The Biasing Effects of Memory Distortions on the Process of Legal Decision-Making

DIEGO NOCETTI
Clarkson University

This article investigates the impact of memory limitations and memory distortions on the process of legal decision-making. I develop a simple framework in which jurors establish interim judgments, which are later used by a memory technology to reconstruct, in an inductive process, evidence related to those judgments. The resulting behavior matches a number of stylized facts that are inconsistent with the standard Bayesian framework. I show that beliefs in a given hypothesis may remain unchanged, and may even be strengthened, in the face of disconfirming evidence. This, in turn, implies that the beliefs of two jurors with different memory technologies may deviate further apart as they receive new information, accounting for heterogeneous, i.e. opposite verdict choices, and strongly held beliefs after a large amount of information is presented. Finally, I show that in this setup the probability of legal errors is highest for moderate strengths of evidence.

1. INTRODUCTION

Bayesian inference has not earned much empirical support as a descriptive theory of juror decision-making. This should not be too surprising, since jurors are prone to the same biases usually found in other contexts of decision-making under uncertainty. One possible explanation for the failure of Bayesian inference lies in the jurors’ limited availability of information previously encoded in memory. Simply put, some evidence may be forgotten and the evidence that is recalled may be a biased sample of the evidence presented during the trial. Furthermore, beyond ‘sampling selection bias,’ the recalled history may be a distorted version of the factual history, resulting in ‘non-sampling errors’ that systematically distort beliefs. The aim of this article is to investigate the impact of memory deficits and memory distortions in the process of legal decision-making.

*I thank Bill Smith, Luciana Echazu, and two anonymous referees for many helpful comments.*
Specifically, we conjecture that when a juror evaluates competing hypotheses (e.g., whether the defendant is guilty or not guilty) she retrieves evidence from memory and, given the recalled history, applies the Bayes theorem to judge the likelihood of each alternative. Some evidence is readily available, either because it has just been presented or because of its high level of activation in long-term memory (e.g., a salient piece of evidence). Other events are not simply discarded. Instead, a reconstructive process takes place. Our core assumption is that the likelihood that the juror retrieves one particular realization of an episode (e.g., the witness saw or she did not see the defendant carrying a gun) is determined by the diagnosticity of the information available.

In a simple framework we show that this process of memory reconstruction generates a situation that is similar to that in models of informational herding (e.g., Banerjee 1992; Bikhchandani et al., 1992). In that literature, the actions of a few initial decision-makers introduce an informational externality to subsequent decision-makers who optimally decide to follow their actions. In the proposed framework, the juror establishes interim judgments which are assumed to be more salient than the information leading to them. As a result, if the informational content of the initial judgments is greater than that of the newly acquired evidence, the memory system will tend to recreate memories consistent with those beliefs and the juror’s current judgment will tend to replicate that of her predecessors (selves). In essence, a ‘confirmatory bias’ (Pitz et al., 1967; Lord et al., 1979) arises endogenously. We dub this process of imperfect information aggregation intra-personal herding/cascades.

In contrast to the standard Bayesian framework that predicts that the beliefs of different jurors will tend to converge when they receive more information, our framework accounts for heterogeneous (i.e., opposite verdict choices) and strongly held beliefs after a very large amount of information is presented. Furthermore, our model predicts that any preference/bias towards a particular hypothesis may produce initial attitudes that do not change, and may be strengthened, over the course of the trial. For instance, it is well known that pre-trial publicity and the defendant’s criminal history can influence jurors’ ultimate verdict (see, e.g., Carroll et al., 1986; Devine et al., 2001). We show that these extra-evidentiary factors will have the most influence when the evidence presented during the trial is more ambiguous, providing a memory-based rationale for Kalven and Zeisel’s (1966) ‘liberation’ hypothesis.

1 Nickerson (1998) provides a full review of belief perseverance and confirmatory bias phenomena.
2 For example, Roberts (1987:71) argues that over 80 percent of jurors form unchanging beliefs about cases immediately after the opening statements (see also Wenner and Cusimano, 2000).
In addition to explaining the proclivity to maintain incorrect judgments, our analysis offers interesting insights about the interaction between the decision-making environment and the dynamics of beliefs. Specifically, we show that, from an unbiased pre-trial perspective, the probability of wrongful conviction and of wrongful acquittal is highest for moderate strengths of evidence. Intuitively, it is highly unlikely that a juror maintains an incorrect attitude if she receives unequivocal evidence. Yet, although more informative evidence reduces the probability that a juror comes to believe that the incorrect hypothesis is more likely, the probability of recreating evidence that coincides with the judgments that trigger an erroneous belief is also higher when the incoming information is less equivocal (since in this case judgments contain higher diagnosticity).

