Proceedings of Machine Learning Research 58 (2016) 111-127

Hindsight Bias Impedes Learning

Shaudi Mahdavi SHAUDI@WHARTON.UPENN.EDU
Wharton School, University of Pennsylvania, Philadelphia, PA 19148.

M. Amin Rahimian MOHAR@SEAS.UPENN.EDU
Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19148.

Editor: Tatiana V. Guy, Miroslav Kárný, David Rios-Insua, David H. Wolpert

Abstract

We propose a model that addresses an open question in the cognitive science literature: How can we rigorously model the cognitive bias known as hindsight bias such that we fully account for critical experimental results? Though hindsight bias has been studied extensively, prior work has failed to produce a consensus theoretical model sufficiently general to account for several key experimental results, or to fully demonstrate how hindsight impedes our ability to learn the truth in a repeated decision or social network setting. We present a model in which agents aim to learn the quality of their signals through repeated interactions with their environment. Our results indicate that agents who are subject to hindsight bias will always believe themselves to be high-type "experts" regardless of whether they are actually high- or low-type.

Keywords: Hindsight Bias, Bounded Rationality, Sequential Decision-making, Non-Bayesian Social Learning

1. Introduction

Hindsight bias, also known as the curse of knowledge, is a pervasive cognitive bias that captures the difficulty of faithfully reconstructing one's foresight knowledge in retrospect, after one has learned actual outcome or correct answer of a decision-making task (Fischhoff and Beyth, 1975; Camerer et al., 1989). Hindsight bias has been shown to contaminate judgment and decision-making on tasks ranging from predicting outcomes to arithmetic calculation to knowledge retrieval, and in people of all ages, skill and knowledge levels, intelligence, and cultures (Pohl et al., 2002; Bernstein et al., 2011; Pohl et al., 2010; Coolin et al., 2014).

Hindsight bias has been observed in many domains, in both the laboratory and the real world. In politics, pundits and laypeople alike commonly forget their pre-election predictions after a surprising election result, and report feeling that they had in fact anticipated the actual outcome all along. In healthcare settings, diagnosing physicians often experience exaggerated feelings of having been able to anticipate patient outcomes after the fact, or cannot objectively retroactively assess colleagues' diagnostic competence in cases where they begin knowing the patient's fate. Analogously, lawyers commonly claim to have known the outcome all along in the aftermath of a verdict, regardless of what they actually predicted before and during the trial. In gambling, sports bettors' mistaken recollection of having predicted shocking game results has been shown to give some gamblers unearned confidence to take increasingly risky future bets. Similar examples abound in educational, research, business, financial, and scientific decision making (Fischhoff and Beyth, 1975; Fischhoff, 2007; Fiedler et al., 2015; Arkes, 2013; Harley, 2007). As we discuss in this paper, the effects of hindsight bias, such as overconfidence, misattribution of blame to external factors, and normalized perception of unlikely tail events, can gravely distort heuristics and future decisions.

Ubiquity aside, hindsight bias is a compelling topic for study because of the striking difficulty of eliminating hindsight bias from people's decision-making. Indeed, Fischhoff (1975) and many others have deemed the so-called curse of knowledge an inevitability. This supposed inevitability bears out intuition: Once an outcome or answer is revealed, a person's outlook, and modes of interpretation, and judgment can be fundamentally changed in complex ways. Just as knowledge cannot be willed away, in the absence of costly, scrupulous bookmarking of knowledge and thoughts at various points in time, it can be extremely challenging to faithfully reconstruct one's prior, naive memory, judgments, and thought processes from a new position of information privilege.

We contribute to the substantial literature on hindsight bias by introducing a new, more complex, more general model of hindsight biased decision-making than currently exists. Using this model, we examine biased decision-making in two novel settings: when people can make infinitely many repeated judgments, and when they make decisions while communicating with a large social network of heterogeneous neighbors.

1.1 What is Hindsight Bias?

First documented by Fischhoff and Beyth (1975), hindsight bias formalizes the well-known "I knew it all along" distortion of cognition and information processing. This highly pervasive bias clouds a decision maker's ability to recall her prior expectations for an outcome after learning new information such as when the true state of the world, or to reliably anticipate the decisions of people she knows are not privy some knowledge she is. Outcomes and answers that seemed impossible to anticipate can suddenly seem obvious once the truth is revealed, because these truths either distort original memories, bias the reconstruction of foresight (pre-answer) knowledge, or both.

Interestingly, unless the revealed outcome or correct answer is subsequently debunked, most laboratory interventions attempting to mitigate hindsight bias have failed. In fact, several experiments have shown that even informing people they are prone to such bias, and instructing them to avoid allowing it to contaminate their answers during a range of decision-making tasks, fails to prevent hindsight from contaminating their answers (Fischhoff, 1976; Pohl and Hell, 1996; Bernstein et al., 2011).

Hindsight bias is generally identified and measured within two main experimental paradigms. The first, known as memory design, involves asking participants questions about a relatively obscure topic. These questions may involve judging the probability of a certain outcome, or answering a factual question (often with a numerical answer), such as the number of satellites of a planet. To assess hindsight bias, researchers first pose such a question to participants, reveal the answer or true outcome, and then again ask what the participants' original answer was. Hindsight bias is observed when participants' post-revelation answer moves away from the original response and towards the correct answer. The number of times and distance a subject's answer move towards the correct answer determine the magnitude of her bias. The second, hypothetical design, involves first informing experimental subjects of the answer to a relatively obscure question, then asking how they would have answered before learning the answer, or how a hypothetical naïve peer would have answered the same question. In this setup, hindsight bias would be expected to cause subjects to report hypothetical answers closer to the correct answer when they are informed of the answer than when they are not (Coolin et al., 2016; Leary, 1982)

Hindsight bias affects judgment in a broad range of settings. For example, Biais and Weber (2009) study hindsight bias exhibited by stock market traders. Traders need to update their risk assessments when volatility is stochastic, using realized returns. On observing unexpectedly positive or negative returns, rational agents should raise their volatility estimates. Hindsight biased traders will form inaccurate beliefs compared to fully rational agents, because in retrospect, they perceive their prior beliefs as having accounted for this tail event. In doing so, they fail to consider such returns were unexpected, and thus underestimate variances. This failure to correctly assimilate information naturally leads to inferior trades, and thus suboptimal financial performance, as they fail to hedge appropriately, fail to reject analyses that are challenged by the facts, and fail to differentiate between their information and that of their peers.

