Exemplars as a least-committed alternative to dual-representations in learning and memory

Inauguraldissertation
zur Erlangung des
Doktorgrades
der Humanwissenschaftlichen Fakultät
der Universität zu Köln
nach der Promotionsordnung 10.5.2010
vorgelegt von

Frederik Aust
aus
Quimper, Frankreich

August, 2020

Abstract

Despite some notable counterexamples, the theoretical and empirical exchange between the fields of learning and memory is limited. In an attempt to promote further theoretical exchange, I explored how learning and memory may be conceptualized as distinct algorithms that operate on a the same representations of past experiences. I review representational and process assumptions in learning and memory, by the example of evaluative conditioning and false recognition, and identified important similarities in the theoretical debates. Based on my review, I identify global matching memory models and their exemplar representation as a promising candidate for a common representational substrate that satisfies the principle of least commitment. I then present two cases in which exemplar-based global matching models, which take characteristics of the stimulus material and context into account, suggest parsimonious explanations for empirical dissociations in evaluative conditioning and false recognition in long-term memory. These explanations suggest reinterpretations of findings that are commonly taken as evidence for dual-representation models. Finally, I report the same approach provides also provides a natural unitary account of false recognition in short-term memory, a finding which challenges the assumption that short-term memory is insulated from long-term memory. Taken together, this work illustrates the broad explanatory scope and the integrative and yet parsimonious potential of exemplar-based global matching models.

This dissertation was accepted by the Faculty of Human Sciences at the University of Cologne in January 2021.

Table of contents

Introduction

Chapter 1

Learning is not memory

FREDERIK AUST

> *learning and memory researchers, collectively, seem to be conducting a gigantic, double-blind "conceptual replication." The dominant positions on several central issues in [research on memory] have returned to points that would have given the neobehaviorists and functionalists well-founded feelings of déjà vu. (pp. 360-361, Hintzman, 1993)*

Everyday language suggests that *learning* and *memory* can be used interchangeably to refer to the psychological study of how experiences are retained and shape our behavior. For example, students similarly complain about what they are required to *learn* or *memorize*. This misconception caused me quite some confusion during the first sitting of my introductory course entitled "Learning". The course would not, as I had expected, cover how to effectively prepare for my exams, that is how to memorize effectively. Rather the course would primarily cover animal research on conditioning. I remember being annoyed about the misleading course title and jargon. Ironically, I later found out that James Mazur, the author of the recommended course book, was similarly confused when he signed up for his first courses on learning (p. 1, Mazur, 2012). On the one hand, this anecdote may be taken as an illustration of how jargon can obscure conceptual distinctions, especially when it repurposes everyday language expressions. At a deeper level this anecdote also illustrates that how counterintuitive and debatable the partition of psychological research into learning and memory is. Despite broad substantive overlap and parallels in theoretical debates the fields differ in experimental methodology and theorizing.

The common definitions of learning and memory reveal some of these differences. In psychological science the term memory may refer to one of three concepts (Spear & Riccio, 1994): (1) the mental representation of a particular past experience, also refereed to as memory trace, (2) the structure that stores the memory traces, and (3) the process by which memory traces are formed, stored, and retrieved. As these definitions illustrate, memory research aims to investigate unobservable processes and structures that serve to preserve information over time. In this sense, memory mediates the effect of past experiences on behavior. In stark contrast, definitions of learning are first and foremost concerned with observable events. Consider the following common textbook definition[1]:

> *Learning refers to a relatively permanent change in behavior as a result of practice or experience (Lachman, 1997)*

This definition demands an observable change in behavior for learning to have occurred. It makes no reference to unobservable mental effects of past experiences. Hence, research on learning requires no assumptions about or interest in processes that mediate effects of past experiences on behavior (De Houwer et al., 2013).

Although there is some theoretical convergence of research on learning and memory, the illustrated disconnect largely persists. I will argue that both fields have much to gain from theoretical exchange and suggest a broad theoretical perspective to facilitate integrative theory building. Before addressing the mutual benefits of theoretical exchange between the fields of learning and memory, it is useful to understand how these fields diverged despite their common research goal. In the following I will, therefore, briefly sketch the historical origins and epistemological differences between research on learning and memory.

[1] Lachman (1997) criticizes this definition and proposes a process-focused alternative. However, an improved functional definition has recently been proposed that avoids referring to unobservable mental processes (De Houwer et al., 2013). I present the less than perfect textbook definition here because it is simple, common, and serves to illustrate the traditional focus on observables in learning research.

1.1 Historic origins

The divergence between the fields of learning and memory has historical reasons that can be traced back to the dominant lines of thinking at the time they emerged. Learning was first intensely studied when behaviorism dominated experimental psychology starting in the 1910s. Research on memory only gained traction when behaviorism was challenged during the cognitive revolution starting in the 1950s.

1.1.1 Behaviorist origins of learning research

> *behaviorism is the only road leading to science. [...] behaviorism must be looked upon as the rough scientific clay which all must shape (p. vii, Watson, 1924)*

The advent of behavioristic psychology—behaviorism—in the early 20th century steered the focus of psychological research away from introspective methods and towards behavioral observations (Watson, 1924). The central tenet of behaviorism was that the scientific method demands intersubjective verifiability. Thus, Watson argued that psychological science should analyze observable events—stimuli and responses—rather than unobservable events accessible only by introspection. The influential learning researcher Burrhus Skinner made an even stronger antimentalist assertion: Skinner argued that, in addition to being unserviceable as data, unobservable events or latent states have no merit in predicting behavior and unnecessarily complicate psychological theory. By the virtue of parsimony, psychological theory should, therefore, explain behavior as a function of tangible factors (Skinner, 1950; Skinner, 1985). Anything in between stimulus and response was to be treated as a black box. Skinner's position was extreme even among behaviorists (e.g., Miller, 1959), but it serves to highlight that important research on learning was conducted without particular concern for unobservable entities such as memory.

The behaviorist preoccupation with observable stimuli and response regularities also shaped the experimental methodology. The stated goal was to identify general principles of learning that apply broadly across learning situations and species (e.g., Skinner, 1938). To this end, behaviorists relied heavily on animal research. The use of animals permitted greater control over the experimental environment, subject, and experimenter effects (but see Rosenthal & Fode, 1963). It was further assumed that animal research facilitated discovery of general principles due to reduced complexity of the organism compared to humans.

Consequently, much of behaviorist learning research focused on basic learning procedures (Jenkins, 1979). In classical conditioning procedures, conditioned stimuli (CS) were paired with unconditioned stimuli (US) that elicited inherent responses (referred to as unconditioned responses, UR). Following repeated pairings the CS started to elicit a similar response as the US, which was refereed to as the conditioned response (CR; for a review see Bouton, 2007). The classic example is the procedure used in the seminal studies by Pavlov (1927/1960) in which he sounded a bell (CS) when feeding (US) dogs, which caused the dogs to salivate (CR) in response to the bell sound even when no food was present. In operant conditioning procedures, voluntary behavior is shaped by reinforcement, that is, the positive or negative outcome of an action (for reviews see Bouton, 2007; Pierce & Cheney, 2013).

However, in addition to these famous learning phenomena, behaviorists also studied verbal learning in humans, the direct precursor to subsequent memory research (Tulving & Madigan, 1970). Behaviorists interested in verbal learning studied how verbal information is retained and how it affects behavior. They built on the experimental methodology of rote learning spearheaded by Ebbinghaus (1885) and largely understood rote learning to be a form of stimulus-response (S-R) learning (Bower, 2000). For example, forgetting was understood to be a form of extinction or renewal of preexperimental associations. Paying tribute to the behaviorist tradition, verbal learning researchers employed experimental paradigms with clearly defined stimuli and associated responses. In serial learning paradigms, participants learned lists of stimuli in order to later freely reproduce them in the correct order. It was assumed each list element acted as stimulus and prompted as response the production of the subsequent stimulus. In the paired-associate paradigm participants studied lists of stimulus pairs and were later cued with one element of each pair and asked to reproduce the other. Especially the paired-associate paradigm bears obvious resemblance to the well-known conditioning paradigms.

Much of behaviorist learning research assessed the effects of, for example, stimulus characteristics, stimulus-response contingencies, and time on the frequency of responses (Skinner, 1950). The study of animals precluded assessment of introspective reports and metacognitive judgments, but this was no loss to behaviorists. And the behaviorist approach proved to be very successful. Behaviorist learning researchers discovered fundamental relations between environmental factors and behavior that guide the development of interventions to modify behavior to this day.

1.1.2 Cognitivist revolution and memory research

> *Behaviorism [...] must accommodate itself to accepting the importance of what goes on inside the "black box", especially since we now have methods for investigating its contents (p. 432, Shevrin & Dickman, 1980).*

Notwithstanding its success the primacy of behaviorism toppled in the so-called cognitive revolution. The work by Edward Tolman is often considered an early precursor of cognitivist psychology (Tolman & Honzik, 1930; Tolman, 1948; Tolman et al., 1946). For example, Tolman et al. (1946) repeatedly placed a food reward at a designated finish location in a simple maze until the rats learned the route. During a subsequent test run the previously learned route to the food reward was blocked and the rats were forced to choose one of 18 new alleys. In contrast to the dominant view that learning creates associations between stimuli (S-S learning) or S-R learning, most rats chose the alley that provided the shortest route to the food reward—no relearning was necessary. Tolman concluded that the rats had formed cognitive maps of the maze in which particular locations were associated with rewards. Hence, he was convinced that his data could not be explained by simple operant conditioning and necessitated consideration of an unobservable spatial memory.

The advent of cognitivist psychology changed researchers' view on human experience and behavior as well as the questions they sought to answer. In response to the seminal work by Miller (1956) and Neisser (1967) among others, researchers started to think about mental processes in terms of information theory (Garner, 1962)—a view still commonplace today. Using computers as metaphor, humans were construed as capacity-limited information processors that analyze and encode information into neural activity which is processed by mental programs, stored, and later retrieved to guide behavior.