In a closely related paper, Rabin and Schrag (1999) present a framework of belief formation with confirmatory bias. In addition to the application of confirmatory bias to juror decision-making and its related implications, the emphasis that I place on memory reconstruction has the advantage of formalizing an explanation of the underlying mechanism of the bias and offers new insights. For example, the endogenous characteristic of confirmatory bias permits establishing a link between the strength of the bias and the strength of the evidence. Notably, I show that there is a positive relationship between the two, which leads to the result mentioned in the previous paragraph.

The rest of the paper is organized as follows. In the next section I present the basic setup and provide a brief review of the relevant literature on memory reconstruction and memory distortions which provides the motivation for the formal reconstruction mechanism proposed. In section 4 I derive the main results of the paper. Section 5 concludes with a discussion of the limitations and possible extensions of the paper.

2. MODEL

2.1. SETUP

The basic framework draws on Rabin and Schrag’s (1999) model of confirmatory bias and on the literature of informational cascades/herding (see, e.g., Banerjee, 1992; Bikhchandani et al., 1992). Specifically, consider two states of the world \( x \in \{G, NG\} \), where \( G \) and \( NG \) represent the mutually exclusive hypotheses that the defendant is guilty or not guilty. A representative juror holds prior beliefs \( \text{prob}(x = G) = \mu_0 \) and \( \text{prob}(x = NG) = 1 - \mu_0 \). The trial presentation phase lasts for \( T \) periods, during which the juror receives independent and identically distributed signals (e.g., a signal represents the presentation and cross-examination of witnesses) of
the defendant’s guilt, $e_t \in \{g, ng\}$. These signals are correlated with the true state: $prob(e_t = g|G) = prob(e_t = ng|NG) = \lambda > 0.5$, where $\lambda$ is the strength of the signals/evidence.

I assume that in each period the juror employs the likelihood ratio to establish an attitude (which I also refer to as a judgment or interim verdict) towards the defendant’s innocence or guilt, $\nu_t \in \{\Gamma, N\}$: $\nu_t = N$ if $\Omega_t \leq 1$ and $\nu_t = \Gamma$ if $\Omega_t > 1$,

$$\Omega_t = \frac{prob(x = G|n_g, n_{ng})}{prob(x = NG|n_g, n_{ng})} = \frac{\mu_0}{1 - \mu_0} \left(\lambda^{n_g - n_{ng}}\right),$$

and $n_g$ ($n_{ng}$) is the number of recalled signals that support hypothesis $G$ (NG). Finally, after all evidence is presented, the juror establishes a (pre-deliberation) verdict given a standard of proof: convict if $\Omega_t \geq \Omega^*$.

### 2.2. Memory Reconstruction

#### 2.2.1. Evidence of Memory Reconstruction

Give us a dozen healthy memories, well-formed, and our specified world to handle them in. And we’ll guarantee to take any one at random and train it to become any type of memory that we might select – hammer, screwdriver, wrench, stop sign, yield sign, Indian chief – regardless of its origin or the brain that holds it. - E.F. Loftus and H.G. Hoffman (1989:103)

Roediger and McDermott (1995:812) indicate that “All remembering is constructive in nature...The illusion of remembering events that never happened can occur quite readily.” Although Loftus and Hoffman’s quote is a parody of a famous passage by John B. Watson (1939:104), they make clear this pervasive malleability of memory.

Bartlett (1932) was the first to provide clear evidence that memory is not reproductive. He suggested, instead, that individuals reconstruct events from the past using abstract representations (i.e., mental models or, as Bartlett labeled it, schemata) of how the world works. For instance, in Bartlett’s most influential work, participants were asked to read a story based on a Native American legend, ‘The War of the Ghosts,’ and later to reproduce it. Systematically, participants omitted parts of the story, transformed unfamiliar elements into familiar ones.

---

3 See Schacter (1996) for a full review.
(e.g., ‘hunting seals’ became ‘fishing,’ a ‘canoe’ became a ‘boat’) and rationalized the story to make it more coherent with their own cultural background (e.g., ghosts became imagined by the Indian character when wounded).

Similar schema-induced errors were demonstrated by Pennington and Hastie (1988) in the context of juror decision-making, where mock jurors were shown to reconstruct the evidence to coincide with the archetypal crime story. Likewise, Cohen (1981) showed that individuals’ recall of the behavior and appearance of another person depends on the category stereotypes (schema) of that person; specifically, after watching a video, participants that were told that the woman character was a waitress were more likely to describe her as blond and as drinking beer while those told that she was a librarian described her as brown-haired and as drinking wine.

In addition and in conjunction with schema-induced reconstructive errors, memory distortions have been shown to depend on a variety of external and internal sources of misinformation. In a classic experiment, Loftus and Palmer (1974) showed traffic safety films to college students depicting car accidents. One of the questions asked for an estimate of the car’s speed: “How fast were the cars going when they --- each other? With the verb that described the collision varied for different groups, using smashed, collided, bumped, hit, and contacted. Unambiguously, those who got the stronger verbs gave the higher estimates of speed. A follow-up question asked whether there was broken glass (there was no broken glass) and almost one-third of the subjects who had the verb smashed said yes. Furthermore, even minor changes in the wording of the question (e.g., “did you see the broken glass” versus “did you see a broken glass”) had large effects in the recall of the experience (those that received a definite article were more prone to errors).

