Literature review

How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives

Valerie F. Reyna*, Jonathan C. Corbin, Rebecca B. Weldon, Charles J. Brainerd

Cornell University, Human Neuroscience Institute, G331, Martha Van Rensselaer Hall, Ithaca, NY 14853, United States

ABSTRACT

Fuzzy-trace theory posits independent verbatim and gist memory processes, a distinction that has implications for such applied topics as eyewitness testimony. This distinction between precise, literal verbatim memory and meaning-based, intuitive gist accounts for memory paradoxes including dissociations between true and false memory, false memories outlasting true memories, and developmental increases in false memory. We provide an overview of fuzzy-trace theory, and, using mathematical modeling, also present results demonstrating verbatim and gist memory in true and false recognition of narrative sentences and inferences. Results supported fuzzy-trace theory's dual-process view of memory: verbatim memory was relied on to reject meaning-consistent, but unpresented, sentences (via recollection rejection). However, verbatim memory was often not retrieved, and gist memory supported acceptance of these sentences (via similarity judgment and phantom recollection). Thus, mathematical models of words can be extended to explain memory for complex stimuli, such as narratives, the kind of memory interrogated in law.

© 2016 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.

Contents

This article provides a review of fuzzy-trace theory's (FTT) theoretical framework that unifies and refines constructivist and information-processing approaches by relying on a distinction between literal, precise memory representations (verbatim traces), and vague, meaning-based memory representations (gist traces), both of which improve with development (Brainerd & Reyna, 2005; Reyna, Wilhelms, McCormick, & Weldon, 2015). In this review, we discuss how verbatim and gist memory accounts for dissociations between true and false memory, distinguishing spontaneous false memory (e.g., people misremember because they make inferences) from misinformation effects (e.g., people misremember because interrogators introduce false information). We then explain how seemingly paradoxical findings in false memory that are not easily accounted for by associationist theories of memory (e.g., Gallo & Roediger, 2003), such as the greater persistence of false as compared to true memory, are predicted by fuzzy-trace theory.

We argue that artificial distinctions between studies of true memory as opposed to false memory are no longer tenable. Indeed, accounts of memory that exclude consideration of meaning-related "false" memories are not only incomplete, they are misleading. To

* Corresponding author. Tel.: +1 607 319 0655.
  E-mail address: vr53@cornell.edu (V.F. Reyna).

illustrate, true recall can be based on readout from verbatim memory or reconstruction from gist memory (followed by a familiarity judgment). For example, given the following study list – bed, rest, dream, snooze, blanket, doze, slumber, snore, nap, peace, yawn, and drowsy – people may correctly recognize "snooze" on a subsequent recognition task because they retrieve a verbatim trace of the presented word, or they could have inferred its presence based on the overall gist of the list as one containing sleep-related terms. Thus, even correct recall (e.g., of "snooze") is not a pure measure of a single memory process, but, rather, has been shown to be produced by dual verbatim and gist processes (e.g., Brainerd, Reyna, & Howe, 2009). Therefore, directly comparing true and false memory – as is done routinely in behavioral and functional magnetic resonance imaging (fMRI) studies – does not necessarily isolate memory processes. People can testify accurately about their memory for events either because they retrieve verbatim memories or because they reconstruct them from gist. The accuracy of the memory report does not unambiguously prove its psychological origin.

Knowing the conditions under which memories were encoded and tested sometimes makes it possible to infer their psychological origin (Reyna, Mills, Estrada, & Brainerd, 2007). In this article, we report the results of an experiment about memory for narratives that illustrates many of these conditions and their effects on memory. Narrative memory, as opposed to memory for words, more closely resembles memory for witnessed events. One such condition is the delay between the witnessed event and testimony about the event. Eyewitness reports that occur immediately after a witnessed crime are more likely to be based on verbatim memory than reports given later in court and, thus, are more likely to be accurate about arbitrary details. However, a legal case often does not turn on arbitrary details. Later testimony in court that requires an inference, for example about whether the accused was acting strangely or was angry on the day of the crime, draws on memory for the gist of events, and such gist generally endures over long delays. The gist of an event is not random and, based on knowledge about the witness (e.g., a rival gang member) and circumstances (e.g., the accused was standing over the decedent and, hence, is remembered as holding the gun), likely hypotheses can be generated about the type of information the witness would be likely to infer. These hypotheses can then be tested with other evidence.

