar to previous sections, consider a subject with a memory of T signals: $M = \{s_1, \dots, s_T\}$, where $q_t = l_t$. An experimenter tries to elicit the subject's "prior belief", denoted by p_T^1 , by presenting a prompt $\alpha_0 = (l_{\alpha_0}, q_{\alpha_0})$, e.g., "How likely do you think that your IQ test ranks among the top half of your session?". Naturally, the experimenter tries to avoid biasing the subject's belief towards one state or the other. Thus, I assume that the prompt α_0 is "generic", that $l_{\alpha_0} = 1$ and $r(q_{\alpha_0}, q_t, q_{-t}) = r_0$ for all signals s_t . The expected prior belief of the subject thus follows:

$$E \left[\log \frac{p_T^1}{1 - p_T^1} \right] = r_0 \sum_{t=1}^T \log l_t.$$

Next, the experimenter presents the subject with a new piece of information

$s_{T+1} = (l_{T+1}, q_{T+1})$, which serves as a new prompt, to elicit the subject's "posterior belief" denoted by p_{T+1}^1 . Without loss of generality, assume $l_{T+1} > 1$ such that the new information supports state 1. Here, I assume that information supporting the same state is "homogeneous", such that a new signal supporting state ω (weakly) likely invokes retrieval of signals supporting the same state. Moreover, the new signal is (weakly) more "thought-provoking" than the generic prompt α_0 , such that it invokes more retrieval of memory in general. Together, mathematically, I assume that:

$$r(q_{T+1}, q_t, q_{-t}) = \begin{cases} \beta r_H & \text{if } l_t > 1; \\ r_H & \text{otherwise,} \end{cases}$$

where $\beta \in [1, \frac{1}{r_H}]$ and $r_H \geq r_0$. β measures how homogeneous signals supporting the same state are, and $r_H - r_0$ measures how thought-provoking the new piece of information is. The expected posterior belief of the subject thus follows:

$$E \left[\log \frac{p_{T+1}^1}{1 - p_{T+1}^1} \right] = K \log l_{T+1} + \beta r_H \sum_{t: l_t > 1} \log l_t + r_H \sum_{t: l_t < 1} \log l_t$$

I first present a result on under- and overreaction. The subject on average overreacts to the new information if

$$E \left[\log \frac{p_{T+1}^1}{1 - p_{T+1}^1} \right] - E \left[\log \frac{p_T^1}{1 - p_T^1} \right] > \log l_{T+1},$$

and on average underreacts to the new information if the reverse inequality holds. Consider two subjects, A and B , with memory $M^A = (s_1^A \cdots, s_{T_A}^A)$ and $M^B = (s_1^B \cdots, s_{T_B}^B)$ and that on average they have the same prior, i.e.,

$$r_0 \sum_{t=1}^{T_A} \log l_t^A = r_0 \sum_{t=1}^{T_B} \log l_t^B.$$

I say that subject A has more information than subject B if

$$\sum_{t: l_t^A > 1} \log l_t^A > \sum_{t: l_t^B > 1} \log l_t^B,$$

or equivalently, $\sum_{t: l_t^A < 1} \log l_t^A > \sum_{t: l_t^B < 1} \log l_t^B$. In other words, subject A has more information supporting both states, which cancel each other such that she has the same expected prior as subject B .

Proposition 3. *Assume that $\beta > 1$ such that signals supporting the same state are strictly homogeneous. Suppose subject A and B have the same expected prior belief, but subject A has more information. If subject B on average overreacts to the new information, subject A also on average overreacts to the new information. The reverse is not necessarily true.*

Note that the Proposition does not require the new information to be strictly thought-provoking, i.e., r_H could be equal to r_0 . The Proposition provides a simple explanation for the results in the literature where underreaction is commonly observed in experimental data, but overreaction is more predominant in empirical data. As lab experiments tend to involve abstract states, e.g., guessing which urn a ball is drawn from, subjects do not have much related pre-existing memory. Less memory retrieval is invoked by the new information, and thus underreaction is more likely. Moreover, the homogeneous assumption is more likely to fail in lab experiments: the abstract information subjects receive in lab experiments, e.g., draws from an urn, is likely to be unfamiliar and dissimilar to information in the subjects' memories.

The following shows how two subjects with different prior beliefs react to the new information differently, and how a piece of public new information can polarize the beliefs of the two subjects.

Proposition 4. *Assume that $r_H > r_0$ such that the new information is strictly thought-provoking. Suppose subject A's expected prior $E(p_T^{1A}) > 0.5$, subject B's expected prior $E(p_T^{1B}) < 0.5$, and $l_{T+1} > 1$, i.e., the new information supports state 1. Subject A's expected posterior always favors state 1 more, i.e., $E(p_{T+1}^{1A}) > E(p_T^{1A})$, while subject B's expected posterior favors state 2 more if her prior is extreme enough, i.e., $E(p_{T+1}^{1B}) < E(p_T^{1B})$ if $E(p_T^{1B})$ is small enough.*

Note that the Proposition does not require information supporting the same state to be strictly homogeneous i.e., β could be equal to 1. The assumption that the new signal is strictly thought-provoking is more likely to hold when the issue of uncertainty is engaging, for example with political issues or when the issue is self-related. It is more likely to hold if the new signal is more detailed as it contains more cues that could invoke memory retrieval. The Proposition explains results in [Bail et al. \(2018\)](#) which shows that exposure to counter-attitudinal information backfires and increases polarization. The intuition is that the new information invokes memory retrieval of subjects, which likely contains information that supports their prior belief. The mechanism resonates with existing Cognitive and Neuroscience studies which show that a higher working memory ability is associated with stronger confirmation bias ([Doll et al. \(2011, 2014\)](#)).