Next, consider the decisions in the context of medicine (Danz et al., 2015). First, we note that after a tumor has been diagnosed, in hindsight, 90 percent of lung cancers and 70 percent of breast cancers that had previously not been recognized are observed on radiographs (Berlin, 2003). Suppose that in the past, a trainee doctor had seen a radiograph of an patient but overlooked the tumor. Suppose also, as in the situation posed above, that the tumor was discovered later. A supervising doctor not involved with the diagnosis, who, armed with the knowledge of the existence of this tumor, is hindsight biased, unfairly evaluates the subordinate's ability to diagnose cancer. In other words, being informed about the location of the tumor, the supervisor overestimates the likelihood that she would herself have found the tumor initially. Consequently, she might underestimate the abilities of the young trainee doctor, against the backdrop of her own inflated sense of ability (Madarász, 2011).

1.2 Origins of Hindsight Bias

Despite a substantial amount of research devoted to the topic, theoretical modeling work is sparse, and no consensus model of hindsight bias has been proposed – likely because generalizing across the incredible heterogeneity of settings and tasks addressed by experimental work on the subject appears daunting. The prevailing theories of hindsight bias can be hierarchically grouped into three categories of theoretical constructs (Roese and Vohs, 2012; Blank et al., 2008; Nestler et al., 2008; Bernstein et al., 2015).

The first theoretical umbrella relates to distorted memories, which can cause people to incorrectly recall that they had correctly guessed the actual outcome or answer. Within this category, the first memory-based theory is *automatic assimilation*, or the idea that the brain encodes the correct answer automatically, and with negligible effort (Fischhoff, 1982; Bernstein et al., 2012). Automatic assimilation is consistent with the observation that informing participants of their hindsight bias does not reduce bias of their responses. The next theory is the *trace-strength hypothesis*, which asserts that memory traces, or records, for the original and correct answers possess different strengths, or degrees of accessibility, that determine the magnitude of the hindsight bias (Hell et al., 1988; Wood, 1978; Harley et al., 2004). The third memory construct is *recollection bias*, which describes the phenomenon in which learning the correct answer makes it more difficult to recall the original answer – either because the new memory overwrites or becomes more accessible than the previous memory (Erdfelder and Buchner, 1998; Erdfelder et al., 2007).

The second theoretical umbrella relates to the *reconstruction bias*, which arises when people attempt to reconstruct their naive foresight knowledge by some process that is, to some extent, biased or weighted toward the correct answer or revealed outcome (Schwarz and Stahlberg, 2003). The first theoretical construct in this category is *anchoring and adjustment*, in which participants anchor their response to correct answer, and adjust (perhaps incompletely) toward their original answer to a degree reflecting the magnitude of their bias (Epley and Gilovich, 2006). The second of these theories is *biased memory search*, according to which the correct answer corrupts memory recall by guiding the memory search toward content related to the correct answer (Hoffrage et al., 2000). The final construct, *metacognitive regulation theory*, asserts that adjustment toward the correct answer depends on one's confidence in the original and recalled answers, and on the extent of surprise caused by the newly-learned information (Müller and Stahlberg, 2007).

The final theoretical umbrella is *motivational theory.* In essence, this theory is rooted in the idea that people tend to misremember their original judgments or answers such that they are perceived as wiser or smarter than they actually believe themselves to be. Motivational theory presupposes that humans are motivated to maintain a sense of self-worth, which they derive from markers of intelligence such as decision-making aptitude, level of knowledge, and fidelity of memory. Three theories comprise this category. The first is *retroactive pessimism*, which asserts that people protect their sense of self-worth in the face of negative outcomes by convincing themselves a negative outcome was more likely than they believed it to be in foresight, and perhaps inevitable (Tykocinski and Steinberg, 2005). The second construct is *defensive processing.* According to this theory, people self-protectively perceive negative outcomes as less-predictable in hindsight to absolve themselves of responsibility and guilt for having made bad decisions that led to the outcome in question (Louie, 1999; Mark and Mellor, 1991). The last, *motivated sense-making* combines the two preceding theories, such that people use external and internal factors, depending on the nature and degree of surprise the experience, to explain discrepancies in expectation versus outcome (Pezzo, 2011).

1.3 Motivation and Contributions to the Literature

In this paper, we present a novel model of hindsight-biased learning that is both more generalizable and more complex than past models. First, thanks to its relative complexity, unlike existing models, our model clearly captures elements of all three aforementioned theoretical constructs. Our model accounts for faulty memories as well as faulty memory reconstruction (through anchoring and adjustment) as previous models have done, but with an additional element of randomness that accounts for motivational theory. By introducing this randomness, we account for the tension between a person's drive to produce correct answers to feel competent (motivational theory), and the orthogonal need to stress-test hypotheses and admit heuristic weakness, in order to be able to improve future decision-making.

Second, we incorporate elements that allows us to distinguish among agents heterogeneous in multiple dimensions. We define classes of good decision-making and bad decision-making agents as a broad proxy for inherent differences in skill, domain knowledge, age, intelligence, memory capacity, that influence decision-makers' susceptibility to hindsight bias in the real world, and can account for result discrepancies in experiments administered to different populations of subjects (Roese and Vohs, 2012). We further accommodate subject heterogeneity by distinguishing between signal reception and signal perception. Unlike previous models, we distinguish between decision-makers not only in terms of what signals they receive, but also in how they perceive those signals – an additional layer of complexity that lends itself to increased generalizability.

Thanks to our modeling choice, we can contribute the first study of hindsight biased decision outcomes under infinitely repeated decisions, and in an arbitrarily large networked setting with social learning. More specifically, unlike previous models, which typically consider decision-making as a two-step process, we study the cumulative effects of hindsight bias over repeated decisions, such as those of a stock trader who makes investment decisions every day. We also study the implications of communications with a network of heterogeneous agents for hindsight-biased learning.

2. Modeling Hindsight Bias

There exist several models of hindsight bias in the literature. One of the most prominent is that of Madarász (2011). Here, an agent has a signal realization s, but does not correctly remember it. Instead, her memory is biased towards the state of the world revealed to be true in the meantime. Unlike the approaches used in the earlier models of Camerer et al. (1989) and Biais and Weber (2009) where the parameter $\lambda \in (0, 1)$ is used to form the linear combination of the true signal and the true state of the world, Madarász (2011) regards the parameter λ as the weight of the two beliefs. This is a subtle but highly critical distinction. A λ-hindsight biased agent, in forming her posterior belief, weights the signal realization from the first period with $(1 - \lambda)$ and the true state of the world with λ. Here, parameter λ denotes the strength of the hindsight bias. If we denote the posterior beliefs of the agent by $\pi_+(\cdot)$ and the truth state by ω, the biased agent's posterior belief becomes $\pi_+(\cdot) = \lambda\pi_+(\cdot|\omega) + (1 - \lambda)\pi_+(\cdot|s)$. This is a deviation from the Bayesian posterior $\pi_+(\cdot|s)$, which updates the prior belief, denoted by $\pi(\cdot)$, given the observation s according to the Bayes rule. Another prominent model introduced by Biais and Weber (2009) defines the hindsight bias by forming the reconstructed prior mean of the biased agent as a weighted average of the true prior mean and of the realization of the random variable.