More relevant to our framework, internal states play an important role in memory reconstruction. For example, Read and Rossen (1982) gave subjects an article to read on nuclear power and subsequently tested recognition for true items, false items (both pro and anti-nuclear), and exaggerated distortions (both pro and anti-nuclear). Subjects that had a prior attitude against (in favor of) nuclear power before reading the material were more likely to ‘recognize’ in a subsequent test consistent items, regardless of whether the items were accurate or not.

Similarly, prior decisions are an important source of memory errors. For instance, in the choice-supportive paradigm (e.g., Mather et al., 2000) a participant

---

4 Broadly, errors in memory reconstruction due to external sources of misinformation is what Schacter (1999) calls the memory sin of suggestibility. Self-generated errors (e.g. those influenced by prior decisions, knowledge, or beliefs) are examples of the sin of bias.
is given positive and negative features of two objects (e.g., a job candidate, a blind date) and has to select one alternative. The participant is later presented with features that correspond to one of the choices or to none of them. The typical finding is that participants tend to attribute positive features to the selected alternative and negative features to the rejected alternative, regardless of whether the features are new or belonged to the other object. This is consistent with previous applied research by Pennington and Hastie (1988) who found that mock jurors were more likely to falsely ‘recognize’ lures corresponding to their own verdict choices than to ‘recognize’ lures associated with opposing verdicts.

Loftus (2003) summarizes the research on memory reconstruction as follows: “Who we are may be shaped by our memories, but our memories are shaped by who we are and what we have been led to believe.” In the next sub-section we present a simple framework that seeks to explain the process by which memories are shaped.

2.2.2. General Description of the Memory Reconstruction Mechanism

The evidence just presented suggests that the memory system integrates the information that is readily available to produce memories that are congruent with that information. But, how does the memory system integrate the evidence to recreate the past?

We envision memory reconstruction as a procedure whereby target events not currently available in the decision-makers’ ‘database’ are imputed in an inductive fashion. The process starts when memory is queried about one or multiple targets (e.g., speed of the car, a particular feature of a job candidate, evidence presented during the trial that help the decision-maker evaluate competing hypotheses). Given the query, the memory system assembles a set of cues, which may include contextual knowledge about the problem at hand (e.g., knowledge about the underlying distribution – or semantic category – of the data-generating process, overt beliefs) and also specific pieces of information (e.g., a salient or recent piece of evidence, prior decisions). These cues are then used as inputs into a mental model (a schema) to create a probability distribution of the alternative realizations of the target (e.g., the car was going fast/slow, job candidate A is smart/dumb). The probability assigned to a particular event (i.e., its level of activation) is given by the conditional likelihood of this event given the cues readily available.

---

5 See McClelland (1995) for a similar conceptual model of ‘spreading activation’ used to explain memory reconstruction and distortions.
For instance, denoting the cues as $E$, the factual event as $T$, and a complementary event as $T^C$, the probability assigned to the factual event would be

$$\Pr(T|T) \Pr(T^C|T) + \Pr(T|T^C) \Pr(T^C|T^C).$$

$\Pr(T|T)$ and $\Pr(T|T^C)$ represent the conditional probabilities of the evidence given the alternative realizations of the target event while $\Pr(T)$ and $\Pr(T^C)$ are the base (unconditional) levels of activation. Differences in these base levels of activation may represent, for example, decay, whereby memories that are not rehearsed become less available over time (i.e., $\Pr(T)$ gets closer to 0.5), or differential encoding and maintenance of the incoming information – e.g., events that require a deeper level of analysis at the moment of encoding (Craik and Lockhart, 1972; Craik and Tulving, 1975) and, more generally, that are more salient (Murdock, 1960; Neath, 1993) maintain a high level of activation. Supposedly, the cues that are used to recreate the event represent information with a very high level of unconditional activation.

Finally, the system samples (retrieves) one realization from the distribution and, just like an imputed event in a statistician’s database, the decision-maker takes this information as the true history.

### 2.2.3. Evidence Reconstruction During the Trial

To simplify future discussion, we will refer to each temporal self (juror) by their respective periods of control $(t, t+1,...)$. Let $E_{t+N} = e_t, e_{t+1},..., e_{t+N}$ represent the public history of realized signals from periods $t$ to $t+N$ and let $E^m_{t+N}$ be the subset of evidence that self $t+N$ has readily available in memory. In general, we take $E^m_{t+N} = e_{t+N}$ such that only the last piece of evidence is readily available. Essentially, this is a form of rapid memory decay; in section 3.2 we briefly discuss how more information affects the decision problem. Finally, $V_{t+N-1} = V_t, V_{t+1},..., V_{t+N-1}$ represents the history of interim verdicts established by selves $t$ to $t+N-1$. I assume that these judgments are never forgotten; as suggested before, the intuition of this assumption is that the juror requires a deeper level of analysis to produce those judgments.