Although the conceptual division of learning into acquisition and retrieval was prevalent among behaviorists, they devoted little attention to the act of retrieval:

> *recall was observable behavior whose measurable aspects simply served to provide evidence about strength of associations. Moreover, the act of recall was empirically neutral in that it did not affect the state of the system; it was theoretically uninteresting because it could not be studied independently of acquisition. (p. 352, Tulving & Thomson, 1973)*

In contrast, cognitivists considered information processing to be constructive at all stages—encoding, processing, and retrieval (Neisser, 1967). This view prompted new questions about how the information stored in memory is organized.

In an influential verbal learning experiment, Tulving (1962) investigated the mental organization of studied material that in itself exhibits no obvious learnable structure. Tulving repeatedly presented a set of semantically unrelated words such that across presentations each word was adjacent to each other word equally often. After each presentation, participants were asked to freely recall all words in *any order*- From a verbal learning perspective, it seemed unlikely that performance would improve across repeated presentations due to the absence of repeated S-R contingencies. Tulving, however, found that, across repeated presentations, participants' recalled the words in very similar sequences. And the more consistent their recall sequences the more words participants recalled. Participant actively structured the study material beyond the sequential order, suggesting that encoding entails transformations and interpretations of presented stimuli.

These free recall findings seemed difficult to explain in behaviorist terms. It is not obvious what stimuli participants responded to when producing the consistent recall sequences. Moreover, participants appeared associate stimuli with unobserved responses. Tulving (1962) observed similar consistent recall sequences across subjects, which suggested that participants found structure in the superficially unstructured material. Besides demonstrating the importance of unobserved processes during learning, this work called into question the prevailing view that verbal learning was a form of simple S-R learning. To further probe the structure of memory and the organization of knowledge, researchers started to employ different experimental methods. Free recall paradigms quickly gained popularity to the disadvantage of traditional behaviorist paradigms:

> *In just five years, between 1967 and 1972, the ratio of paired-associative studies to free-recall studies in the index of the Journal of Verbal Learning and Verbal Behavior dropped from 31:9 to 2:32. (p. 370, Hintzman, 1993)*

For example, in a study, which attracted wider interest only three decades later, Deese (1959) investigated the cause of intrusion errors in free recall—the seemingly random erroneous recall of words that had not been on the study list. Revealing important organizational principles of memory, Deese found that intrusions could be predicted from the frequency with which the intruding words were freely produced as associates of the words on the study list. This study was one of the first to investigate false memories, a field of study that later attracted widespread attention and has served to gauge the structure of memory.

Recognition testing was another popularized dependent measure that remains influential until today. In its simplest form, participants are asked to judge whether a stimulus was presented as part of the study list (*old*) or not (*new*). Accordingly, this task is commonly referred to as old-new recognition testing. Compared to recall measures, recognition testing is more sensitive to weak memories and more economic than the even more sensitive relearning method introduced by Ebbinghaus (1885) and Groninger and Groninger (1980).

An important study on recognition of words was reported by Underwood (1965). In addition to the customary old and new words, the list of memory probes encompassed new words that were related to previously studied words. These *lures* were, for example, antonyms (*day* and *night*), converging associates (*bread* and *butter*), or exemplars to a superordinates (*oak* and *tree*). In line with the free recall findings by Deese (1959), Underwood (1965) observed false recognition—increased rates of old-responses to lures compared to unrelated new words. These findings swayed Underwood (1969), a influential memory researcher and S-R learning theorist, towards cognitivist encoding principles.

Another aspect of the computer-inspired information processor metaphor, limited processing capacity, sparked interest in short-term memory (Atkinson & Shiffrin, 1968; Bower, 2000). In his famous paper, Miller (1956) proposed that human short-term memory was limited in capacity to approximately seven informational units, which he called *chunks*. Moreover, Peterson and Peterson (1959) showed that even three-letter consonant syllables were difficult to retain for more than a few seconds while participants counted to keep them from rehearsing the to-be-remembered material. Causes for the fragility of short-term memory remain a central focus of short-term memory research to this day (Cowan, 2001; Cowan et al., 2012; Ma et al., 2014). It is seen as one of the key characteristics that sets short-term memory apart from long-term memory (Cowan, 2008). I will later return to this discussion in [Short-term memory].

Cognitivist ideas sparked new lines of research that adopted a different vocabulary, metaphors, and experimental methodology. These shifts are exemplified by the fact that the *Journal of Verbal Learning and Verbal Behavior* was continued as *Journal of Memory and Language* in 1985. Yet, despite its name, the cognitive revolution lacks properties of a scientific revolution as commonly defined in the philosophy of science. The behaviorist paradigm saw

> *no falsification, no drowning in a sea of anomalies, no ad hoc strategies to save a degenerating research paradigm, and no inferior empirical and conceptual problem-solving capacity (p. 105, O'Donohue et al., 2003).*

Rather, it reflects a widespread change in research trends and interests (also see Hintzman, 1993; Roediger, 2004), which lead to a partitioning of the field. As outlined above, research into memory became a new field that fully embraced cognitivism and probed, among other things, the structure of memory and knowledge representation. Learning researchers continued to address questions about conditioning and the effects of predictive stimulus relations on behavior. Behaviorist (or functional) learning research is being continued (Roediger, 2004; for a brief primer on current developments see Stewart, 2016), but over time the majority of learning theorists have adopted a more cognitivist perspective (e.g., p. 106, Holland, 1990). Nonetheless the partitioning into learning and memory persists and has seen only scant theoretical exchange (Hintzman, 1993; Tulving & Madigan, 1970).

1.2 Epistemic differences

The diverging pretheoretical convictions, interests, and methods of traditional learning and memory research also entail epistemological differences. That is, learning and memory research sought explanations at different levels of analysis. It has been argued that these levels of analysis are linked only loosely and are best addressed independently. By extension, theoretical exchange seems unproductive. In the following I will briefly review the concept of levels of analysis and why it provides no compelling argument for independent theorizing. Finally, I will discuss different approaches to bridge levels of analysis.

1.2.1 Marr's levels of analysis

Marr (1982) proposed that complex information-processing systems may be explained at different, loosely related, levels. Drawing on the cognitivist information-processor metaphor, he distinguished between the *computational*, *representational and algorithmic*, or *implementation* level[2](Figure 1.1; see Skinner, 1950, for a similar distinction). Computational level analyses aim at providing a precise, ideally formal, description of the system's input-output mapping. Crucially, such a description comprises what is computed and why the computation is appropriate (e.g., in the environmental context; Bechtel & Shagrir, 2015; Shagrir & Bechtel, 2018). Representational and algorithmic level analyses are concerned with how a system represents input (and output) and what algorithms realize the computation to transform one into the other. Analyses at the implementation level seek to explain how the system is implemented in a physical medium, such as neurons.

To illustrate his distinction, Marr (1982) uses the example of a cash register (p. 20ff.). At the computational level, the device performs addition of positive and negative numbers. Answering *why* a cash register adds product prices rather than multiplying them, requires an analysis of the task as well as the environment and its affordances. Marr lists requirements that

[2]Marr (1982) was neither the first (e.g., Dennett, 1981; Glass et al., 1979; Newell, 1982) nor the last (e.g., Anderson, 1990; Pylyshyn, 1984) to make this or similar distinctions. van der Helm (2012) even likens Marr's to Aristotle's distinction between goal, method, and means. In the following I'll focus on Marr's formulation because it is the best known in cognitive science.

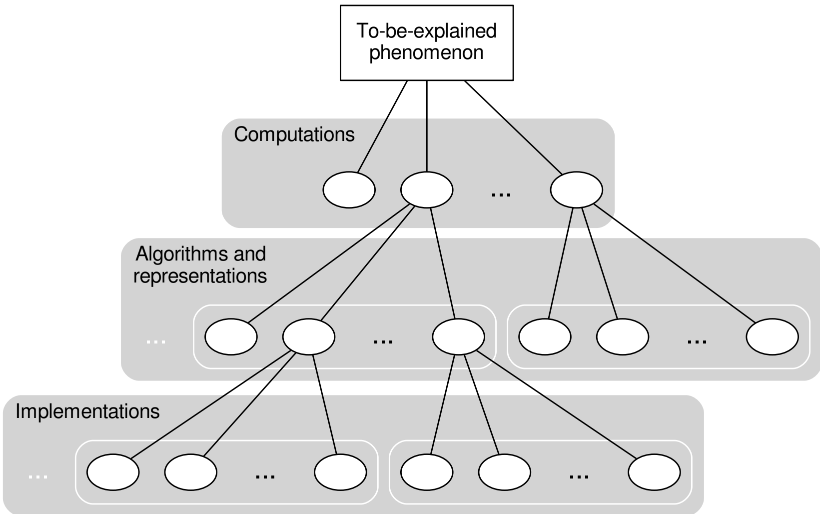

Figure 1.1: Graphical representation of to top-down one-to-many hierarchy of explanations according to Marr's levels of analysis. Round nodes represent explanations and edges indicate the nesting between explanations at different levels.

we intuitively feel must be met when determining a final bill and derives constraints that precisely define the computation:

> 1. *If you buy nothing, it should cost you nothing [. . .].*
> 2. *The order in which goods are presented to the cashier should not affect the total. [. . .]*
> 3. *Arranging the goods into two piles and paying for each pile separately should not affect the total amount you pay. [. . .]*
> 4. *If you "buy" an item and then return it for a refund, your total expenditure should be zero. (pp. 22-23, Marr, 1982)*

At the representational and algorithmic level, product prices and final bill are represented, for example, in the Arabic numeral system (rather than Roman or binary) and addition is carried out by starting with the least significant digit and "carrying" if the sum exceeds 9 (p. 23, Marr, 1982). Marr notes that possible algorithms may be heavily constrained by the representation and for any representation there may be multiple possible algorithms that differ with respect to, for example, efficiency or robustness. The implementation level may put additional constraints on the algorithm. For example, a classic computer architecture may necessitate an algorithm that operates serially rather than in parallel. But again, different physical implementations need to be considered. Besides elucidating the differences between explanations at each level, the example highlights levels mutually constrain each other.