In contrast to the above authors' approaches, we model the cognition of the hindsight-biased agent as a random process, such that: with probability $\lambda \in (0, 1]$ the biased agent believes she has observed the correct signal (coinciding with the truth), irrespective of of her actual observation. We thus distinguish between the agents' perceptions r_t and observation s_t by modeling both as random processes, which may or may not agree with the true state ω_t.

The observations s_t are a sequence of independent random variables[1], and the perceptions r_t are such that $r_t = s_t$ with probability $1 - \lambda$ and $r_t = \omega_t$ (the truth state) with probability λ; we thus investigate the dynamic effects as the agents receives a sequence of observations, while updating their beliefs based on their perceptions:

$$r_t = \begin{cases} \omega_t & \text{with probability} \qquad \lambda, \\ \\ s_t & \text{with probability} \qquad (1-\lambda). \end{cases} \tag{1}$$

The above equation captures the fact that the agent has a limited memory of her past signal. The first term represents the potentially false memory from projecting her knowledge of the state of the world onto her past signal, while the second represents a factual memory of the signal observed in the past. Stated differently, in this formulation, the agent's recollection of the signal realization, given in the second term, is corrupted by the agent's memory limitation and cognitive bias, which leads her to project her current information ω_t onto her past signal s_t.

We can also justify the model and the bias parameter λ from a motivational theory perspective: Every time that our decision-maker makes a mistake, she pays an emotional (fixed) cost of believing herself mistaken. On the other hand, every time she remembers her choice correctly, she benefits by learning about the quality of her choice (informational benefit), so λ could balance the expected (emotional) cost of making mistakes against the informational advantage of new correct samples. For if λ were such that the expected cost of mistake – in other words, the expected negative emotions caused by perception of a mistake – exceeded the informational benefits of a correct sample, the agent would be always flipping her choice to avoid the negative emotions; while if the expected informational benefits of a new sample were more valuable to her, then she would always be forcing herself to remember her samples correctly so that she could learn from them. The fact that the agent randomizes between the two strategies implies that the bias parameter λ should equalize the expected costs and benefits of the incoming samples. It is worthwhile to highlight two main assumptions that underlie equation (1): first the bias parameter λ is fixed over time (time-invariance); secondly, the agent is unaware of her hindsight bias and the underlying parameter λ as well as the stimuli random process s_t are not observable to her.

3. How does Hindsight bias limit rational learning?

Inspired by Danz et al. (2015), we assume that agents are heterogeneous in some characteristics that govern their decision-making abilities. We model this as follows: we assume nature draws the type of the agent (at random), and the agent's type can be either high (i.e., good) denoted by $\bar{\theta}$ or low (i.e., bad), denoted by $\underline{\theta}$. We further assume the prior probability of being a high-type agent is $\pi_0(\bar{\theta})$. The agent does not know her own type, but does have a prior $\pi_0(\cdot)$ about the type distributions. The latter represents the agent's knowledge of the distribution of the types before she makes any personal observations about her own type;

1. The independence assumption models repeated interactions of the agent with her environment, for example examinations of radiographs of different patients by a doctor. This assumption may be relaxed to some extent as long as the arguments pertaining to the application of the law of large numbers to the sum of random variables in the appendix remain valid.

for example, $\pi_0(\bar{\theta})$ can be induced by the prior knowledge about the fraction of radiologists who have superior skills in diagnostic radiography. At the beginning of each period t, nature draws the state of the world ω_t, which can be thought to represent a property of widespread interest such as the health status of a patient. States of the world are drawn independently across periods, and the two possible realizations of the state of the world, 0 and 1, are equally likely. The latter assumption is only for the sake of clarity and the ensuing analysis does not rely on this assumption because the inferences of the agent are based only on the agreements or disagreements between the perceived random process r_t and the truth states ω_t; such agreements and disagreements are in turn dependent on the received signals s_t through (1). The quality of the signal, i.e., the probability ρ that the signal corresponds to the true state of the world, depends on the type θ of the agent: A high type ($\bar{\theta}$) agent receives signals that are correct with probability $\bar{\rho}$ whereas a low type ($\underline{\theta}$)'s signals are only correct with probability $\underline{\rho}$, where $0.5 \leq \underline{\rho} < \bar{\rho} \leq 1$.

We begin with the assumption of independence across time. At every t, we have ω_t which takes either of the two values $\{0, 1\}$ with equal probability. Thus in lack of any side information, s_t and r_t also takes values $\{0, 1\}$ with equal probabilities. Here, the learning goal is to determine whether the agent's type, denoted as θ, is low ($\underline{\theta}$) or high ($\bar{\theta}$); symbolically, whether $\theta = \underline{\theta}$ or $\theta = \bar{\theta}$.

Concretely, in a real-world setting, a high-type agent is a person with the capacity to make better inferences about the state of the world than her low-type counterparts. For modeling purposes, we could use the type parameter θ to more generally capture heterogeneity of any of a vast number of qualities that influence judgment and decision-making – for example, domain knowledge, skill, or expertise; memory size or reliability; memory retrieval; intelligence, attention span; arithmetic aptitude; age to the extent that it influences cognitive processes; confidence; and many others.

3.1 Description of the Dynamic Model

At every round state ω_t is drawn and a signal s_t is observed. The likelihood of the signal given the state is determined by the agent's type, which could be high or low. We denote the observed signals of low and high type agents by $\underline{s}_t$ and $\bar{s}_t$, respectively. The signal of a high type agent is distributed according to

$$\bar{s}_t = \begin{cases} \omega_t & \text{with probability} \quad \bar{\rho} \\ 1 - \omega_t & \text{with probability} \quad 1 - \bar{\rho}. \end{cases}$$

while the likelihood of the signals for low type agent is given by

$$\underline{s}_t = \begin{cases} \omega_t & \text{with probability} \quad \underline{\rho}, \\ 1 - \omega_t & \text{with probability} \quad 1 - \underline{\rho}. \end{cases}$$

The signal perceived by the agent differs from her actual observations due to her bias. We denote the perceived signal of a high type agent by $\bar{r}_t$; similarly for a low type agent, we use $\underline{r}_t$.