---

6 Of course, the target event itself could have a very high level of unconditional activation, in which case it would be reproduced directly without the need of the imputation procedure.

7 Specifically, Craik and Lockhart’s (1972) proposed that ‘trace persistence is a positive function of the levels of perceptual analysis, with deeper levels of analysis associated with more elaborate, longer lasting, and stronger traces.’
In line with the ideas presented above, the process of evidence reconstruction – for self \(T+N\) – consists on the transformation of the evidence received in period \(t+i\) \((i=0,1,...,N)\) into \(\hat{e}_{t+i}\), a variable that satisfies

\[
\hat{e}_{t+i} = \begin{cases} 
g & \text{with probability } q_{t+i} = \Pr(e_{t+i} = g \mid V_{t+N}, E_{t+N}^m) \\
g & \text{with probability } 1 - q_{t+i} = \Pr(e_{t+i} = ng \mid V_{t+N}, E_{t+N}^m) \end{cases}
\]

In other words, \(q_{t+i}\) \((1 - q_{t+i})\) is the posterior probability that self \(T+N\) retrieves a signal \(g\) \((ng)\) for the evidence received by self \(t+i\), based on the history of judgments (and the evidence produced given those actions) and on the subset of evidence readily available.

To see the intuition of the memory reconstruction process, consider the case of a juror that receives the evidence \((g, ng)\). Given the first signal, self \(t\) favors a guilty verdict (i.e., \(\nu_t = \Gamma\)). Then, for the first signal, the reconstructive process in any period in the future satisfies

\[
q_t = \Pr(e_t = g \mid \nu_t = \Gamma) = \frac{\Pr(\nu_t = \Gamma \mid e_t = g)}{\Pr(\nu_t = \Gamma \mid e_t = g) + \Pr(\nu_t = \Gamma \mid e_t = ng)} = 1,
\]

that is, having favored hypothesis \(G\) in the first period, there is no doubt that the signal was \(g\). Similarly, since self \(t+1\) retrieves the correct signal received by self \(t\), the new judgment will be \(\nu_{t+1} = N\), which gives perfect information to future selves about the evidence received, i.e., \(q_{t+1} = \Pr(\nu_{t+1} = N \mid \Gamma \& N) = 1\). It follows that, as long as previous decisions are completely informative, the memory technology can separate the realized signal/payoff type and the recalled history is identical to the public history.

3. ANALYSIS

In the following subsections I discuss a number of implications of the preceding framework. Although I place particular emphasis on the implications

\[\text{Notice that the juror’s judgments are based on previous inferences, but not on the strength of his or her beliefs – the likelihood ratio. Since the likelihood ratio is an efficient summary of information, recalling this information would allow her to always aggregate information as if she was recalling information. This assumption is central in economic models of limited memory (see, e.g., Mullainathan, 2002).}\]
for legal decision-making, the results apply to many different situations (e.g.,
voting, consumption and investment decisions).

3.1. **BELIEF PERSEVERANCE**

The human understanding when it has once adopted an opinion draws all
things else to support and agree with it. And though there be a greater
number and weight of instances to be found on the other side, yet these it
either neglects and despises, or else by some distinction sets aside and
rejects, in order that by this great and pernicious predetermination the
authority of its former conclusion may remain inviolate. - Francis Bacon
(1620, 1960)

It has long been recognized that people usually maintain their beliefs despite
evidence to the contrary. Motivated by the literature on informational
herding/cascades, I provide a formal definition of belief perseverance,

**Definition.** Belief perseverance (intra-personal herding/cascades)
occurs if, based on the information provided by previous judgments, the juror’s current
judgment is the same as that of her predecessors (selves), regardless of the actual
realization of the signals received.

The following proposition establishes the first result of the paper,

**Proposition 1.** Memory reconstruction may generate intra-personal
herding.

The intuition is simple. Assume that the defendant is not guilty, that \( \mu_0 = 0.5 \),
and that the jurors receive a sequence of evidence \( (g, g, ng, ng, ...) \). As in the
example in the previous subsection, the two initial judgments are completely
informative, i.e., \( q_t = q_{t+1} = 1 \). Now, however, think about the inductive process
of memory reconstruction for self \( t + 3 \) (after the acquisition of the fourth
signal). In addition to the information provided by the initial decisions, the juror
maintains the same judgment in period \( t + 2 \). This decision, however, is
uninformative, i.e., whether she receives \( g \) or \( ng \) she maintains the same
judgment. Yet, given the recalled history and the current observed payoff, there
is a higher probability that the payoff in period \( t + 2 \) favored state \( G \), i.e.,
\( q_{t+2} > 0.5 \) since

\[
\Pr(v_{t+2} = v_G | \Gamma \Rightarrow v_t = g, \Gamma \Rightarrow v_{t+1} = g, v_{t+3} = ng) > \\
\Pr(v_{t+2} = v_B | \Gamma \Rightarrow v_t = g, \Gamma \Rightarrow v_{t+1} = g, v_{t+3} = ng).
\]
Accordingly, there is a higher probability that the juror retrieves \( g \) for the evidence received in period \( t + 2 \).