Past debates show that the exact level of analysis of an explanation is debatable (pp. 194-195, McClamrock, 1991). It is important to understand, however, that Marr's levels of analysis constitute no postulate about the actual number of levels of organization of any complex information-processing system. They are adopted relative to a to-be-explained phenomenon. Hence, Marr's levels of analysis are probably best understood as an idealization of a multiply nested structure of analysis levels (pp. 194-195, McClamrock, 1991) or a continuous degree of abstraction.

In relative terms, it seems clear that behaviorist learning research, and cognitivist research on learning and memory gravitate towards different levels of analysis. The behaviorist focus on observables orients traditional learning research toward the computational level, whereas memory research that probes the structure of memory and semantic knowledge is focused on the representational and algorithmic level. Cognitivist explanations of learning have trended more towards the implementation level by invoking (abstractions of) neurons and their associations as theoretical building blocks. The affinity towards implementational explanations may be owed to the legacy of influential behaviorist thinkers, who notoriously denied the utility of a cognitive level of analysis:

> *In a behavioral account the whole organism responds, and it responds to the world around it—for reasons which neurology, not cognitive science, will eventually discover (p. 293). [. . .] Cognitive science is often only premature neurology (p. 300, Skinner, 1985).*

1.2.2 Across-level comparisons

Marr (1982) posits that the levels of analysis are linked only loosely because each level requires many considerations that are not informed by explanations at the other levels. This claim has often been cited to argue that theorizing should proceed independently at each level (and sparked debates about the "right" level of analysis; Eliasmith & Kolbeck, 2015; Love, 2015; McClamrock, 1991). With this in mind, the disconnect between learning and memory research may seem unproblematic, sensible even. This strict independence interpretation of Marr (1982), however, forgoes one of the strengths of the approach (Eliasmith & Kolbeck, 2015). As previously noted, Marr envisioned one-to-many correspondences when moving from higher to lower levels of analysis, Figure 1.1. Any explanation at the computational level is consistent with multiple competing representational and algorithmic explanations. But constraints also arise from the bottom up (Bechtel & Shagrir, 2015). The cash register example clearly demonstrates constraints exerted across levels ("of course they are logically and causally related" p. 25, Marr, 1982). It follows that complete understanding of a phenomenon requires adequate explanations at each level.

Rather than theorizing independently, pragmatic epistemological pluralism may be more productive (van der Helm, 2012): Entertaining tentative explanations at each level enables across-level comparisons that may inspire revisions or refinements. Initially, independent theorizing at each level fosters diversity in thought and stimulates new research. After an explanation at any given level survives initial falsification attempts and is tentatively retained, across-level comparisons should follow. In addition to inspiring revisions, across-level comparisons can clarify the relationship between explanations in the one-to-many hierarchy. Explanations that appear to contradict each other may, upon closer inspection, turn out to be closely related, or vice versa. Hence, Marr's distinction between levels of analysis provides no compelling argument for the theoretical separation of learning and memory research. After decades of scant theoretical exchange, tentative explanations have been proposed in both fields—across-level comparisons of learning and memory explanations appear timely and topical.

Which of Marr's levels of analysis to address first and how to approach across-level comparisons is subject to debate. Commonly discussed are top-down (e.g., Griffiths et al., 2010; Marr, 1982) and bottom-up approaches (e.g., Bower & Beeman, 1998; McClelland et al., 2010); more recently an inside-out approach has been suggested (Love, 2015).

Marr (1982) stressed the importance of the computational level and suggested a top-down approach to match the assumed one-to-many hierarchy of explanations. His approach was motivated by the belief that it is not feasible to recover the computed mathematical function from the detailed neural implementation (cf. p. 200, Shagrir & Bechtel, 2018). Marr's suggested focus on a precise formal characterization of phenomena at the computational level and top-down approach bears a remarkable resemblance to Skinner's thinking: Skinner (1950) advocated for formal characterization of the relationship between manipulated variables and performance and remained convinced (1985) that such precise characterizations formed the basis for investigations into the neurological underpinnings of learning. Among contemporary advocates of the top-down approach sensu Marr are cognitive scientists working on a Bayesian probabilistic framework of cognition (Griffiths et al., 2010), functional psychologists (e.g., Fiedler, 2016; Houwer, 2011; Hughes et al., 2016), and some philosophers of science (Shagrir & Bechtel, 2018).

Critics of the top-down approach contend that the implementation level is practically devalued as mere implementation in a physical medium (e.g., Smolensky, 1988). Consequently, the brain is mostly ignored and constraint that arises from the neurological substrate is neglected (e.g., Love, 2015; McClelland et al., 2010). Proponents of the bottom-up approach, such as connectionist cognitive scientists (e.g., McClelland et al., 2010; Rogers &

McClelland, 2014) or computational neurobiologists (e.g., Bower & Beeman, 1998), prefer to let algorithms, representations, and computations emerge from sub-cognitive processes. Through emergence, the bottom-up approach naturally adheres to constraints of the neurological substrate and avoids incorrect assumptions about, for example, representational structures. An important challenge for the bottom-up approach is that it assumes a detailed understanding and implementation of the basic neural processes—incorrect or missing details yield incorrect emergent properties. Moreover, explanations at the implementation level quickly become complicated and offer little direct insight into cognitive phenomena beyond "the outcome was caused by changes to neural connections" (French & Thomas, 2015). The latter critique originally prompted Marr's (1982) distinction between levels of analysis, as reflected in the famous quote:

> *trying to understand perception by studying only neurons is like trying to understand bird flight by only studying feathers: It just cannot be done (p. 27, Marr, 1982).*

Recently, Love (2015) advocated for an inside-out approach—starting across-level comparisons from the middle algorithmic and representation level—to combine the advantages of the top-down and bottom-up approaches. The idea is that the algorithmic and representation level simultaneously grants theoretical proximity to the computational and implementation level and is thus most readily constrained. In downward-directed comparisons, this proximity facilitates incorporating constraints arising from the neurological substrate and select competing theories based on the predicted neural activity (e.g., Mack et al., 2013). At the same time, the algorithmic and representation level may produce more direct and satisfying explanations of cognitive phenomena than the implementation level. In upward-directed comparisons, the proximity facilitates comparisons with functional relationships between the environment and behavior or, for example, normative rational accounts (e.g., Griffiths et al., 2010), compared to the bottom-up approach.

It seems unlikely that any one approach to across-level comparisons is generally superior—a cynical observer might comment that the suggested approaches are rationalizations of their proponents' preferred level of analysis. But given that the levels of analysis are mutually constraining, any attempt to incorporate constraints across levels can be productive (Bechtel & Shagrir, 2015; McClamrock, 1991; van der Helm, 2012).

When attempting to connect largely disconnected fields of research, such as learning and memory, an argument can be made for the inside-out approach: The algorithmic and representation level lends itself to relate explanations of distant phenomena and explore commonalities. Considering that computational-level explanations serve to delineate to-be-explained phenomena (Shagrir & Bechtel, 2018), the precise mappings from input to output, by necessity, have a narrow scope.[3] For example, a computational level explanation of classical conditioning maps stimulus pairings to conditioned responding. Such an explanation purposely does not apply to associative memory tasks where stimulus pairings need to be mapped to recall or recognition and confidence judgments. At the other end of the spectrum, explanations at the implementation level, due to their complexity may complicate the discovery of commonalities. Thus, the algorithmic and representational level strikes a balance between abstraction and generality that is well suited to develop domain-general explanations.

More specifically with respect to the distinction between learning and memory, an obvious question is whether they may be conceptualized as distinct algorithms that operate on the same representation (cf. p. 707, Nosofsky, 1988). As noted previously, learning and memory algorithms necessarily differ to some extent. The assumed representation of stimuli and pairings, however, is potentially domain-general. Discovery of representational formats that are applicable to a broad range of phenomena would be theoretically interesting because they serve as a basis for broad-scoped, well-constrained, parsimonious explanations. According to Marr's principle of least commitment, a domain-general representation likely stores stimulus information after minimal preprocessing because unprocessed representations are conducive to the flexible deployment of different algorithms to meet the demands of different tasks (pp. 485-486, Marr, 1976). Thus, a promising broadly unifying theoretical approach to learning and memory would be to identify a common representational

[3] Love (2015) contends that explanations at the computational level are not sufficiently constrained. His analysis targets probabilistic models of cognition and their a priori assumption of optimality or rationality in particular. However, as Shagrir and Bechtel (2018) point out computational level analysis do not necessarily presuppose optimal procedures. They only require that the procedures, optimal or not, are grounded in the environment that the information-processing system operates in (p. 203).

format and defer distinct processing steps to the response stage. For example, the stimuli presented in classical conditioning and associative memory paradigms may result in largely the same representations but the way in which this information affects the observed behavior observed in these tasks may differ. The perspective that follows from this simple idea may facilitate integrative theory building and is a recurrent theme in the work presented here. In the next chapter, I will review some examples of theories that moved towards or fully adopted this perspective.

1.3 Nascent convergence

After a long period of scant theoretical exchange, theorizing in learning and memory seem to be reapproaching one another. As previously noted, a majority of learning researchers have become interested in cognitive representational and algorithmic level analysis:

> *Most students of animal conditioning today agree that the more profitable subjects of inquiry are the mental events, structures, and processes that underlie conditioned behavior, and not the conditioned behavior itself. (p. 106; Holland, 1990)*

This shift has resulted in debates about the representation of learned predictive relations (e.g., Hanus, 2016; Mitchell et al., 2009). As the structure of knowledge has been the object of intense study in memory research, some learning theorists have looked to models of episodic memory to develop representational and algorithmic explanations of learning.

In particular, the broader class of global matching memory models, also known as exemplar or instance models, has been the starting point for much theorizing. Variants of global matching models have been successfully applied to a range of learning phenomena, such as associative learning (Jamieson et al., 2012), artificial grammar learning (Jamieson & Mewhort, 2009a; Jamieson & Mewhort, 2010; Jamieson & Hauri, 2012), and serial reaction-time task (Jamieson & Mewhort, 2009b). Moreover, a related approach has been applied to a diverse set of effects from the contingency learning literature (Schmidt et al., 2016). All these proposals aim to develop a unifying theoretical account of learning and memory phenomena and propose that learning phenomena can be understood in terms of encoding and retrieval of episodic memory traces.