The high and low types will subsequently update their beliefs based on their (biased) perceptions rather than their actual observations. Hence for a high type we have $\pi_t(\theta) = P\{\theta|\bar{r}_1, \bar{r}_2, ..., \bar{r}_t\}, \theta \in \{\bar{\theta}, \underline{\theta}\}$, and for a low type we obtain $\pi_t(\theta) = P\{\theta|\underline{r}_1, \underline{r}_2, ..., \underline{r}_t\}, \theta \in \{\bar{\theta}, \underline{\theta}\}$; i.e. the belief at time t is the conditional probability of each state given the perceptions made up until time t. In the case that there is no hindsight bias and the agent is fully rational, the perceptions correspond to the actual observations, and we have:

$$\bar{r}_t = \bar{s}_t \;\; \forall\, t \;\rightarrow \bar{r}_t = \bar{s}_t = \begin{cases} \omega_t & \text{with probability} \quad \bar{\rho} \\ \\ 1 - \omega_t & \text{with probability} \quad 1 - \bar{\rho}. \end{cases}$$

for a rational high type agent and

$$\underline{r}_t = \underline{s}_t \;\; \forall\, t \;\rightarrow \underline{r}_t = \underline{s}_t = \begin{cases} \omega_t & \text{with probability} \quad \underline{\rho} \\ \\ 1 - \omega_t & \text{with probability} \quad 1 - \underline{\rho}. \end{cases}$$

for a rational low type agent. For such a fully Bayesian agent, we have that

$$\pi_t(\theta = \bar{\theta}) \underset{a.s.}{\longrightarrow} \mathbb{1}(\theta = \bar{\theta}) \quad \text{as} \quad t \rightarrow \infty, \tag{2}$$

similarly for $\theta = \underline{\theta}$. Thus rational agents always learn their true types (with probability one).

On the other hand, a biased agent would, with probability $0 < \lambda < 1$, perceive her observed signal as the truth ω_t that is revealed to her, after her experiment with her environment (leading to her observation). Similarly, with probability $1 - \lambda$ she will perceive her signal as it was, i.e. $\bar{s}_t$ for a high type and $\underline{s}_t$ for a low type. Consequently, we can derive the likelihoods of the perceived signals conditional on the state ω_t for low and high types as follows:

$$\bar{r}_t = \begin{cases} \omega_t & \text{with probability} \qquad \lambda + (1-\lambda)\bar{\rho}, \\ \\ 1 - \omega_t & \text{with probability} \qquad (1-\lambda)(1-\bar{\rho}), \end{cases} \tag{3}$$

for a high type, and

$$\underline{r}_t = \begin{cases} \omega_t & \text{with probability} \qquad \lambda + (1-\lambda)\underline{\rho}, \\ \\ 1 - \omega_t & \text{with probability} \qquad (1-\lambda)(1-\underline{\rho}), \end{cases}$$

for a low type. Now, we ask whether, following experimentation with their environment and after updating their beliefs based on their perceptions according to $\pi_t(\theta) = P\{\theta|\bar{r}_1, \bar{r}_2, ..., \bar{r}_t\}, \theta \in \{\bar{\theta}, \underline{\theta}\}$, and $\pi_t(\theta) = P\{\theta|\underline{r}_1, \underline{r}_2, ..., \underline{r}_t\}, \theta \in \{\bar{\theta}, \underline{\theta}\}$: can agents correctly learn their types, despite their hindsight bias?

Recall from (2) that rational agents learn their types, almost surely; in fact, one can show that this learning occurs at a rate that is asymptotically exponentially fast, cf. Section III.A of (Rahimian and Jadbabaie, 2015). The situation for hindsight biased agents is different:

Theorem 1 (Bayesian Learning with Hindsight) *Regardless of type, an agent with hindsight believes herself to be of a high type. In particular, low-type hindsight biased agents always mislearn their true states.*

Theorem 1 establishes that low type agents may come to incorrectly believe themselves to be of a higher type due to their hindsight following their experimentations with their environment. As illustrated by this case, hindsight bias prevents rational processing, and therefore, learning from past data. As Fischhoff (1982) wrote, hindsight bias prevents agents from rejecting their hypotheses about the world. Biased agents' memories and reconstruction processes are contaminated by the actual outcome information. Therefore, they underestimate past surprises by mistakenly factoring current information into the original expectation about the state of the world. As a result, agents use weak tests to evaluate hypotheses, and fail to properly update original hypotheses that they might have about the state of the world during experimentation.

In practical terms, the inability to be surprised, learn from the past and and reject hypotheses can be very destructive. As an illustration, stock traders burdened by such limitations will fail to recognize and revise incorrect hypotheses about the market, and thus fail to cut losses, sell investments, or avoid risky trades when a fully rational agent would. Also, hindsight-biased investors might incorrectly (irrationally) incorporate new informational content like signals from sources like earnings announcement or macroeconomic news, contaminating their memory reconstruction, giving them a falsely optimistic view of the optimality and value of their trading heuristics.

3.2 How is social learning affected by the hindsight bias?

Next, we consider the dynamic model of Section 3 in a network setting. Suppose that there are n agents. There is a type variable that is common to all agents, and determining this is the objective of the learning process. This type could encapsulate any number of characteristics of this society, such as the performance of the economy: if the agents are high type, then they live in a good economy; whereas if the agents are low-type, then they live in a bad economy. In this case, based on their experiences in their environment and on their interactions with each other, agents form an opinion about the state of their economy. Starting from a prior $\pi_0(\bar{\theta})$ at beginning of each round, neighboring agents communicate with each other and update their beliefs following a log-linear update with positive weights that sum to one, cf. (Rahimian and Jababaie, 2015) and Section 4 of (Rahimian and Jadbabaie, 2016) for behavioral and cognitive foundations of log-linear belief updates. After this initial communication they interact with their environment in the following manner: at each time t, nature draws a binary state $\omega_{i,t}$ for each of the n agents. Each agent observes a signal ($\bar{s}_{i,t}$ or $\underline{s}_{i,t}$ depending on society's type) conditional on her state and after the state is revealed to her, she changes her perception of her observation into a biased version ($\bar{r}_{i,t}$ or $\underline{r}_{i,t}$ depending on society's type) and she then updates her belief about her type based on her perception and using the Bayes rule. At the initiation of each future epoch, the agents communicate their beliefs following the same log-linear rule and interact with their environment in the manner described above. The states $\omega_{i,t}$, signals $s_{i,t}$, and perceptions $r_{i,t}$ encode the essential aspects of the agents' interactions with their environment: for example in a good economy, each agent $i \in [n]$ can make better judgments about her state $\omega_{i,t}$ and thus makes decisions that lead to increased prosperity.