The situation is similar to that studied in models of informational herding. Initial decisions generate useful information that the memory technology uses to reconstruct the evidence. If, following the actual evidence, the juror’s two initial judgments are the same, her new judgments are uninformative and ‘hard’ information stops accumulating. At this point, even if the memory system were to discard all events that cannot be reproduced with certainty, the juror would maintain the same belief. Our assumption that the memory system reconstructs all the evidence previously received predicts, in addition, that the juror will strengthen her beliefs in the initially preferred alternative as she receives more information. \(^9\)

Therefore, in contrast to the common prediction in the standard Bayesian framework that beliefs should tend to equalize as more information arrives, our model says that the beliefs of two jurors receiving the same evidence may become polarized (see, e.g., Lord et al., 1979).\(^{10}\) For example, while a juror with limited recall may be in an incorrect cascade and will tend to confirm her possibly incorrect beliefs when new evidence is presented, a juror with no memory limitations will, on average, believe in the correct hypothesis.\(^{11}\)

3.2. Jurors’ Delusions: The Verdict After an Infinite Number of Signals

After a large amount of evidence a juror that has no memory limitations will become almost certain in the correct hypothesis. This is also true for a juror that retrieves a limited but unbiased amount of information. This fails to account for the regularity of heterogeneous beliefs after a large amount of

\[^9\] That is, on expectation she will believe more strongly that the true state is \( G \) after receiving the fourth (\( n \)) piece of evidence, i.e. since

\[
\lambda \left[ g_{n+2} \lambda + (1-g_{n+2})(1-\lambda) \right] > \lambda,
\]

where the LHS of the inequality is the expected likelihood ratio of self \( t+3 \) from the perspective of an outside observer that knows the juror’s process of memory reconstruction and the RHS is the likelihood ratio prior to receiving the fourth payoff.

\[^{10}\] This prediction is consistent with ‘mere thought’ effects in attitude polarization (see e.g. Tesser, 1978, Tesser and Leone, 1977), whereby simply thinking about an object may polarize one’s attitudes towards that object. For example, retrieval of information can strengthen existing beliefs or make new attitude-consistent beliefs more accessible.

\[^{11}\] Polarization may also occur when some jurors utilize memory-based judgments and others utilize on-line judgments (see Hastie and Park, 1986), whereby the latter implies that jurors have a clear recall of their previous beliefs and, as a result, no bias would exist.
evidence has been presented; most notably, hung juries. The following proposition establishes that the process of evidence reconstruction may generate strong beliefs in the incorrect hypothesis.

**Proposition 2.** There is a positive probability that a cascade lasts forever and the juror becomes (almost) certain of the incorrect hypothesis.

Assume that the defendant is not guilty, that $\mu_0 = 0.5$, and that a $G$ cascade occurs in period $t+j$, i.e., period $t+j+1$ is the first period in which the judgment is uninformative. Denote $k \equiv n_g - n_q$ and notice that the fact that a cascade is triggered in period $t+j$ implies that $k \geq 2$ and $q_{t+j} = q$, the posterior probability of retrieving a $g$ signal, is higher than 0.5. Then, the probability that a $G$ cascade lasts forever is $P = \left[1 - \left(1 - q/q \right)^k \right]$. 

The term $P$ increases with $k$ and with $q$. Once a cascade occurs the probability that it lasts forever will, on average, increase over time, as $k$ does. Intuitively, as long as the juror is in a cascade, the externality of the judgments initiating the cascade will affect more previous selves. This, in turn, implies that if a cascade is not dislodged the juror will become (almost) certain of the incorrect hypothesis.

Possibly more surprising is that $P$ increases with the strength of the evidence, since $q$ increases with $\lambda$. This implies that, if a $G$ cascade is triggered the innocent defendant may be better off with weaker evidence in his favor. The reason for this quite unintuitive result lies in the way that memory reconstructs the evidence. If a cascade is triggered, only a few judgments are completely informative. In trying to infer the evidence received by previous selves, the memory technology weights the information provided by the judgments initiating the cascade against the most recent information received. If the initiating signals favored a $G$ hypothesis the juror will more likely maintain her beliefs in this hypothesis if the probability of receiving a $g$ signal is low.

One may expect that, at least from an unbiased pre-trial perspective, the probability of legal errors decreases with the strength of the evidence. The following proposition establishes that such intuition only holds for a certain range of values of $\lambda$.

---

12 Note that $q$ is not the same in each period as it depends on the last realized signal.

13 This is a restatement of a general result due to Feller (1968) and used in the context of confirmatory bias by Rabin and Schrag (1999).
Proposition 3. From a pre-trial perspective, the probability of wrongful conviction and of wrongful acquittal after an infinite number of signals (and for any standard of proof) is highest for moderate strengths of evidence.