Conversely, memory theorists have, for example, taken inspiration from the concept of prediction error, which is central to many learning theories. At least since the influential Rescorla-Wagner model of classical conditioning (Miller et al., 1995; Rescorla & Wagner, 1972; Siegel & Allan, 1996), the idea that performance in learning paradigms is driven by the discrepancy between predicted and observed events has been very influential in theories of learning and far beyond. Early on, the Rescorla-Wagner model was considered as a model of performance in paired-associate tasks but was abandoned when memory researchers focused on other tasks (Siegel & Allan, 1996). More recently, the concept of prediction errors has been adopted to explain how the continuous stream of events we experience is segmented into separable episodes which are then stored in episodic long-term memory (Zacks et al., 2007). In short, an episode is conceptualized as a predictable set of events and that unexpected events signal the beginning of a new episode.

Theorists gravitating more towards implementation level analyses have also attempted to provide a unified perspective on learning and memory. Connectionist or parallel distributed processing models rely heavily on the insight that prediction error can guide learning of association strengths between neuron-like representational units (Rumelhart et al., 1987). Some of the first applications of these models were in the domains of memory, categorization, and grammar learning (McClelland et al., 1987). Since then the basic ideas of connectionist models have been developed into the influential Complementary Learning Systems framework (CLS; McClelland et al., 1995; O'Reilly et al., 2014) and elaborated into neurophysiologically-underpinned models of learning (e.g., O'Reilly & Rudy, 2001; Hebscher et al., 2019) and episodic memory (Norman & O'Reilly, 2003; Hebscher et al., 2019; Schapiro et al., 2017). Despite these impressive efforts, theoretical exchange across the domains of learning and memory is scant.

The relationship between traditional research into learning and memory was aptly summarized by Tulving and Madigan (1970):

> *the two subcultures share a common goal, but they talk different languages, ask different questions, use different methods, and have sworn allegiance to different pretheoretical assumptions (p. 439, Tulving &*

Madigan, 1970)

These differences originate from diverging research interests that emerged during the cognitive revolution. Learning research continued to focus on stimulus-response contingencies and employed performance measures. Memory research, on the other hand, conceptually separated encoding, maintenance, and retrieval processes and probed the structure of knowledge representations. To that end, memory researchers adopted new experimental paradigms and measures, such as free recall, recognition, and meta-cognitive judgments.

As learning researchers have become more interested in the representation of learned information, mutual theoretical exchange with memory researchers has become more attainable. The algorithmic and representational level of analysis lends itself to explore the domain generality of the assumed representations. I believe the fields of learning and memory have much to gain from theoretical exchange. The behaviorist tradition of learning research may motivate in-depth considerations about characteristics of the to-be-retained material and environment, which, in turn, may simplify assumptions about the structure of memory. Similarly, tested formalized theories of long-term memory, when applied to findings from learning research, can inform debates about the complexity of learning processes that parallel those in research on memory.

In what follows, I will first briefly review recent theoretical discussions regarding the complexity of learning and memory processes (Chapter 2). I will then present two cases in which insights from computational-level analyses (i.e., regard for characteristics of the stimulus material and context) combined with tested representational explanations of memory suggest parsimonious explanations for empirical dissociations in evaluative learning (Chapter 3) and episodic long-term memory (Chapter 4). These explanations suggest reinterpretations of findings that are commonly taken as evidence for dual-process models. Finally, I report an experiment that shows that the model I apply to false recognition in long-term memory also accounts for false recognition in short-term memory (Chapter 5). These findings are consistent with unitary memory models and oppose the assumption of a short-term memory that is insulated from long-term memory and operates on distinct representations. Taken together, this work illustrates the broad explanatory scope of exemplar-based global matching models.

References

Anderson, J. R. (1990). *The adaptive character of thought*. L. Erlbaum Associates.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (pp. 89–195). Academic Press.

Bechtel, W., & Shagrir, O. (2015). The Non-Redundant Contributions of Marr's Three Levels of Analysis for Explaining Information-Processing Mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322. https://doi.org/10.1111/tops.12141

Bouton, M. E. (2007). *Learning and behavior: A contemporary synthesis*. Sinauer Associates.

Bower, G. H. (2000). A brief history of memory research. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 3–32). Oxford University Press.

Bower, J. M., & Beeman, D. (1998). *The book of GENESIS: Exploring realistic neural models with the GEneral NEural SImulation System* (2nd ed.). TELOS.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In J.-C. Lacaille, W. S. Sossin, V. F. Castellucci, & S. Belleville (Eds.), *Progress in Brain Research* (pp. 323–338). Elsevier.

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, *119*(3), 480–499. https://doi.org/10.1037/a0027791

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*, *20*(4), 631–642. https://doi.org/10.3758/s13423-013-0386-3

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dennett, D. C. (1981). *Brainstorms: Philosophical essays on mind and psychology*. Bradford Books.

Ebbinghaus, H. (1885). *Über das Gedächtnis [On memory]*. Duncker & Humblot.

Eliasmith, C., & Kolbeck, C. (2015). Marr's Attacks: On Reductionism and Vagueness. *Topics in Cognitive Science*, *7*(2), 323–335. https://doi.org/10.1111/tops.12133

Fiedler, K. (2016). Functional research and cognitive-process research in behavioural science: An unequal but firmly connected pair. *International Journal of Psychology*, *51*(1), 64–71. https://doi.org/10.1002/ijop.12163

French, R. M., & Thomas, E. (2015). Interactive Effects of Explicit Emergent Structure: A Major Challenge for Cognitive Computational Modeling. *Topics in Cognitive Science*, *7*(2), 206–216. https://doi.org/10.1111/tops.12135

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. Wiley.

Glass, A. L., Holyoak, K. J., & Santa, J. L. (1979). *Cognition*. Addison-Wesley.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

Groninger, L. K., & Groninger, L. D. (1980). A comparison of recognition and savings as retrieval measures: A reexamination. *Bulletin of the Psychonomic Society*, *15*(4), 263–266. https://doi.org/10.3758/BF03334527

Hanus, D. (2016). Causal reasoning versus associative learning: A useful dichotomy or a strawman battle in comparative psychology? *Journal of Comparative Psychology*, *130*(3), 241–248. https://doi.org/10.1037/a0040235

Hebscher, M., Wing, E., Ryan, J., & Gilboa, A. (2019). Rapid Cortical Plasticity Supports Long-Term Memory Formation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.09.009

Hintzman, D. L. (1993). Twenty-Five Years of Learning and Memory: Was the Cognitive Revolution a Mistake? In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 359–391). MIT Press.

Holland, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, *37*(1-2), 105–131. https://doi.org/10.1016/0010-0277(90)90020-K

Houwer, J. D. (2011). Why the Cognitive Approach in Psychology Would Profit From a Functional Approach and Vice Versa. *Perspectives on Psychological Science*, *6*(2), 202–209. https://doi.org/10.1177/1745691611400238

Hughes, S., Houwer, J. D., & Perugini, M. (2016). The functional-cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology*, *51*(1), 4–14. https://doi.org/10.1002/ijop.12239

Jamieson, R. K., Crump, M. J. C., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning & Behavior*, *40*(1), 61–82. https://doi.org/10.3758/s13420-011-0046-2

Jamieson, R. K., & Hauri, B. R. (2012). An exemplar model of performance in the artificial grammar task: Holographic representation. *Canadian Journal of Experimental Psychology*, *66*(2), 98–105. https://doi.org/10.1037/a0027023

Jamieson, R. K., & Mewhort, D. J. K. (2009a). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *The Quarterly Journal of Experimental Psychology*, *62*(3), 550–575. https://doi.org/10.1080/17470210802055749

Jamieson, R. K., & Mewhort, D. J. K. (2009b). Applying an exemplar model to the serial reaction-time task: Anticipating from experience. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1757–1783. https://doi.org/10.1080/17470210802557637

Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial-grammar task: String completion and performance on individual items. *The Quarterly Journal of Experimental Psychology*, *63*(5), 1014–1039. https://doi.org/10.1080/17470210903267417

Jenkins, H. M. (1979). Animal learning and behavior theory. In E. Hearst (Ed.), *The first century of experimental psychology* (pp. 177–228). Erlbaum Hillsdale, NJ.

Lachman, S. J. (1997). Learning is a Process: Toward an Improved Definition of Learning. *The Journal of Psychology*, *131*(5), 477–480. https://doi.org/10.1080/00223989709603535

Love, B. C. (2015). The Algorithmic Level Is the Bridge Between Computation and Brain. *Topics in Cognitive Science*, *7*(2), 230–242. https://doi.org/10.1111/tops.12131

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), nn.3655. https://doi.org/10.1038/nn.3655

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the Brain's Algorithm for Categorization from Its Neural Implementation. *Current Biology*, *23*(20), 2023–2027. https://doi.org/10.1016/j.cub.2013.08.035

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *275*(942), 483–519. https://doi.org/10.1098/rstb.1976.0090

Marr, D. (1982). *Vision*. W. H. Freeman and Company.

Mazur, J. E. (2012). *Learning and Behavior* (7th ed.). Pearson.

McClamrock, R. (1991). Marr's three levels: A re-evaluation. *Minds and Machines*, *1*(2), 185–196. https://doi.org/10.1007/BF00361036

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356. https://doi.org/10.1016/j.tics.2010.06.002

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1987). *Psychological and Biological Models*. MIT Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, N. E. (1959). Liberalization of basic S-R Concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science* (pp. 196–202). McGraw-Hill.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386. https://doi.org/10.1037/0033-2909.117.3.363

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183–198. https://doi.org/10.1017/S0140525X09000855

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts.

Newell, A. (1982). The knowledge level. *Artificial Intelligence*, *18*(1), 87–127. https://doi.org/10.1016/0004-3702(82)90012-1

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

O'Donohue, W., Ferguson, K. E., & Naugle, A. E. (2003). The structure of the cognitive revolution: An examination from the philosophy of science. *The Behavior Analyst*, *26*(1), 85–110.