Our next theorem addresses the asymptotic outcome of the social learning with hindsight; in particular, whether the agents reach consensus and if they do, is it a consensus on truth or not. What if different agents have different λ_i? How does the heterogeneity in individual biases affect the learning outcome? All these questions are addressed by the following theorem.

Theorem 2 (Social Learning with Hindsight) *Consider a society of agents with a common type θ, which could be high or low, and at every epoch they communicate their beliefs about this common type. In addition, the agents also have a sequence of individual interactions with their environment and make private observations accordingly. If agents are subject to hindsight bias with parameter λ_i for each agent i, then regardless of type, a society of agents with hindsight believe themselves to be of a high type. In particular, a society of low-type hindsight biased agents always mis-learn their true types. This is true even if all but one agents are rational, as long as $\sum_{i=1}^{n} \lambda_i > 0$ and the agents communicate in a connected social network some agents' hindsight prevents all of them from correctly learning their true types.*

Theorem 2 reinforces the conclusions of Theorem 1 in that the low types agents' inference is compromised by their hindsight, causing them to incorrectly believe themselves to be of a higher type. Furthermore, the proof of Theorem 2 demonstrates that the hindsight bias for the network as a whole is simply the average hindsight over all agents in the network. Hence, interacting with biased agents contaminates the opinion of unbiased ones and subjects them to the same kinds of bias. Putting highly biased agents into such a setting with a large number of less-biased or unbiased agents would be an effective strategy for mitigating hindsight bias within a networked setting.

4. Conclusions

We have modeled the effect of hindsight bias on the inferences of a rational observer and also on the social learning outcomes in a network of arbitrary size. We introduced a more generalizable, more nuanced model that can capture multiple dimensions of agent heterogeneity, and has a demonstrable connection to each of three major hindsight bias theoretical constructs – both of which previous models lack. Our results also help verify the intuition that people become overconfident because they do not or cannot faithfully retain their past track record of wrong predictions, or correctly reconstruct earlier thought processes.

Acknowledgments

The authors gratefully acknowledge the editors of the NIPS 2016 Workshop Imperfect Decision Makers: Admitting Real-World Rationality for their comments on an earlier version of this paper.

Appendix A. Proofs of the Main Results

A.1 Proof of Theorem 1

For a high-type hindsight biased agent, we have, for the probability it puts on being high-type:

$$\pi_t(\bar{\theta}) = \mathbb{1}(\bar{r}_t = \omega_t)\frac{\pi_{t-1}(\bar{\theta})\bar{\rho}}{\pi_{t-1}(\bar{\theta})\bar{\rho} + \pi_{t-1}(\underline{\theta})\underline{\rho}} + \mathbb{1}(\bar{r}_t \neq \omega_t)\frac{\pi_{t-1}(\bar{\theta})(1-\bar{\rho})}{\pi_{t-1}(\bar{\theta})(1-\bar{\rho}) + \pi_{t-1}(\underline{\theta})(1-\underline{\rho})}.$$

and analogously, for the probability the same agent puts on being low-type,

$$\pi_t(\underline{\theta}) = \mathbb{1}(\bar{r}_t = \omega_t)\frac{\pi_{t-1}(\underline{\theta})\underline{\rho}}{\pi_{t-1}(\bar{\theta})\bar{\rho} + \pi_{t-1}(\underline{\theta})\underline{\rho}} + \mathbb{1}(\bar{r}_t \neq \omega_t)\frac{\pi_{t-1}(\underline{\theta})(1-\underline{\rho})}{\pi_{t-1}(\bar{\theta})(1-\bar{\rho}) + \pi_{t-1}(\underline{\theta})(1-\underline{\rho})}$$

whence we calculate the following ratio:

$$\frac{\pi_t(\bar{\theta})}{\pi_t(\underline{\theta})} = \begin{cases} \frac{\pi_{t-1}(\bar{\theta})}{\pi_{t-1}(\underline{\theta})}\frac{\bar{\rho}}{\underline{\rho}}, \text{ if } r_t = \omega_t \\ \\ \frac{\pi_{t-1}(\bar{\theta})}{\pi_{t-1}(\underline{\theta})}\frac{(1-\bar{\rho})}{(1-\underline{\rho})}, \text{ if } r_t \neq \omega_t \end{cases} \tag{4}$$

It follows that we have the recursive relation

$$\frac{\pi_t(\bar{\theta})}{\pi_t(\underline{\theta})} = \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_t=\omega_t)}\left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_t\neq\omega_t)}\frac{\pi_{t-1}(\bar{\theta})}{\pi_{t-1}(\underline{\theta})}$$

and hence, iterating on t, we easily see that:

$$\begin{aligned} \frac{\pi_t(\bar{\theta})}{\pi_t(\underline{\theta})} &= \prod_{\tau=1}^{t}\left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_t=\omega_t)}\left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_t\neq\omega_t)}\frac{\pi_0(\bar{\theta})}{\pi_0(\underline{\theta})} \\ &= \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\sum_{\tau=1}^{t}\mathbb{1}(\bar{r}_t=\omega_t)}\left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{\sum_{\tau=1}^{t}\mathbb{1}(\bar{r}_t\neq\omega_t)}\frac{\pi_0(\bar{\theta})}{\pi_0(\underline{\theta})}. \end{aligned} \tag{5}$$

Now, it follows easily from the law of large numbers that: $\sum_{\tau=1}^{t}\mathbb{1}(\bar{r}_t = \omega_t) \underset{a.s.}{\longrightarrow} t\big(\lambda+(1-\lambda)\bar{\rho}\big)$ and $\sum_{\tau=1}^{t}\mathbb{1}(\bar{r}_t = \omega_t) \underset{a.s.}{\longrightarrow} t\big((1-\lambda)(1-\bar{\rho})\big)$. Finally, using the latter two limits in the above equation, we obtain, for a high-type agent:

$$\frac{\pi_t(\bar{\theta})}{\pi_t(\underline{\theta})} \underset{a.s.}{\longrightarrow} \left(\left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{(\lambda+(1-\lambda)\bar{\rho})}\left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{(1-\lambda)(1-\bar{\rho})}\right)^t\frac{\pi_0(\bar{\theta})}{\pi_0(\underline{\theta})}.$$