The appendix contains a formal proof, but the intuition is also simple. On the one hand, the probability of retrieving evidence that coincides with the judgments that initiate a cascade is higher when the signals are more informative, but, on the other hand, more informative evidence reduces the probability that an incorrect cascade will be triggered. When signals are very noisy the former effect dominates since memory reconstruction is more likely to terminate a cascade. When the signals are very informative the latter effect will dominate since the probability that a wrong cascade will occur is small. When $\lambda$ is at moderate levels, none of these effects will dominate and the probability of legal errors will be highest.

It is important to notice that, although a juror may come to believe with near certainty in the incorrect hypothesis, beliefs are fragile. Why? First, although the evidence retrieved is biased in favor of the initial judgments, at any point in time a cascade may be dislodged due to the stochastic nature of memory retrieval. As the $DM$ receives more information, however, it is less likely that she will abandon the incorrect hypothesis as the evidence retrieved will approach the true (biased) distribution that the memory technology creates. Second, and possibly more important, as it occurs in models of informational herding, it takes only a piece of additional 'hard' information to burst the cascade.

This does not mean, however, that intra-personal cascades cannot occur when $V_{r+1}^{m}$ contains more information. Indeed, if the additional information that enters the memory reconstruction process is unbiased, the probability that an incorrect cascade occurs will be lower. Yet, the effect of additional information is similar to the case of more informative evidence. More information decreases the probability that an incorrect cascade is triggered but, once a cascade is triggered, more previous decisions will be informative and abandonment will be less likely. Of course, more information would be detrimental if memory is also selective, such that the information readily available correlates with the judgments previously adopted or if the juror selectively searches and samples information, such that the underlying distribution is biased.

For instance, if the last $b$ realized signals are readily available, a sufficient condition for a cascade to occur is a run of $b+2$ signals of the same type. That is, if the true state is $B$ and the last $b$ signals are readily available, given $b+2$ signals, the probability that a ‘$G$’ cascade occurs is $(1 - P)^{b+2}$ and the probability of a ‘$B$’ cascade is $P^{b+2}$.

Review of Law & Economics, © 2008 by bepress
Similarly, it is straightforward that attitudinal (and, in general, extra-evidentiary) factors may exaggerate the biasing effects of memory reconstruction. For instance, suppose that the juror initially believes that type $G$ is more likely, $\Pr(G) = P_0 > \Pr(NG)$, and let $P_0/(1-P_0) > \lambda/(1-\lambda)$, i.e., the prior likelihood has a larger weight than a given realized signal. In this case, judgments are uninformative since the juror will maintain her beliefs regardless of initial evidence presented. As a result, when prior attitudes are so strongly entrenched that the initial judgments are uninformative, and assuming that these overt beliefs are readily available, memory reconstruction will be intrinsically biased (by those attitudes) to reproduce evidence in favor of the a priori preferred alternative. Moreover, these extra-evidentiary factors will have the most influence when the evidence presented during the trial is very noisy/weak (i.e., the more informative the evidence, the lower the weight of the prior beliefs), which is consistent with a memory-based explanation of Kalven and Zeisel’s (1966) ‘liberation’ hypothesis.\footnote{MacCoun (1989) suggests that this may reflect the influence of extra-evidentiary factors in the stringency of the standard-of-proof thresholds rather than the diagnosticity of the evidence.}

3.3. Summary of Legal Implications

In the standard (Bayesian) model of legal decision-making, jurors use all the available evidence to reach a verdict. That framework provides a strong prediction: given sufficient information, all jurors should have (approximately) the same – correct – beliefs regarding the defendant’s culpability, regardless of any initial bias. Our analysis shows, however, that those predictions may not hold when jurors have limited memory. Specifically, the reconstructive nature of memory may generate a long-lasting tendency to maintain incorrect beliefs in the face of disconfirming evidence. Therefore, given different memory capabilities, two jurors may reach completely opposite conclusions after being exposed to the same evidence.

Furthermore, our model predicts that any preference/bias towards a particular hypothesis may produce initial attitudes that do not change, and may be strengthened, over the course of the trial. So, for instance, a juror may form unchanging beliefs about a case immediately after the opening statements (Roberts, 1987:71), after learning the defendant’s criminal record (Barnett, 1985) or after being exposed to pre-trial publicity (see, e.g., Carroll et al., 1986; Devine et al., 2001).

Of course, the voir dire process is specifically designed to eliminate jurors with prior biases and judges specifically instruct jurors not to base their decisions on extra-evidentiary factors. Yet, our model of (unconscious) memory reconstruction says that the influence of prior biases may be more subtle than
their direct effect on overt beliefs. For instance, a highly salient piece of extralegal information (e.g., a prosecutor that mentions the defendant’s criminal record) will influence not only a juror’s beliefs but also how the juror recalls the evidence presented. Similarly, voir dire itself may change jurors’ beliefs and the subsequent recollection of evidence (see, e.g., Dexter et al., 1992).