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary Learning Systems. *Cognitive Science*, *38*(6), 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311–345. https://doi.org/10.1037/0033-295X.108.2.311

Pavlov, I. P. (1927/1960). *Conditioned reflex: An investigation of the physiological activity of the cerebral cortex*. Dover Publications.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Pierce, W. D., & Cheney, C. D. (2013). *Behavior Analysis and Learning* (5th ed.). Psychology Press.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Roediger, H. L. (2004). What Happened to Behaviorism. *APS Observer*, *17*(3).

Rogers, T. T., & McClelland, J. L. (2014). Parallel Distributed Processing at 25: Further Explorations in the Microstructure of Cognition. *Cognitive Science*, *38*(6), 1024–1077. https://doi.org/10.1111/cogs.12148

Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, *8*(3), 183–189. https://doi.org/10.1002/bs.3830080302

Rumelhart, D. E., James L. McClelland, & PDP Research Group. (1987). *Foundations* (12. Aufl). MIT Press.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schmidt, J. R., De Houwer, J., & Rothermund, K. (2016). The Parallel Episodic Processing (PEP) model 2.0: A single computational model of stimulus-response binding, contingency learning, power curves, and mixing costs. *Cognitive Psychology*, *91*, 82–108. https://doi.org/10.1016/j.cogpsych.2016.10.004

Shagrir, O., & Bechtel, W. (2018). Marr's Computational Level and Delineating Phenomena. In D. M. Kaplan (Ed.), *Explanation and Integration in Mind and Brain Science* (pp. 190–214). Oxford University Press.

Shevrin, H., & Dickman, S. (1980). The psychological unconscious: A necessary assumption for all psychological theory? *American Psychologist*, *35*(5), 421–434. https://doi.org/10.1037/0003-066X.35.5.421

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review*, *3*(3), 314–321. https://doi.org/10.3758/BF03210755

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, *57*(4), 193–216.

Skinner, B. F. (1985). Cognitive science and behaviourism. *British Journal of Psychology*, *76*(3), 291–301. https://doi.org/10.1111/j.2044-8295.1985.tb01953.x

Skinner, B. (1938). *The Behavior of Organisms: An Experimental Analysis*. Appleton-Century.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*(01), 1. https://doi.org/10.1017/S0140525X00052432

Spear, N. E., & Riccio, D. C. (1994). *Memory: Phenomena and principles*. Allyn & Bacon.

Stewart, I. (2016). The fruits of a functional approach for psychological science. *International Journal of Psychology*, *51*(1), 15–27. https://doi.org/10.1002/ijop.12184

Tolman, E. C., & Honzik, C. H. (1930). "Insight" in rats. *University of California Publications in Psychology*, *4*, 215–232.

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. I. Orientation and the short-cut. *Journal of Experimental Psychology*, *36*(1), 13–24. https://doi.org/10.1037/h0053944

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*(4), 344–354. https://doi.org/10.1037/h0043150

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, *21*, 437–484. https://doi.org/10.1146/annurev.ps.21.020170.002253

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*(1), 122–129. https://doi.org/10.1037/h0022014

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*(6), 559–573. https://doi.org/10.1037/h0028143

van der Helm, P. A. (2012). Cognitive architecture of perceptual organization: From neurons to gnosons. *Cognitive Processing*, *13*(1), 13–40. https://doi.org/10.1007/s10339-011-0425-9

Watson, J. B. (1924). *Psychology from the Standpoint of a Behaviorist* (2nd ed.). J. B. Lippincott.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Chapter 2

Exemplars as least-committed representations

FREDERIK AUST

One of the fundamental questions in the study of cognitive processes pertains to their dimensionality: Is the observed phenomenon driven by one or more latent causes? This question has been posed for phenomena as diverse as learning, memory, reasoning, and decision making (Evans, 2008; Mitchell et al., 2009; Yonelinas, 2002). Most theories can be broadly assigned to one of two categories: single- and dual-process theories, Figure 2.1. Single-process theories posit that all presented information enters into one cognitive process that determines the appropriate response, Figure 2.1A. In contrast, dual-process theories posit that the presented information engages one or both of two separable cognitive processes, which may work in concert or opposition to shape behavior, Figure 2.1B and C.

Common among several influential dual-process theories is the assumption that the two processes implement different trade-offs between informational fidelity and processing speed. The faster of the two processes operates on abstract, lossy information that becomes available quickly and enables rapid responding. The slower, more sophisticated process operates on richer information, is generally more accurate and accompanied by a richer subjective experience, but takes longer to complete. Hence, although the fast process yields accurate responses in many situations, it is more automatic or heuristic and potentially error-prone than the slow process. Given sufficient time to respond, the two processes may produce response conflicts. Such conflicts are typically resolved in favor of the slow

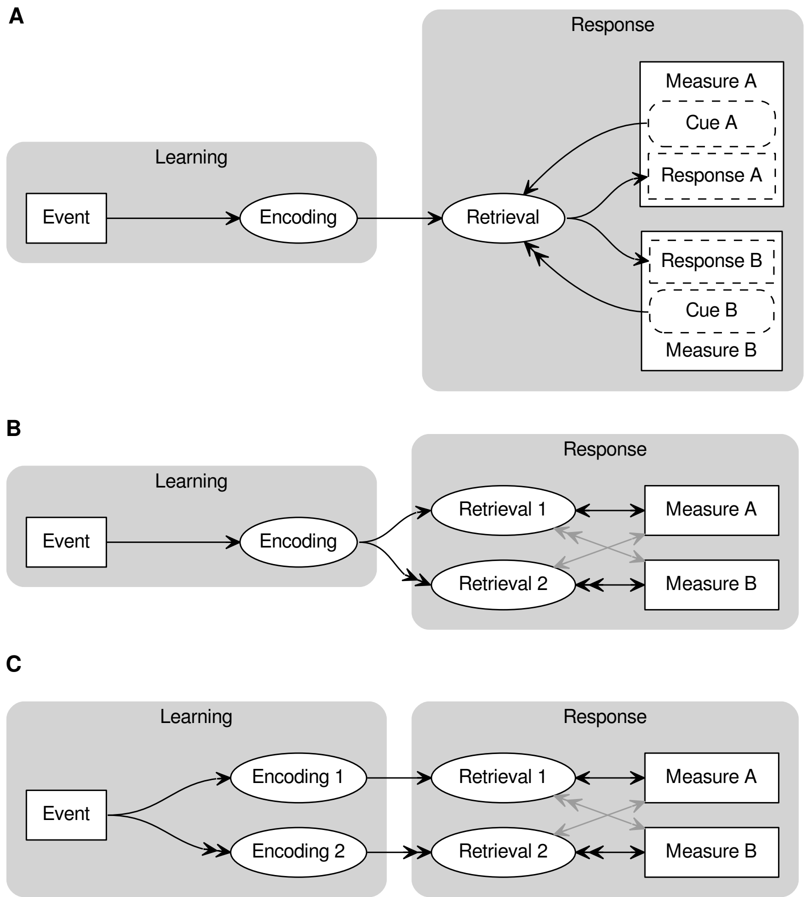

Figure 2.1: Visualization of the general structure of (**A**) single-process, (**B**) dual-retrieval, and (**C**) dual-representation explanations in learning and memory. Rectangles represent observables, ovals represent unobservable cognitive events, and the rounded boxes indicate that retrieval cues may be composed of a mixture of observable and unobservable features. Double arrows indicate relatively fast processing; gray arrows indicate possible effect mechanisms assumed to be absent when inferring multiple latent processes from functional dissociations. The comprehensive visualization of interactions between measures and retrieval processes in panel **A** is abbreviated to bidirectional arrows in panels **B** and **C**.

process. In some instances these conflicts have been described pointedly as two minds in one brain (Evans, 2008; Rydell et al., 2006; p. 87, Brainerd & Reyna, 2005) or, more martial, as a brain at war with itself (Stanovich, 2005; cited from Evans, 2008).

In the following I will review the dominant single- and dual-process perspectives in learning and memory and highlight the similarities in the theoretical debates. To organize the debate, I will distinguish between single-process models and two families of dual-process models—*dual-retrieval* and *dual-representation* models. That is, I organize the models along the processing stage to which they attribute the observed dissociations: the information used to cue retrieval (single-process), distinct retrieval processes (dual-retrieval), or distinct encoding processes (dual-representation). I will argue that models that attribute dissociations to later stages are more parsimonious and should be preferred until refuted. In particular, dual-representation assumptions require evidence of dissociations attributable only to encoding processes. Based on this discussion and my review of the literature I conclude that the available evidence currently does not necessitate dual-representations. I furthermore highlight a common element in several theories of learning and memory: They postulate a rich unitary representation of past events combined with parallel similarity-based retrieval. That is, these theories attribute empirical dissociations to the retrieval response stage. Based on these considerations, I suggest that global matching models of memory and their exemplar representation are promising candidates in the search for common theoretical ground of learning and memory.

To illustrate, I focus on two specific phenomena, namely evaluative conditioning and false recognition. For both, influential dual-representation explanations are being discussed and both phenomena have had considerable impact on theoretical developments in their respective field at large.

Evaluative conditioning is a change in liking following pairings of the to-be-evaluated object (conditioned stimulus, CS) and a valent event (unconditioned stimulus, US; De Houwer, 2007). In recent years, evaluative conditioning has featured prominently in debates about the broad distinction between single- and dual-process models of learning. Proponents of single-process models have acknowledged that some of the most convincing evidence for dual-process models has been reported in evaluative conditioning (Mitchell et al., 2009). Dual-process theories typically assume that CS-

US pairings cause the formation of simple associative links that connect the CS and the US. As this link is strengthened, for example through repeated pairings, the liking of the CS converges to the valence of the US. These associative links are abstract and lossy in the sense that their strength is a single summary statistic of the co-occurrence history. Qualifying information, such as the relationship between CS and US, is lost. Via the so-formed links, the CS elicits an immediate, intuitive response—either positive or negative depending on the valence of the US. According to dual-process theories, CS-US pairings additionally leave propositional representations that subserve more complex reasoning processes. In contrast to associative links, these propositions do include qualifications of the CS-US relationship and are verifiable in the sense that they can be true or false. That is, propositional representations enable more nuanced, contextualized attitudes. For an illustration of how these processes interact, consider an upcoming visit to the dentist. The thought of visiting the dentist may evoke an immediate negative response because such visits are associated with pain from drilling or root canal treatments. Although dental treatments can be painful, they undoubtedly serve to *relieve* acute pain and prevent future pain, more aggressive treatments, and tooth loss. In light of these facts your overall attitude towards regular visits to the dentist may be positive, despite the initial negative gut response.