The base of the exponential term for all range of parameters $0.5 < \underline{\rho} < \bar{\rho} < 1$ and $0 < \lambda < 1$ is always strictly greater than one[2], indicating that the belief ratio $\pi_t(\bar{\theta})/\pi_t(\underline{\theta})$ for high type

2. To see why, consider the first-order partial derivative of the function

$$f(\lambda, \bar{\rho}, \underline{\rho}) = \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{(1-\lambda)\bar{\rho}+\lambda}\left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{(1-\lambda)(1-\bar{\rho})}, \text{ with respect to } \lambda:$$

agent always converges almost surely to $+\infty$ leading to her learning her true type which is high. However, the case for a low type agent is different, following the same steps as above and using $\underline{r}_t$ instead of $\bar{r}_t$ for the perceived signals of a low type agent we obtain the following almost sure limit for the belief ratio of a low type agent:

$$\frac{\pi_t(\bar{\theta})}{\pi_t(\underline{\theta})} \xrightarrow[a.s.]{} \left(\left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{(\lambda+(1-\lambda)\underline{\rho})} \left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{(1-\lambda)(1-\underline{\rho})} \right)^t \frac{\pi_0(\bar{\theta})}{\pi_0(\underline{\theta})}$$

If a low type agent is to learn her type correctly then we need the above belief ratio to converge to 0 almost surely or equivalently we need the parameters $\underline{\rho}$, $\bar{\rho}$ and λ be such that

$$\left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{(\lambda+(1-\lambda)\underline{\rho})} \left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{(1-\lambda)(1-\underline{\rho})} < 1, \tag{6}$$

There are no parameters $0.5 < \underline{\rho} < \bar{\rho} < 1$ and $0 < \lambda < 1$ for which the above inequality is not violated[3], causing a low type agent to always incorrectly believe she is of a higher type; thence, the ability of a low type agent to learn her type is compromised by her hindsight.

$$\frac{\partial f(\lambda,\bar{\rho},\underline{\rho})}{\partial \lambda} = (\bar{\rho}-1)\left(\frac{\bar{\rho}-1}{\underline{\rho}-1}\right)^{(\lambda-1)(\bar{\rho}-1)} \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\lambda(-\bar{\rho})+\lambda+\bar{\rho}} \left(\log\left(\frac{\bar{\rho}-1}{\underline{\rho}-1}\right) - \log\left(\frac{\bar{\rho}}{\underline{\rho}}\right)\right).$$

The latter is positive everywhere and therefore increasing monotonically over the domain, which implies that for any fixed choice of $\bar{\rho},\underline{\rho}$, $f(\lambda,\bar{\rho},\underline{\rho})$ is minimized at $\lambda^* = 0$. Now, we seek values of $\bar{\rho},\underline{\rho}$ that minimize

$$f(\lambda^*,\bar{\rho},\underline{\rho}) = \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\bar{\rho}} \left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{(1-\bar{\rho})}, \text{ to satisfy the condition } f(\lambda^*,\bar{\rho},\underline{\rho}) < 1.$$

Given our assumption of positive $\bar{\rho}$ and $\underline{\rho}$, we have

$$\frac{\partial f(\lambda^*,\bar{\rho},\underline{\rho})}{\partial \underline{\rho}} = \frac{(\bar{\rho}-1)\bar{\rho}(\bar{\rho}-\underline{\rho})\left(\frac{\bar{\rho}-1}{\underline{\rho}-1}\right)^{-\bar{\rho}}\left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\bar{\rho}-1}}{(\underline{\rho}-1)^2\underline{\rho}^2} < 0,$$

which is monotonic decreasing in $\underline{\rho}$, which implies that for any fixed $\bar{\rho}$, $f(\lambda^*,\bar{\rho},\underline{\rho})$ is minimized by maximal $\underline{\rho}$. Hence, over the domain of f,

$$f(\lambda,\bar{\rho},\underline{\rho}) \geq f(\lambda^*,\bar{\rho},\underline{\rho}) \geq \bar{\rho}\left(\frac{\bar{\rho}}{\bar{\rho}}\right)^{\bar{\rho}}\left(\frac{1-\bar{\rho}}{1-\bar{\rho}}\right)^{(1-\bar{\rho})} = 1 \implies \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{(1-\lambda)\bar{\rho}+\lambda}\left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{(1-\lambda)(1-\bar{\rho})} \geq 1,$$

and hence a high type agent always learns her type correctly.

3. In more details (6) is satisfied, only if

$$\lambda < \left(\frac{\log(\frac{\underline{\rho}-1}{\bar{\rho}-1})}{\log(\frac{\underline{\rho}-1}{\bar{\rho}-1}) + \log(\frac{\bar{\rho}}{\underline{\rho}})} - 2\underline{\rho}\right)(1-\underline{\rho})^{-1}, \text{ which can never be since } \lambda > 0,\ 2\underline{\rho} > 1 \text{ and}$$

$$\frac{\log(\frac{\underline{\rho}-1}{\bar{\rho}-1})}{\log(\frac{\underline{\rho}-1}{\bar{\rho}-1}) + \log(\frac{\bar{\rho}}{\underline{\rho}})} < 1, \text{ for all } 0.5 < \underline{\rho} < \bar{\rho} < 1.$$

A.2 Proof of Theorem 2

Let $\mathcal{G} = ([n], \mathcal{E}, T)$, $[n] = \{1, \ldots, n\}$ be the weighted graph of social network influences, with $T = [T_{ij}]_{i,j\in[n]}$ being the matrix of influence weight and $T_{i,j} \geq 0$ being the influence of agent j on i, and $\mathcal{E} \subset [n] \times [n]$ is the set of edges corresponding to the non-zero entries of matrix T. We assume T to be a doubly stochastic matrix. At every round t, let $\pi_{i,t}(\bar{\theta})$ and $\pi_{i,t}(\underline{\theta})$ be the beliefs of agent i on the false and true states after communicating with her neighbors and experimenting with her environment at the particular round. The neighborhood of agent i is the set of all agents who have a strictly positive influence on her: $\mathcal{N}(i) = \{j : T_{i,j} > 0\}$. At all epochs t, define $\hat{\pi}_{i,t}(\bar{\theta})/\hat{\pi}_{i,t}(\underline{\theta})$ to be the updated belief of agent i after communicating with her neighbors but before experimenting with her environment, i.e. before observing $\bar{s}_{i,t}$ or perceiving $\bar{r}_{i,t}$ for a high type society (we consider the case of high-type agents and the low-type case is similar). The log-linear belief updates prescribe the updated belief of the agent before experimenting with her environment $\hat{\pi}_i(\bar{\theta})/\hat{\pi}(\underline{\theta})$ in terms of the reported beliefs of her neighbors:

$$\frac{\hat{\pi}_{i,t}(\bar{\theta})}{\hat{\pi}_{i,t}(\underline{\theta})} = \prod_{j\in\mathcal{N}(i)} \left(\frac{\pi_{j,t-1}(\bar{\theta})}{\pi_{j,t-1}(\underline{\theta})}\right)^{T_{ij}}. \tag{7}$$