Also in contrast to the Bayesian framework, our model predicts that the probability of legal errors does not decrease monotonically with the strength of the evidence and may be influenced by the order of evidence presentation. Specifically, we confirmed the common belief that jurors’ first impressions (together with the most recent information presented) may be a deciding factor in the final verdict. In contrast to our framework with an exogenous data-generating process, this suggests that counsel may present the evidence strategically to manipulate jurors’ memory distortions (e.g., the prosecution or the defense may refrain from presenting evidence that favors their case if they believe that such evidence will strengthen the opposite party’s case – proposition 1, or they may try to alter the strength and order of the signals to increase/decrease the probability of legal errors – proposition 3). Embedding the proposed process of memory reconstruction in a model of endogenous information generation may prove fruitful for future research.

Overall, our analysis provides some support for those who express reservations about the jurors’ ability to evaluate and integrate trial evidence in an efficient and unbiased manner. Although judicial decision-making is not immune to decision-making biases (e.g., Robbennolt, 2005), judges’ expertise at identifying relevant information and, more importantly, the possibility to access the trial transcripts would eliminate most if not all memory biases described here. Of course, jurors’ limited capabilities and limited understanding of complex evidence is one of the oldest arguments in the “trial-by-jury versus trial-by-judge” debate, and our framework simply formalizes one aspect of this argument. Without trying to settle a debate that involves much broader issues than the ones analyzed here, our analysis provides a clear policy implication for the current system of juror decision-making: procedural rules governing the adjudicative process should be targeted at enhancing jurors recall. For example, it has been shown that giving preliminary instructions on the applicable law (ForsterLee and Horowitz, 2003), note-taking (Rosenhan et al., 1994; ForsterLee and Horowitz, 2003), providing notebooks with relevant trial documents (Dann et al., 2004), and Judge’s summation and direct questioning of witnesses (Collet and Kovera, 2003) enhance a juror’s ability to process and recall relevant evidence. Similarly, in light of our result that cumulative evidence may strengthen jurors’ beliefs in a given hypothesis, particular attention should be paid to evidence rules regarding the relevancy of the information
presented during trial and, in particular, the exclusion of cumulative evidence (e.g., rule 403, Federal Rules of Evidence).

4. CONCLUSIONS

This paper proposed a theory that synthesizes existing evidence from the psychology literature on memory distortions and economic theories of informational herding. The process of decision-making that we propose is closely related to the *story model* of juror decision-making (Pennington and Hastie, 1981, 1988, 1992). Both frameworks share the assumption that jurors base their judgments on a story that is constructed in memory given some background knowledge of the case. The formal model that I present permits investigating the dynamics of belief formation and the circumstances under which biases and legal errors are more likely to occur.

Our results regarding jurors’ evaluation of evidence are also in line with the findings of the ‘cascaded inference’ research (see Schum and Martin, 1982, for a review). In a typical experiment, mock jurors are asked to judge the likelihood of the evidence in three equivalent ways: by aggregating all the evidence into a single likelihood ratio, by evaluating the likelihood ratio of each piece of evidence, or by evaluating the conditional probabilities of the hypothesis given the evidence. Schum and Martin (1982:105) conclude that “Results show that persons, when required to mentally combine a large amount of probabilistic evidence, exhibit certain inconsistencies such as treating contradictory testimony as corroborative testimony and double-counting or overvaluing redundant testimony. However, when people are asked to make assessments about the fine-grained logical details of the same evidence, these inconsistencies do not occur.” This is precisely what our model predicts.16 The reason for the differences, according to our model, would be biases in memory reconstruction that occur when individuals have to aggregate a large amount of evidence.

At a more general level, our model is related to the explanation of belief perseverance as a byproduct of cognitive dissonance. Originally developed by Festinger (1957), the theory of cognitive dissonance proposes that individuals have an innate *preference* for preserving cognitive and behavioral consistency and, therefore, when two or more cognitions are inconsistent the cognitive system adapts those cognitions to reduce the level of dissonance. A similar line of argument comes from the psychology literature on self-serving biases (e.g.,

---

16 In our model, since by assumption jurors behave in a consistent Bayesian fashion, being able to ‘make assessments about the fine-grained logical details of the same evidence’ refers to the situation in which the juror is able to recall all pieces of evidence.
Pyszczynski and Greenberg, 1987; Pyszczynski et al., 1985; Sanitioso et al., 1990). As Sanitioso et al. (1990) explain, "People attempt to construct a rational justification for the conclusions that they want to draw. To that end, they search through memory for relevant information, but the search is biased in favor of information that is consistent with the desired conclusions. If they succeed in finding a preponderance of such consistent information, they are able to draw the desired conclusion while maintaining an illusion of objectivity."