False recognition is the assertion that an object has been encountered before, when, in fact, it has not. In contrast to random guessing, false recognition is caused by retrieval of fabricated, distorted, or misinterpreted information and may be accompanied by strong subjective feelings of confidence (Brainerd & Reyna, 2005; Gallo, 2006). As outlined in the previous chapter, false recall and false recognition have shaped theories of the organization of memory since the advent of the cognitive revolution (Underwood, 1965; Deese, 1959) and remain influential today. Especially, high-confidence false memories continue to stimulate new theoretical developments (e.g., Diana et al., 2006; Brainerd et al., 2015; Brainerd & Reyna, 2018; Brainerd et al., 2020). Dual-process theories commonly assume that experienced events leave gist-like memory traces that represent an episode's conceptual, semantic, and associative features (e.g., Reyna & Brainerd, 1995b). These gist-like traces are abstract and lossy in the sense that details of the specific episode, such as accompanying sensory impressions and context, are lost. When retrieved, gist-like traces cause an immediate sense of familiarity but provide little to no information about the origin of said familiarity. In addition to gist-like traces, experienced events may also leave a more detailed trace of the episode, including perceptual and contextual details. When retrieved, these detailed traces induce a vivid remembrance of the past episode including sensory impressions, such as smells. In contrast to gist-like traces, these detailed traces can serve to distinguish between previous episodes from similar but new events. For an illustration of how these processes interact, again, consider a visit to the dentist. When asked it may seem that, as usual, one attended the professional tooth cleaning immediately prior to the prophylactic dental checkup. Only upon further reflection—and retrieval of a detailed trace—one may remember that in fact the professional tooth cleaning was canceled the day before the appointment.

2.1 From single- to dual-representations and back

The above examples make evident the similarities of dual-process assumption in evaluative conditioning and false recognition. In both cases dual-process theories assume a faster process that operates on abstract, lossy information and a slower process that operates on richer information. Conflicts between these processes are typically resolved in favor of the slower process. These similarities are, in part, attributable to parallels in the iterative refinement of the theories in response to new evidence (p. 195, Mitchell et al., 2009; p. 60 ff., Brainerd & Reyna, 2005; Brainerd & Reyna, 1998). Both phenomena were initially explained solely by a fast process operating on lossy representations. To accommodate for new empirical evidence, an additional slower process operating on more detailed representations was assumed. Given the assumption of detailed representations the necessity of assuming additional abstract, lossy representations became debatable—leading to the current discussion between single- and dual-process proponents. As a result there is a close relationship between competing single- and dual-process theories: dual-process theories often extend a single-process theory. That is, they postulate one process that is identical or closely related to the process posited by a competing single-process theory.

Consider the case of evaluative conditioning. Initial theories of evaluative conditioning were strongly inspired by contemporary theories of Pavlovian conditioning. Indeed, EC, like other forms of learning, was initially understood in terms of simple associative links (Martin & Levey, 1978; De Houwer et al., 2001; De Houwer, 2007; Hofmann et al., 2010). Hence, it was assumed that EC was independent of more complex processes, such as reasoning—the change in liking was conceived to be entirely intuitive. The assumption of associative links was so deeply rooted that it was often conflated with the observable effect and thought of as a defining feature of EC (De Houwer, 2007). With heightened interest in cognition and new empirical evidence, it has become increasingly clear that learning, including Pavlovian conditioning, more broadly involves richer representation and more complex processing (De Houwer, 2009; Shanks, 2007). The prominent dual-process view that has come out of these considerations is that evaluative conditioning involves both associative links and propositional representations of CS-US pairings (e.g., Bar-Anan & Moran, 2018; De Houwer, 2007; Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; McConnell & Rydell, 2014; Smith & DeCoster, 2000). The competing single-process view fully embraces propositional representations and posits that the additional assumption of simple associative links is unnecessary (De Houwer, 2009; Mitchell et al., 2009; De Houwer, 2018).

Now consider the theoretical developments in false recognition. In the constructivist tradition, false recognition was initially thought to involve only abstract lossy representations (Reyna & Brainerd, 1995b; p. 60 Brainerd & Reyna, 2005). The central assumption was that an experienced event is integrated with associated semantic information and inferences. Through integration, the resulting representations emphasized the gist of the experience at the expense of detail (Reyna & Brainerd, 1995a; Reyna & Brainerd, 1995b). However, dissociations between memory and reasoning (that is, inferential processes) as well as findings suggesting that rich, detailed representations of experiences are retained in memory, motivated the development of fuzzy-trace theory, an influential dual-process model of false recognition (Reyna & Brainerd, 1995a; Reyna & Brainerd, 1995b; Brainerd & Reyna, 2002; Brainerd et al., 1999). The theory posits the parallel formation of rich verbatim in addition to gist-like representations. The competing single-process view, again, rejects the assumption of abstract, gist-like representations and posits that only detailed representations of each episode are retained (e.g., Arndt & Hirshman, 1998).

Thus, contrary to initially dominant theoretical positions, empirical findings on learning and memory indicate that episodic memory retains rich, detailed representations of each experience. This insight sparked the ongoing debate between single- and some dual-process theorists about the necessity of abstract, inherently lossy representations given their redundancy with the rich, detailed representations.

2.2 Telling dual-representations from dual-retrieval

Before I review the current theoretical debates in learning and memory with a focus on representational assumptions, I will briefly discuss what constitutes evidence for multiple latent causes in general and multiple representations in particular In doing so, I elaborate on the distinction between dual-retrieval and dual-representation models. As is typically the case in psychology, understanding the processes that underlie learning and memory is challenging because the constructs of interest are unobservable. Therefore, robust conclusions must combine (1) functional investigations of the number of irreducible dimensions underlying an observable phenomenon and (2) theory-guided interpretation of the identified dimensions. On their own, both approaches are limited.

A common approach to investigate the number of latent causes of a phenomenon is to produce functional or task dissociations (e.g., Yonelinas, 2002). The idea is to identify separable cognitive processes by exploring their profiles of functional characteristics (e.g., fidelity, forgetting rate, or processing speed). As such, this approach presupposes tasks or experimental conditions that quantify the operation of the cognitive processes. Given a set of two such tasks, functional dissociations—and by extension separable latent processes—are concluded from double dissociations: For each task there is at least one manipulation that selectively influence performance on that task, but not the other.

However, this logic of inferring latent multidimensionality from functional dissociations is flawed for two reasons (Dunn & Kirsner, 1988; Dunn & Kalish, 2018): First, double dissociations do not, in general, necessitate multiple latent causes (Dunn & Kirsner, 1988; Newell & Dunn, 2008). Second, the logic underlying this approach rests on two critical assumptions that are unlikely to hold in most applications (p. 96, Dunn & Kirsner, 1988). Selective influence presupposes that a manipulation affects only one underlying process and that this process in turn contributes to only one of the two tasks. Despite these flaws, there are numerous prominent examples of selective influence assumptions in theoretically influential work in both learning (e.g., Rydell et al., 2006; Hütter et al., 2012; Gawronski et al., 2014) and memory (e.g., Dunn, 2008; Wixted & Mickes, 2010).

State-trace analysis is a logically consistent and less assumption-laden approach to inferring multiple latent causes by mapping the functional relationship between two measures that ostensibly reflect the same process (Bamber, 1979; Dunn & Kirsner, 1988; Dunn & Kalish, 2018). In contrast to functional dissociations, none of the assumptions related to selective influence of the manipulation or underlying processes are required. State-trace analysis only presupposes that the two measures are monotonically related to the assumed single latent dimension. If this assumption holds, both measures must be monotonically related; any departure from monotonicity constitutes evidence for multiple latent dimensions.

State-trace analysis may be viewed as a computational-level analysis (Marr, 1982): a characterization of the functional relationship between two observable quantities. Nonetheless, it provides insights into the latent dimensionality of a phenomenon—constraining algorithmic and representational explanations. State-trace analysis in itself, however, yields no insight into the nature of the latent dimensions. That is, it cannot differentiate among any of the algorithmic and representational explanations consistent with the functional-level analysis (p. 100, Dunn & Kirsner, 1988). What is more, and will be central for the following discussion, not all results that necessitate rejections of unidimensionality are relevant to the single- vs. dual-process debates. As Dunn and Kirsner (1988) point out:

> *In one sense, rejection of the single-process model is trivial. It must surely always be the case that two tasks, sufficiently different to be interesting, cannot have all their component processes in common. (p. 100)*

Hence, whether a rejection of unidimensionality informs an algorithmic and representational explanation depends on the combination of measures and manipulations and cannot be judged without references to available theories. As noted by Tulving and Bower (1974), "this restriction holds regardless of what method is used" (pp. 296–297).

2.2.1 Dissociations are ambiguous

The ambiguity inherent in the inferred multidimensionality has contributed to the development of, as I will argue, unnecessarily complex dual-representational assumptions to explain evaluative conditioning and false recognition. Specifically, many dissociations that are interpreted as evidence for multiple independent representations may similarly be explained by distinct algorithms that operate on a common representational substrate (e.g., Hintzman, 1990). As Hintzman (1993) put it,

> *functional dissociations and stochastic independence between memory tasks have been taken as evidence that different memory systems underlie the tasks [. . .]—but both functional dissociations and stochastic independence can be easily derived from models assuming a single memory system, so their diagnosticity is an illusion (p. 381)*

Multidimensionality can be attributed to different stages of the cognitive processing cascade. First consider single-process models, which posit that stimulus information is represented in a unitary memory system and accessed via one retrieval mechanism, Figure 2.1A. Here, multidimensionality may arise, for example, because the tasks or conditions used to probe the cognitive architecture function as different cues to the retrieval mechanism. Clark and Gronlund (1996) termed this *information dissociations* (p. 56). For example, the measures may (inadvertently) add different contextual information to the retrieval cue and thereby promote retrieval of different subsets of available episodic representations. As noted by Dunn and Kirsner (1988), such multidimensionality is trivial—it contributes to our understanding of the measures, but does not constrain the representations and algorithms. To give a concrete example, information dissociations have been suggested to underlie dissociations between explicit and implicit memory tasks (Humphreys, Bain, et al., 1989). In Chapter 3, I present evidence that a context-based information dissociation may drive dissociations between liking and expectancy, which some consider evidence for dual-representation models of evaluative conditioning.