Consider the agent in a society of high type agents, after communicating with her neighbors the agent engages with her environment and (4) still holds true in her updating of her beliefs following her observation of the private signal $\bar{s}_{i,t}$ and after her true state $\omega_{i,t}$ us revealed to her, subjecting her to a hind sight bias. Following (5) and (7) we can write:

$$\begin{aligned}\frac{\pi_{i,t}(\bar{\theta})}{\pi_{i,t}(\underline{\theta})} &= \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_{i,t}=\omega_{i,t})} \left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_{i,t}\neq\omega_{i,t})} \frac{\hat{\pi}_{i,t}(\bar{\theta})}{\hat{\pi}_{i,t}(\underline{\theta})} \\ &= \left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_{i,t}=\omega_{i,t})} \left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{\mathbb{1}(\bar{r}_{i,t}\neq\omega_{i,t})} \prod_{j\in\mathcal{N}(i)} \left(\frac{\pi_{j,t-1}(\bar{\theta})}{\pi_{j,t-1}(\underline{\theta})}\right)^{T_{ij}}\end{aligned} \tag{8}$$

Define the log of the belief ratios for every agent as $\phi_{i,t} = \log(\pi_{i,t}(\bar{\theta})/\pi_{i,t}(\underline{\theta}))$, and concatenate the log of the belief ratios for all agents into a single vector ϕ_t defined as $\phi_t = (\phi_{1,t}, \ldots, \phi_{n,t})$. Similarly, define the log likelihood ratio of the perceived signals as

$$\bar{\ell}_{i,t} = \mathbb{1}(\bar{r}_{i,t} = \omega_{i,t}) \log(\bar{\rho}/\underline{\rho}) + \mathbb{1}(\bar{r}_{i,t} \neq \omega_{i,t}) \log((1-\bar{\rho})/(1-\underline{\rho})) \tag{9}$$

and let $\bar{\ell}_t = (\bar{\ell}_{1,t}, \ldots, \bar{\ell}_{n,t})$ be their concatenation. Taking logarithms of both sides in (8) and using the concatenations ϕ_t and $\bar{\ell}_t$, we obtain the following linearized and vectorized update:

$$\phi_t = T\phi_{t-1} + \bar{\ell}_t = \sum_{\tau=1}^{t} T^{t-\tau}\bar{\ell}_\tau + \phi_0.$$

Next note that for a doubly stochastic matrix $\lim_{\tau\to\infty} T^\tau = (1/n)\mathbb{1}\mathbb{1}^T$. Hence, we can invoke the Cesàro mean together with the strong law applied to the i.i.d. sequence $\{\bar{\ell}_t, t \in \mathbb{N}_0\}$ to conclude that (Rahimian and Jababaie, 2015),

$$\frac{1}{t}\phi_t \to \frac{1}{n}\mathbb{1}\mathbb{1}^T\mathbb{E}\{\bar{\ell}_t\}$$

or equivalently that $\phi_{i,t} \to (t/n)\sum_{i=1}^{n} \mathbb{E}\{\bar{\ell}_{i,t}\}$ with probability one, as $t \to \infty$. Next using (3) and (9) we can write

$$\mathbb{E}\{\bar{\ell}_{i,t}\} = (\lambda_i + (1-\lambda_i)\bar{\rho}) \log\left(\bar{\rho}/\underline{\rho}\right) + (1-\lambda_i)(1-\bar{\rho}) \log\left((1-\bar{\rho})/(1-\underline{\rho})\right),$$

and the conclusion for high type agents follows because $\mathbb{E}\{\bar{\ell}_{i,t}\} > 0$ for all agents $i \in [n]$ with $0.5 < \underline{\rho} < \bar{\rho} < 1$ and $0 < \lambda_i < 1$; and in particular: $(1/n)\sum_{i=1}^{n} \mathbb{E}\{\bar{\ell}_{i,t}\} > 0$ implying that $\phi_{i,t} \to +\infty$ leading to all high type agents to learn their true type correctly. The situation is distinctively different in a society of low type agents. Repeating the above analysis using $\underline{r}_{i,t}$ instead of $\bar{r}_{i,t}$ we obtain that for all low type agents $i \in [n]$: $\phi_{i,t} \to (t/n)\sum_{i=1}^{n} \mathbb{E}\{\underline{\ell}_{i,t}\}$ with probability one, as $t \to \infty$, where:

$$\mathbb{E}\{\underline{\ell}_{i,t}\} = (\lambda_i + (1-\lambda_i)\underline{\rho}) \log\left(\bar{\rho}/\underline{\rho}\right) + (1-\lambda_i)(1-\underline{\rho}) \log\left((1-\bar{\rho})/(1-\underline{\rho})\right).$$

Defining $\lambda_{avg} = (1/n)\sum_{i=1}^{n} \lambda_i$ we get that if

$$(\lambda_{avg} + (1-\lambda_{avg})\underline{\rho}) \log\left(\bar{\rho}/\underline{\rho}\right) + (1-\lambda_{avg})(1-\underline{\rho}) \log\left((1-\bar{\rho})/(1-\underline{\rho})\right) > 0$$

or equivalently if

$$\left(\frac{\bar{\rho}}{\underline{\rho}}\right)^{(\lambda_{avg}+(1-\lambda_{avg})\underline{\rho})} \left(\frac{1-\bar{\rho}}{1-\underline{\rho}}\right)^{(1-\lambda_{avg})(1-\underline{\rho})} > 1, \tag{10}$$

then the society of high type agents reach a consensus on a false state; incorrectly believing themselves to be of a higher type. An analysis identical to the one in footnote 3 in Appendix A.1 reveals that (10) is in fact always true for any choice of parameters $0 < \lambda_{avg} < 1$, and $0.5 < \underline{\rho} < \bar{\rho} < 1$.