6 Conclusion

This paper proposes a simple model of memory-based belief formation and three applications of the baseline model. Given that memory research has attracted and is still attracting massive attention across different disciplines, the model could leverage the many documented characteristics of memory retrieval to explain behavioral

phenomena and analyze applied problems like persuasion with competition, reducing the impact of fake news ([Kemp et al. \(2022\)](#)), and how false memory formation can exacerbate errors in beliefs ([Loftus and Pickrell \(1995\)](#); [Chew et al. \(2020\)](#)).

References

- Bail, Christopher A, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky (2018) “Exposure to opposing views on social media can increase political polarization,” *Proceedings of the National Academy of Sciences*, Vol. 115, No. 37, pp. 9216–9221.
- Benjamin, Daniel J (2019) “Errors in probabilistic reasoning and judgment biases,” *Handbook of Behavioral Economics: Applications and Foundations 1*, Vol. 2, pp. 69–186.
- Bordalo, Pedro, Giovanni Burro, Katherine B Coffman, Nicola Gennaioli, and Andrei Shleifer (2022a) “Imagining the future: memory, simulation and beliefs about COVID.”
- Bordalo, Pedro, Giovanni Burro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2023a) “Imagining the Future: Memory, Simulation and Beliefs.”
- Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer (2023b) “Memory and probability,” *The Quarterly Journal of Economics*, Vol. 138, No. 1, pp. 265–311.
- Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2020) “Belief overreaction and stock market puzzles.”
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2022b) “Overreaction and diagnostic expectations in macroeconomics,” *Journal of Economic Perspectives*, Vol. 36, No. 3, pp. 223–244.
- Chew, Soo Hong, Wei Huang, and Xiaojian Zhao (2020) “Motivated false memory,” *Journal of Political Economy*, Vol. 128, No. 10, pp. 3913–3939.
- Connors, Michael H and Peter W Halligan (2015) “A cognitive account of belief: a tentative road map,” *Frontiers in psychology*, Vol. 5, p. 1588.
- Dessaint, Olivier and Adrien Matray (2017) “Do managers overreact to salient risks? Evidence from hurricane strikes,” *Journal of Financial Economics*, Vol. 126, No. 1, pp. 97–121.

- Doll, Bradley B, Kent E Hutchison, and Michael J Frank (2011) “Dopaminergic genes predict individual differences in susceptibility to confirmation bias,” *Journal of Neuroscience*, Vol. 31, No. 16, pp. 6188–6198.
- Doll, Bradley B, James A Waltz, Jeffrey Cockburn, Jaime K Brown, Michael J Frank, and James M Gold (2014) “Reduced susceptibility to confirmation bias in schizophrenia,” *Cognitive, Affective, & Behavioral Neuroscience*, Vol. 14, pp. 715–728.
- D’Acunto, Francesco and Michael Weber (2021) “Memory and beliefs: Evidence from the field.”
- Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann (2020) “Associative memory and belief formation.”
- Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy (2019) “Measuring group differences in high-dimensional choices: method and application to congressional speech,” *Econometrica*, Vol. 87, No. 4, pp. 1307–1340.
- Graeber, Thomas, Florian Zimmermann, and Christopher Roth (2022) “Stories, statistics, and memory.”
- Huang, Qianqian, Feng Jiang, Yuhai Xuan, and Tao Yuan (2020) “Do Banks Over-react to Disaster Risk?” *Available at SSRN 3871505*.
- Kemp, Paige L, Vanessa M Loaiza, and Christopher N Wahlheim (2022) “Fake news reminders and veracity labels differentially benefit memory and belief accuracy for news headlines,” *Scientific Reports*, Vol. 12, No. 1, p. 21829.
- Leung, Benson Tsz Kin (2020) “Limited cognitive ability and selective information processing,” *Games and Economic Behavior*, Vol. 120, pp. 345–369.
- Li, King King (2013) “Asymmetric memory recall of positive and negative events in social interactions,” *Experimental Economics*, Vol. 16, pp. 248–262.
- Loftus, Elizabeth F and Jacqueline E Pickrell (1995) “The formation of false memories.”
- Roos, Michael and Matthias Reccius (2021) “Narratives in economics,” *Journal of Economic Surveys*.
- Schwartzstein, Joshua (2014) “Selective attention and learning,” *Journal of the European Economic Association*, Vol. 12, No. 6, pp. 1423–1452.
- Shiller, Robert J (2017) “Narrative economics,” *American economic review*, Vol. 107, No. 4, pp. 967–1004.