In this paper, instead, we do not assume a taste for consistency or any motivational bias. The cognitive mechanism of memory reconstruction that we propose generates endogenously a tendency to preserve cognitive and behavioral consistency over time. This is in line with Bem’s (1967, 1972) self-perception theory, in which individuals “learn” their own attitudes by inferring them from past behavior.17 For example, in our context of legal decision-making, a juror’s past attitudes (interim verdicts) work as signals in the reconstruction of the trial evidence. When the informational content of the initial attitudes is greater than that of the newly acquired evidence, the memory system recreates memories consistent with those beliefs and the juror’s current judgment tends to replicate that of her predecessors (selves). It should be clear, however, that either a distaste for dissonance (e.g., the information readily available correlates with the judgments previously adopted) or a self-serving bias (e.g., the juror selectively searches and samples – or pays attention to – information consistent with the desired conclusions) would enhance the effects of memory reconstruction discussed here.18

This paper’s main focus has been the pre-deliberation phase of the trial. One may argue that, even if memory distortions are a problem during this phase, once jurors are allowed to deliberate they will share information and arrive at the correct conclusion. An extensive line of research has shown, however, that information sharing within groups is far from perfect.19 For example, many times discussions only focus on a limited number of facts that are already available to the majority (Janis, 1982; Stasser, 1992; Pritchard and Keenan, 2002).20 Also,

---

17 Specifically, Bem (1972:2) argues that “Individuals come to ‘know’ their own attitudes, emotions, and other internal states partially by inferring them from observations of their own overt behavior and/or the circumstances in which this behavior occurs. Thus, to the extent that internal cues are weak, ambiguous, or un-interpretable, the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual’s inner states.”

18 For example, a biased level of activation in judgment-consistent memory traces would act essentially as a biased prior belief in the hypothesis that such judgment favors.

19 For a review of the literature on group versus individual biases in judgment, see Kerr et al. (1996).

20 In our framework, for example, discussion of evidence presented at the beginning or the end of the trial would have no effect since this information is already available to all jurors.
a juror may perceive the attitudes of the rest of the group as informative, which may confirm and strengthen her beliefs if her beliefs are those of the majority (and not necessarily the true state) or disconfirm and weaken her own attitude if her beliefs are different from the majority (see, e.g., Myers and Kaplan, 1976). Finally, in close connection with the framework analyzed here, memory distortions may be ‘contagious’ (e.g., Roediger et al., 2001; Gabbert et al., 2003) — i.e., the result of other individuals’ faulty memory. As a result, an individual’s distorted view of the past may permeate other people’s memories, possibly exaggerating a pre-discussion bias.

Empirically testing the model would be an important step forward. To do so, the experimenter would need to vary evidence strength and the order of evidence presentation and compare recollection of facts and accuracy of memory with strength of beliefs. As one referee noted, even if such tests provide support for the model, one should be cautious interpreting the results. We are not proposing here that actual calculations of Bayesian updating are taking place. Instead, based on the evidence on memory reconstruction presented above, we suggest that the output of the reconstruction mechanism is observationally equivalent to what a Bayesian statistician who needs to impute missing data would do. Furthermore, we show that, given this imputation procedure, the sequential presentation of evidence may lead to a confirmation bias. It would be difficult, however, to differentiate and isolate the purely cognitive bias proposed here with a motivational bias that may result using the same procedures.

We believe that it is also worth investigating the possibility of more and less sophisticated processes of memory reconstruction. On the one hand, the imputation procedure that we propose is largely efficient, since the memory technology integrates the information readily available as a Bayesian observer would. On the other hand, memory reconstruction is myopic, as it does not take into account the externality that current attitudes impose on future selves. Instead, a ‘rational’ memory technology would weight the costs of deviations from the current expected optimum with the benefits of improving information to future selves.
Maintaining the assumption that NG is the true state, consider the probability that a G cascade occurs after the initial two signals and the juror becomes (almost) certain of this hypothesis, i.e., the probability of wrongful conviction for any standard of proof. First note that the probability of a G cascade after two signals is \((1-\lambda)^2\). Also, the posterior probability of retrieving a g signal once a cascade is triggered is 
\[
q = \lambda^n/(1-\lambda)^n + \lambda^n
\]
with \(n = 1\) if the last realized signal is ng, which occurs with probability \(\lambda\), and \(n = 3\) if the last realized signal is g, which occurs with probability \(1-\lambda\). Therefore, from a pre-trial perspective, 
\[
q^* = \left[\left(1-\lambda\right)^3/\left(1-\lambda\right)^3 + \lambda^3\right] + \lambda^2
\]
is the probability that the juror will retrieve a g signal if a G cascade is triggered. It follows that the probability that a G cascade is triggered after two signals and that this cascade determines a guilty verdict for any standard of proof is:
\[
(1-\lambda)^2 \times \left[1 - \left(1 - \frac{1-q^*}{q^*}\right)^2\right].
\]

It is easy to show that this term is zero when \(\lambda\) is 0.5 (completely uninformative evidence), and it increases initially and then decreases.

REFERENCES