References

Hal R Arkes. The consequences of the hindsight bias in medical decision making. *Current Directions in Psychological Science*, 22(5):356–360, 2013.

Leonard Berlin. Statement of leonard berlin, m.d., to the u.s. senate committee on health, education labor and pensions: Mammography quality standards act reauthorization. 2003. Available online.

Daniel M Bernstein, Edgar Erdfelder, Andrew N Meltzoff, William Peria, and Geoffrey R Loftus. Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2):378, 2011.

Daniel M Bernstein, Alexander Maurice Wilson, Nicole LM Pernat, and Louise R Meilleur. Auditory hindsight bias. *Psychonomic bulletin and review*, 19(4):588–593, 2012.

Daniel M Bernstein, André Aßfalg, R Kumar, and Rakefet Ackerman. Looking backward and forward on hindsight bias. *Handbook of metamemory*, (November), 2015. ISSN 00010782. doi: 10.1145/1516046.1516071.

Bruno Biais and Martin Weber. Hindsight bias, risk perception, and investment performance. *Management Science*, 55(6):1018–1029, 2009.

Hartmut Blank, Steffen Nestler, Gernot von Collani, and Volkhard Fischer. How many hindsight biases are there? *Cognition*, 106(3):1408–1440, 2008.

Colin Camerer, George Loewenstein, and Martin Weber. The curse of knowledge in economic settings: An experimental analysis. *The Journal of Political Economy*, pages 1232–1254, 1989.

A Coolin, DM Bernstein, AE Thornton, and WL Thornton. Inhibition and episodic memory impact age differences in hindsight bias. *Experimental Aging Research*, 40:357–374, 2014.

Alisha Coolin, Edgar Erdfelder, Daniel M Bernstein, Allen E Thornton, and Wendy Loken Thornton. Inhibitory control underlies individual differences in older adults' hindsight bias. *Psychology and aging*, 31(3):224, 2016.

David Danz, Dorothea Kübler, and Julia Schmid. On the failure of hindsight-biased principals to delegate optimally. *Management Science*, 2015.

Nicholas Epley and Thomas Gilovich. The anchoring-and-adjustment heuristic why the adjustments are insufficient. *Psychological science*, 17(4):311–318, 2006.

Edgar Erdfelder and Axel Buchner. Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(2):387, 1998.

Edgar Erdfelder, Martin Brandt, and Arndt Bröder. Recollection biases in hindsight judgments. *Social Cognition*, 25(1):114, 2007.

K Fiedler, M von Sydow, MW Eysenck, and D Groome. Heuristics and biases: beyond tversky and kahneman's (1974) judgment under uncertainty. *Cognitive Psychology: Revisiting the Classical Studies, eds Eysenck MW, Groome D., editors.(Los Angeles, US: Sage*, pages 146–161, 2015.

Baruch Fischhoff. Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, 1(3):288, 1975.

Baruch Fischhoff. The perceived informativeness of factual information. Technical report, DTIC Document, 1976.

Baruch Fischhoff. For those condemned to study the past: Heuristics and biases in hindsight. *D. Kahneman, P. Slovic, A. Tversky, eds. Judgment Under Uncertainty: Heuristics and Biases.*, page 332–351, 1982.

Baruch Fischhoff. An early history of hindsight research. 2007.

Baruch Fischhoff and Ruth Beyth. I knew it would happen: Remembered probabilities of once—future things. *Organizational Behavior and Human Performance*, 13(1):1–16, 1975.

Erin M Harley. Hindsight bias in legal decision making. *Social Cognition*, 25(1):48, 2007.

Erin M Harley, Keri A Carlsen, and Geoffrey R Loftus. The" saw-it-all-along" effect: demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5):960, 2004.

Wolfgang Hell, Gerd Gigerenzer, Siegfried Gauggel, Maria Mall, and Michael Müller. Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, 16(6):533–538, 1988.

Ulrich Hoffrage, Ralph Hertwig, and Gerd Gigerenzer. Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3):566, 2000.

Mark R Leary. Hindsight distortion and the 1980 presidential election. *Personality and Social Psychology Bulletin*, 8(2):257–263, 1982.

Therese A Louie. Decision makers' hindsight bias after receiving favorable and unfavorable feedback. *Journal of Applied Psychology*, 84(1):29, 1999.

Kristóf Madarász. Information projection: Model and applications. *The Review of Economic Studies*, 2011.

Melvin M Mark and Steven Mellor. Effect of self-relevance of an event on hindsight bias: The foreseeability of a layoff. *Journal of Applied Psychology*, 76(4):569, 1991.

Patrick A Müller and Dagmar Stahlberg. The role of surprise in hindsight bias: A metacognitive model of reduced and reversed hindsight bias. *Social Cognition*, 25(1):165, 2007.

Steffen Nestler, Hartmut Blank, and Gernot von Collani. Hindsight bias doesn't always come easy: causal models, cognitive effort, and creeping determinism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5):1043, 2008.

Mark V Pezzo. Hindsight bias: A primer for motivational researchers. *Social and Personality Psychology Compass*, 5(9):665–678, 2011.

Rüdiger F Pohl and Wolfgang Hell. No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, 67(1): 49–58, 1996.

Rüdiger F Pohl, Michael Bender, and Gregor Lachmann. Hindsight bias around the world. *Experimental psychology*, 49(4):270, 2002.

Rüdiger F Pohl, Ute J Bayen, and Claudia Martin. A multiprocess account of hindsight bias in children. *Developmental Psychology*, 46(5):1268, 2010.

M. A. Rahimian and A. Jababaie. Learning without recall: A case for log-linear learning. *IFAC-PapersOnLine*, 48(22):46–51, 2015.

M. A. Rahimian and A. Jadbabaie. Learning without recall in directed circles and rooted trees. In *2015 American Control Conference (ACC)*, pages 4222–4227, July 2015. doi: 10.1109/ACC.2015.7171992.

M. A. Rahimian and A. Jadbabaie. Bayesian heuristics for group decisions. *arXiv:1611.01006*, November 2016.

Neal J Roese and Kathleen D Vohs. Hindsight bias. *Perspectives on Psychological Science*, 7(5):411–426, 2012.

Stefan Schwarz and Dagmar Stahlberg. Strength of hindsight bias as a consequence of meta-cognitions. *Memory*, 11(4-5):395–410, 2003.

Orit E Tykocinski and Noa Steinberg. Coping with disappointing outcomes: Retroactive pessimism and motivated inhibition of counterfactuals. *Journal of Experimental Social Psychology*, 41(5):551–558, 2005.

Gordon Wood. The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, 4(2):345, 1978.