Now consider dual-retrieval models, which again posit that stimulus information is represented in a unitary system but accessed via two distinct retrieval mechanisms, Figure 2.1B. Here, multidimensionality may additionally be caused by two independent retrieval mechanisms or by one retrieval mechanism that produces two independent pieces of information. Several memory theories propose that recognition memory is driven by a memory-strength signal—experienced as unspecific familiarity—and recollection—experienced as availability of specific memory content (e.g., Yonelinas, 2002; Malmberg, 2008; Wixted & Mickes, 2010). Although the strength of memory and the retrieved memory content may be products of the same retrieval process they are theoretically independent properties (Hintzman, 1990). In fact, if pitted against one another memory-strength and -content may promote opposite responses. Such is the case, when in false recognition experiments a lure produces great memory-strength, but there is a decisive mismatch between the lure and the retrieved memory contents (e.g., Malmberg, 2008). This multidimensionality cannot be cast as a trivial difference between tasks and informs our understanding of representations or algorithms.

Finally, consider dual-representation models, which posit that stimulus information engages two encoding mechanisms, which produce different representations each of which is accessed by distinct retrieval mechanisms, Figure 2.1C. Here, multidimensionality may additionally emerge as information is being encoded into qualitatively different representations. As outlined above, theories on evaluative conditioning and false recognition share the assumption of two independent representations, one abstract and lossy, one rich in detail. Distinguishing between dual-representation and dual-retrieval models requires specific tests of separate encoding processes. For example, some dual-representation theories of evaluative learning posit that encoding of abstract link-based associations requires next to no cognitive resources and proceeds even in the absence of stimulus awareness. I will return to this issue shortly in my review of the current theoretical debate in evaluative conditioning. Another approach to testing dual-representation models is to preclude information or retrieval dissociations by the design of the study. With this approach the robustness of the conclusions critically hinge on knowledge about the to-be-tested processes. I used this approach in the experiments discussed in Chapter 4 and Chapter 5.

In sum, when reviewing the evidence in support of dual-process models, it is important to clearly distinguish between dual-retrieval and dual-representation models, Figure 2.1. This distinction helps to organize the available evidence along the processing stages and outlines a path for incremental theory extension from late to early processing stages: For each piece of evidence one should first consider whether it may be explained as an information dissociation. If not, one should consider whether it can be explained by different retrieval processes that operate on a unitary representation. Only if neither of these explanations seem tenable should dual-representations be invoked. Hence, the distinction between dual-retrieval and -representation models focuses attention on specific evidence for distinct representations as opposed to retrieval mechanisms. More generally, this distinction is useful for theorizing because it demands specification of the representational assumptions. Lastly, it motivates scrutiny of the measures to better understand the cues they may provide to the retrieval mechanism(s).

2.3 Least-committed representations

Special attention to the evidence for dual representations is warranted because they constitute a strong, lavish assumption. Rich unprocessed representations, as typically assumed by single-process and dual-retrieval models, allow the individual to meet the shifting demands of a complex environment (and are therefore conducive to integrative theorizing). Lossy abstract representations on the other hand are less versatile. This is articulated in Marr's principle of least commitment (pp. 485–486, 1976; cited from pp. 598-599, 2008): At the time of stimulus encoding it is difficult to foresee all relevant environmental demands and tasks. To some extent preprocessing or information compression operations reflect a commitment to a set of anticipated task demands. Such anticipatory commitment occasions disproportionate costs if it needs to be undone in light of unforeseen task demands. In this sense, it is adaptive and efficient to encode experiences in as rich a representation as the capacity of the system permits. Thus, the principle of least commitment implies that assumptions of abstract lossy representations require a strong theoretical rational. For example, computational-level analyses may reveal environmental factors that demand routine, fast, resource-saving responding. Such optimized responding may be appropriate because the respective stimulus occurs frequently or because it is critical for survival and must be guaranteed to function when the individual operates at her (e.g. attentional) capacity limit. As alluded to before, another reason to assume more committed representations is a system's limited storage or processing capacity.

To focus attention on representational assumptions, in the remainder I will refer to single-process and dual-retrieval models as *least commitment models*, because they assume rich detailed representations, and to dual-representation models as *precommitment models*[1], because they posit the formation of lossy abstract representation.

2.4 Evidence for precommitted representations

In the previous section, I have discussed the distinction between dual-retrieval and dual-representation models and the importance of specific evidence for multiple representations in general. Next, I will review the literature on learning and memory with a focus on representational assumptions. In doing so, I highlight a trend towards least commitment explanations in both learning and memory.

2.4.1 Learning

As noted above, the assumption that learning is mediated by simple associative links has for a long time been ingrained in theories of learning. The paradigm-shifting success of the famous Rescorla-Wagner model of conditioning (Miller et al., 1995; Rescorla & Wagner, 1972) appears to support this assumption. Mitchell et al. (2009) and McLaren et al. (2014) comprehensively discuss the arguments in favor of and against associative links. Of particular relevance to my work is the observation that link-based explanations of learning are fundamentally incompatibility with remembering (e.g., Bouton, 1993; Jozefowiez, 2018; Lovibond & Shanks, 2002; Miller & Escobar, 2001; Mitchell et al., 2009). Miller and Escobar (2001), for example, distinguished between acquisition- and performance-focused models—which roughly correspond to pre- and least-commitment

[1] I borrow the term precommitment from research into decision making and negotiation (Schelling, 1981; Strotz, 1955). It refers to the act of taking actions now to deliberately limit future choice options. As such, precommitment is a strategy to safeguard the execution of a rationally devised plan despite subsequently arising adverse conditions that distract from the committed-to goal, such as changes in motivation or shifts in value perception. Not unlike a general who burns a bridge after his troops have crossed over into enemy territory (Schelling, 1981), precommitment limits the available options to ensure the staunch response execution.

models—and noted that associative links are compressed summaries of the learning history and fundamentally incompatible with episodic memory:

> *people's sense of remembering specific events is not illusionary. Performance-focused models assume that these memories guide behavior. Alternatively, people might retain summary statistics as well as memories of specific past events, but use only summary statistics to guide behavior. However, this would leave memories of specific events as useless artifacts (which they might well be . . .). But such a conclusion is implausible given what is known about natural selection. (p. 145, Miller & Escobar, 2001)*

As I will note below, this reasoning bears a striking resemblance to recent arguments made in favor of recent dual-retrieval models of recognition memory (Wixted & Mickes, 2010).

But how can this call for performance-focused models, and least-committed representations, be reconciled with the success of link-based models such as the Rescorla-Wagner model? Jozefowiez (2018) has recently argued convincingly that this seeming contradiction can be resolved by recasting the model as a computational-level explanation of predictive behavior. According to this view, the Rescorla-Wagner model does not commit to an associative link representation:

> *A cognitive event cannot be both the retrieved memory of a past event and, at the same time, the expectation of a future event. [. . .] predictions rely on memory but are not identical to them. [. . .] Associations in the Rescorla-Wagner model are predictions in disguise. (p. 23, Jozefowiez, 2018)*

The most prominent alternative to associative links in theories of learning is a propositional representation (Mitchell et al., 2009). The underlying assumption is that learning is a consequence of more basic cognitive processes, such as attention, memory, and reasoning—that is, learning is a memory-based reasoning process. Hence, these propositional explanations inherit their assumption about representations and retrieval processes from theories of memory. In particular, Mitchell et al. (2009) noted that MINERVA 2 (Hintzman, 1988; Hintzman, 1984; Hintzman, 1986), a simple but powerful model of episodic memory, may be the simplest model consistent with their proposal (p. 187). In general, however, propositional models are rather unspecific verbal models that do not commit to a particular memory model. In Chapter 3, I present a first draft of a formalized explanation of evaluative conditioning based on MINERVA 2.

As noted above, most theorists now agree that human learning involves elaborate processing based on rich representations, typically assumed to be in propositional format. Thus, the current single- vs. dual-process debate revolves around the existence of associative links (Mitchell et al., 2009; McLaren et al., 2014).

2.4.1.1 Evaluative conditioning

Evaluative conditioning has received special attention in recent years because proponents of propositional models have acknowledged that here some of the most convincing evidence for associative links, and thus dual-representations, has been reported (p. 191, Mitchell et al., 2009). Some of the most influential dual-process theories of attitude learning are precommitment models and assume that stimulus pairings cause the formation of associative links as well as propositions at the time of encoding (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004; Wilson et al., 2000). To varying degrees these theories lean on four central pieces of evidence: (1.) Dissociations between explicit (e.g., self-report) and implicit measures of liking (e.g., inferred from response-latency), (2.) evaluative conditioning in the absence of conscious awareness of CS-US contingencies during learning (Lovibond & Shanks, 2002; Sweldens et al., 2014), (3.) evaluative conditioning effects that are not fully moderated by relational information, and (4.) dissociations between expectancy and liking (e.g., De Houwer et al., 2001; Sweldens et al., 2014).