Therefore, there are two kinds of accuracy that are relevant in the real world and that must be distinguished: literal verbatim accuracy and substantive accuracy about the gist of events. A witness can be right about the gist of an event (e.g., that they were robbed or molested), but wrong about details (the color of a shirt or a car), and vice versa. We use the term "true" memory in its traditional sense in the literature to refer to memory for actual events as they were experienced (i.e., verbatim memory), as contrasted with "false" memory which can be substantively accurate or not. The law assumes that a witness's memory for actual events can be absent, weak, or mistaken, but, when witnesses claim to remember, they are supposed to testify only to what they have experienced. The law fails to accommodate the fact that people routinely "remember" a gist of events that they never experienced.

The accuracy of memory reports not only depends on the quality of verbatim and gist memories, but also on how questions about memory are asked. For example, memory reports are influenced by recognition probes: Asking witnesses about an event that really happened can trigger verbatim memories, but asking them about a true inference (e.g., an event they know to be true but could not have been directly observed) can elicit memories for the gist of the event. Fortunately, these ambiguities in memory reports have been addressed in research that we now discuss, which helps to clarify otherwise mysterious memory phenomena, including dissociations between true and false memory, false memories outlasting true memories, and developmental increases in false memory.

1. Dual processes in memory

1.1. What is false memory?

Studies of autobiographical and everyday episodic memories have revealed that a surprising number of real-life "memories" are false in the sense that they were never experienced (Frenda, Nichols, & Loftus, 2011). Instead, memories can be suggested (e.g., by viewing family photographs), imported from one context in which they did occur to another in which they did not occur (source misattribution e.g., when events from a movie are remembered as personally experienced), or, most commonly, they can represent semantic and inferential extensions of experienced events (e.g., seeing the accused carrying a shirt into a laundry room and hearing a washing machine turn on in that room soon afterwards can be misremembered as having seen the accused put the shirt into the washer; Brainerd, Reyna, Holliday, & Nakamura, 2012; Ceci & Bruck, 1993; Dodson & Shimamura, 2000). Laboratory research has focused mainly on two paradigms: misinformation in which never-experienced events are *suggested* (e.g., through suggestive questions or direct misinformation) and *spontaneous* false memories that occur as a result of ordinary processes of understanding and inference, namely, gist extraction (for a more complete taxonomy of types of false memory, see Table 17.1 in Reyna et al., 2007).

A popular false-memory paradigm is the Deese–Roediger–McDermott (DRM) task in which presented lists consist of words that share numerous semantic relations with one another (e.g., bed, rest, dream, snooze…), relations that go beyond mere word association or contiguity frequency (Brainerd, Yang, Reyna, Howe, & Mills, 2008; Robinson & Roediger, 1997). In recognition tests, presented words and unpresented but semantically related distractors (e.g., sleep) are probed, as well as unrelated distractors that do not share meaning with presented words (e.g., apple). This paradigm typically produces high levels of recognition for the unpresented distractors that are semantically related to multiple words on the list, often rivaling the acceptance rate for presented words (Reyna & Lloyd, 1997).

We concentrate on spontaneous false memory for semantically related words and sentences because these paradigms most clearly reflect gist-based processes in memory falsification. Recent studies explain why this is so; such studies have failed to find correlations between misinformation effects (when subsequently presented misinformation is reported as having occurred as part of an original event) and false memories for semantically related word lists, referred to as DRM effects (e.g., Ost et al., 2013). FTT predicts that both misinformation and DRM effects are products of verbatim and gist memories (Barnhardt, Choi, Gerkens, & Smith, 2006; Reyna & Titcomb, 1997; see Brainerd, Reyna, Wright, & Mojardin, 2003, and Mojardín, 1998, for parameter estimates supporting this prediction). However, verbatim memory (for the misinformation) promotes false-memory acceptance in the misinformation paradigm, whereas it promotes false-memory rejection in the DRM paradigm. Because misinformation is typically gist-consistent in experiments, gist memory promotes false-memory acceptance in both paradigms.

To illustrate these relationships, consider the following example: Imagine that an interrogator inadvertently introduces misinformation that a perpetrator had a hammer (when he actually had a screwdriver). The *better* that witnesses remember the presented word "hammer" from the interrogation, the more likely they are to later *accept* the false statement that the perpetrator had a hammer when questioned in the courtroom (positive correlation). (Misinformation effects occur when people are more likely to remember the manipulated information presented in the interrogation than the information from the original event.)