Wilson, Andrea (2014) “Bounded memory and biases in information processing,” *Econometrica*, Vol. 82, No. 6, pp. 2257–2294.

Wyer, Robert S and Dolores Albarracin (2005) “Belief formation, organization, and change: Cognitive and motivational influences,” *The handbook of attitudes*, Vol. 273, p. 322.

A Proofs

A.1 Proof of Proposition 1

Proof. I first prove the first part of the Proposition. Note that $\hat{p}_\infty^1 = 1$ if and only if $\int_l f^\omega(l)r(l) \log l \, dl > 0$. Also note that when $r(l) = 1$ for all l , the individual is standard Bayesian, $\hat{p}_\infty^1 = 1$ under state 1 and thus $\int_l f^1(l) \log l \, dl > 0$, while $\hat{p}_\infty^1 = 0$ under state 2 and thus $\int_l f^2(l) \log l \, dl < 0$. Now, when r is balanced and strictly positive, $r(l) = r$ for some constant $r > 0$, and $\int_l f^1(l)r \log l \, dl > 0 > \int_l f^2(l)r \log l \, dl$. The first part of the Proposition follows.

Now suppose there exists $\epsilon \in (0, 1]$ and $l \neq l'$ such that $r(l')/r(l) < \epsilon$, or equivalently $r(l') < \epsilon r(l)$. There must exist $l'' > 1 > l'''$ such that $r(l'') < \sqrt{\epsilon} r(l''')$ or $r(l''') < \sqrt{\epsilon} r(l'')$. Without loss of generality suppose $r(l''') < \sqrt{\epsilon} r(l'')$ or equivalently $r(l''')/r(l'') < \sqrt{\epsilon}$, consider the following distribution:

$$\begin{aligned} f^1(l'') &= \frac{l''(1-l''')}{l''-l'''} \text{ and } f^2(l'') = \frac{1-l'''}{l''-l'''}; \\ f^1(l''') &= 1 - f^1(l'') \text{ and } f^2(l''') = 1 - f^2(l''). \end{aligned}$$

Under state 2, the expected posterior belief when M contains T signals follows:

$$\log \frac{E(\hat{p}_T^1)}{1 - E(\hat{p}_T^1)} = \log l_\alpha + T \left[\frac{1-l'''}{l''-l'''} r(l'') \log l'' + \frac{l''-1}{l''-l'''} r(l''') \log l''' \right].$$

As $f^2(l'') \in (0, 1)$ and $\log l'' > 0$, $\log l''' < 0$ are bounded, when ϵ is close to 0, the term in the square bracket is negative. The result follows. $\square$

A.2 Proof of Proposition 2

Proof. Note that the politician prefers the informative prompt α_I to the associative prompt α_A if and only if

$$\log l_{\alpha_I} - \log l_{\alpha_A} > T \int_{l>1} (1-\beta) f^\omega(l) r_H \log l \, dl.$$

The result follows. $\square$

A.3 Proof of Proposition 3

Proof. If subject B on average overreacts to the new information, then

$$(\beta r_H - r_0) \sum_{t:l_t^B > 1} \log l_t^B + (r_H - r_0) \sum_{t:l_t^B < 1} \log l_t^B > (1 - K) \log l_{T+1}.$$

For individual A, we have

$$\begin{aligned} & (\beta r_H - r_0) \sum_{t:l_t^A > 1} \log l_t^A + (r_H - r_0) \sum_{t:l_t^A < 1} \log l_t^A \\ & > (\beta r_H - r_0) \sum_{t:l_t^B > 1} \log l_t^B + (r_H - r_0) \sum_{t:l_t^B < 1} \log l_t^B \\ & > (1 - K) \log l_{T+1}. \end{aligned}$$

The result follows. $\square$

A.4 Proof of Proposition 4

Proof. Subject i 's expected posterior favors state 1 more if and only if

$$K \log l_{T+1} + (\beta r_H - r_0) \sum_{t:l_t^i > 1} \log l_t^i + (r_H - r_0) \sum_{t:l_t^i < 1} \log l_t^i > 0.$$

For subject A, we have

$$\begin{aligned} & K \log l_{T+1} + (\beta r_H - r_0) \sum_{t:l_t^A > 1} \log l_t^A + (r_H - r_0) \sum_{t:l_t^A < 1} \log l_t^A \\ & \geq K \log l_{T+1} + (r_H - r_0) \sum_{t=1}^{T_A} \log l_t^A \\ & > K \log l_{T+1} > 0. \end{aligned}$$

On the other hand, for subject B,

$$E(p_{T+1}^{1B}) - E(p_T^{1B}) = K \log l_{T+1} + (\beta r_H - r_0) \sum_{t:l_t^B > 1} \log l_t^B + (r_H - r_0) \sum_{t:l_t^B < 1} \log l_t^B$$

which is negative when $\sum_{t:l_t^B < 1} \log l_t^B$ is sufficiently negative. The result follows. $\square$