Explicit and implicit measures of liking were long assumed to reflect predominantly propositional and associative processes, respectively. Under this assumption dissociations between explicit and implicit liking were taken as evidence for dual representations. However, as discussed above, this interpretation neglects the inherent ambiguity of these dissociations (Clark & Gronlund, 1996; Hintzman, 1990; see Stewart, 2016, for a similar argument from a functional perspective, p. 24). Accordingly, in a recent review Corneille and Stahl (2019a) conclude

> *The idea that the evaluative outcomes of distinct (propositional vs. associative) learning systems are expressed through different (direct vs. indirect) measures hardly withstands empirical evidence. [. . .] because indirect measures usually involve speeded responses, they involve different retrieval and expression processes. (p. 183, Corneille & Stahl, 2019a)*

Associative link-formation is typically assumed to be fast and effortless and should therefore proceed irrespective of conscious awareness of CS-US contingencies. In contrast to initial reports of unconscious evaluative conditioning, recent follow-up studies found little evidence for such claims when contingency awareness was rigorously controlled (e.g., Heycke et al., 2017; Heycke & Stahl, 2020; Moran et al., in press; Heycke et al., 2018; Heycke et al., in prep; Högden et al., 2018; Pleyers et al., 2007; Stahl & Unkelbach, 2009; Stahl, Haaf, et al., 2016). For example, Rydell et al. (2006) repeatedly presented an image of a fictitious man with valent descriptions of his behavior. Just before his image appeared on screen, a word, the valence of which was opposite to that of the described behavior, was briefly flashed. Because participants did not recognize the words after the learning phase, it was assumed that the words were processed subliminally and, thus, could only engage an effortless associative learning process. As a result of this learning procedure, participants expressed inconsistent attitudes towards the fictitious man: Self-reported liking corresponded to the valence of the behavioral information; in contrast, when assessed with an Implicit Association Test (IAT), liking reflected the opposite valence of the briefly flashed words. This striking example has often been cited to support precommitment claims (e.g., Rydell & McConnell, 2006; Peters & Gawronski, 2011; Sweldens et al., 2014). In stark contrast, each of four replication studies by Heycke et al. (2018) and Heycke et al. (in prep) contradict the original result: Explicit and implicit liking consistently reflected the valence of the described behaviors—there was no effect of the briefly flashed words. Although not all replication studies of unconscious evaluative conditioning yield such indisputable contradictory results (e.g., Moran et al., in press), overall the evidence is weak (Corneille & Stahl, 2019a).

A key property of associative links is that, in contrast to propositions, they ignore qualifying information about the relationship between CS and US or the validity of information. Recently, several studies have reported EC effects that are not (fully) moderated by relational information (e.g., Heycke & Gawronski, 2020; Hütter et al., 2012; Kukken et al., 2020; Moran & Bar-Anan, 2013a; Moran et al., 2016), which may indicate a contribution of unqualified associative links. For example, Moran and Bar-Anan (2013a) found that alien characters who stopped an unpleasant noise were evaluated more favorably than alien characters who ended pleasant music. However, response times in the IAT indicated opposite attitudes: participants preferred alien characters that co-occurred with pleasant music to those that co-occurred with an unpleasant noise—regardless of the qualifying relational information. Although these results appear to support dual-representations, an follow-up study, again, suggests that this post-learning dissociation between measures most likely reflects different response strategies afforded by the tasks (Bading et al., 2020)—a form of information dissociation. Other authors have developed formal measurement models (i.e., multinomial processing tree models, MPT) to more directly measure the contribution of relationally qualified and unqualified liking components (Heycke & Gawronski, 2020; Hütter et al., 2012; Kukken et al., 2020). But again, an as of yet unpublished study from our lab indicates that key assumptions of the measurement procedure are violated and that the results of these studies provide no evidence for associative link formation.

Finally, dissociations between expectancy and liking in EC are theoretically interesting because the expressed expectancies are assumed to be intimately related to other predictive learning phenomena, such as Pavlovian conditioning (Lovibond, 2004) and human associative learning more generally (Mitchell et al., 2009). Hence, dissociations between expectancy and liking indicate unique qualities of EC, which have been cited as evidence for associative link-formation (e.g., Baeyens et al., 2009a; Gawronski et al., 2014; Gawronski & Bodenhausen, 2006). The most commonly cited expectancy-liking dissociation is the apparent resistance of EC to extinction. EC effects that survive repeated presentations of the CS in isolation suggest that temporo-spatial co-occurrences between CSs and USs, but not the predictive value of a CS, drive evaluative conditioning (e.g., De Houwer et al., 2001; Sweldens et al., 2014). Accordingly, precommitment theories posit that only CS-US contiguity drives associative link-formation (Sweldens et al., 2014).

In Chapter 3, I report a series experiments that test a single-process explanation of expectancy-liking dissociations: Lipp et al. (2010) argue that representations CS-US pairings include the (temporal) context. Critically, they propose that evaluations of CSs are, by default, not contextualized and reflect the entire learning history. In contrast, judgments of US expectancy, by default, are (temporally) contextualized and most strongly reflect the recent learning history. Similar context-based explanations of extinction have previously been proposed for Pavlovian conditioning and oppose more traditional accounts that posit unlearning of associative links (e.g., Amsel, 1967; Bouton, 1993; Capaldi, 1967; Rescorla, 2003). The default judgment strategies presumably are afforded by the tasks and, hence, constitute an information dissociation. Additionally, I explore whether this information dissociation account is consistent with the specific representational and retrieval assumptions of MINERVA 2. As argued by Jozefowiez (2018), my results confirm that expectancy and liking can dissociate despite being based on a unitary memory system (Hintzman, 1990) and support recent conceptualizations of EC as memory-based judgments (Aust et al., 2018; Gast, 2018; Stahl & Aust, 2018).

The above discussion shows that evidence for precommitment models of EC is weak. Under scrutiny, some findings appear to be artifacts of the methodology, others cannot be replicated, and still others can be more parsimoniously explained in terms of information dissociations. Accordingly, dual-process proponents are beginning to acknowledge the explanatory scope of the least-commitment perspective and the importance of clearly distinguishing between pre- and least-commitment explanations (e.g., Bar-Anan & Moran, 2018; Gawronski & Bodenhausen, 2018; Hütter & Rothermund, 2020). Thus, evaluative conditioning currently appears to be well within the scope of least-commitment (and single-process) models.

2.4.2 Recognition memory

The characterization of familiarity and recollection in dual-process models of recognition memory is quite varied (for an overview see Yonelinas, 2002). Not unlike learning theories, precommitment models of recognition memory posit that qualitatively different information drives familiarity and recollection (e.g., Reyna & Brainerd, 1995b; Mandler, 1980): However, in contrast to learning theories, there is some disagreement about the mapping of the information to these processes. Earlier models assumed that sensory or perceptual similarity between probe and memory trace drives familiarity whereas conceptual or elaborate information drives recollection (Mandler, 1980). Current theories of false recognition posit the opposite (Brainerd et al., 1999). Yonelinas (2002), however, notes that the available evidence suggests that familiarity may be driven by both conceptual and perceptual information and, hence, simple information-process mappings may be inadequate (p. 480). Studies that report perceptually false recognition further support this assertion (e.g., Brainerd et al., 1995; Koutstaal et al., 1999; Ly et al., 2013; Stahl, Henze, et al., 2016).

Single-process models assume that familiarity is a unidimensional memory strength signal. From a single-process perspective, the view that memory strength is exclusively based on abstract conceptual overlap is problematic. Reminiscent of the associative-link critique by Miller and Escobar (2001), Anderson and Bower (1972) noted that

> *The important feature to note about the strength theory of recognition is that it is "ahistorical"; that is, it assumes that a subject makes recognition decisions about an item not on the basis of detailed memory of the past history [. . .] It is this ahistorical character of strength theory which is the source of all its weaknesses. (p. 98, cited from Wixted & Mickes, 2010)*

Hence, contemporary single-process theories propose a more complex interpretation of familiarity: Rather than simply reflecting the strength of a memory it is construed as an evidence signal that contains historical information about past events. That is, as in learning theory, single-process proponents assume rich rather than abstract lossy episodic representations and that (some of) this information is accessible. Wixted and Mickes (2010) have argued that this view of familiarity aligns better with the dual-process characterization of recollection than familiarity (p. 1026). Thus the current debate revolves around the contribution of a lossy, ahistoric memory-strength signal and precommitted representations (p. 205, Norman et al., 2008). For recent discussions of the evidence for dual-processes in recognition memory see Diana et al. (2006), Yonelinas and Parks (2007), Wixted (2007), Parks and Yonelinas (2007), and Malmberg (2008). Although there are several dual-retrieval models that assume least-committed representations (e.g., Hintzman et al., 1992; Humphreys, Bain, et al., 1989; Malmberg et al., 2004; Malmberg, 2008), there are also several influential precommitment models (e.g., Brainerd et al., 2015; Brainerd et al., 2019; McClelland et al., 1995; Norman & O'Reilly, 2003; Schapiro et al., 2017). In the remainder, I will focus on the evidence for dual-representations from research into false recognition studies.

2.4.2.1 False recognition

> *Often the errors a subject makes were an important clue to how the subject encoded the stimulus. (Miller et al., 1960)*

As in research on perception, memory errors have informed theories of memory (e.g. Underwood, 1965; Miller et al., 1960) and, in particular, the finding of high-confidence false memories continue to challenge theories of episodic memory (e.g., Brainerd et al., 2015; Diana et al., 2006; Brainerd & Reyna, 2018; Malmberg, 2008).

Fuzzy-trace theory posits that two opponent processes contribute to true and false recognition (Brainerd et al., 2019; Reyna & Brainerd, 1995a; Reyna & Lloyd, 1997). Like most dual-process theories, one process, referred to as *gist* memory, is conceptualized as unspecific familiarity, which promotes true and false recognition and is driven by semantic, or conceptual similarity between the probe and memory traces.[2] The second process, referred to as *verbatim* memory is thought of as a recollective process that evokes a vivid and detailed remembrance of the past experience. As such, it promotes true recognition but inhibits false recognition. Fuzzy-trace theory makes the strong precommitment assumption that two encoding processes operate independently, in parallel, and leave (multiple) gist and a verbatim memory trace. At test, gist and verbatim traces presumably are retrieved independently.

Three central empirical results are central to justify the assumption of independent gist and verbatim representations (chapter 3 and 4, Brainerd & Reyna, 2005): Experimental research seems to indicate that (1.) meaning (gist) can be processed faster than the surface features of a stimulus (e.g., Draine & Greenwald, 1998; Reicher, 1969; Wheeler, 1970) . This finding, however, can be accounted for in single-process models that factor in dynamic perceptual processes at test (Brockdorff & Lamberts, 2000; Cox &

[2]Under some circumstances, strong gist activation can result in an erroneous reinstatement of contextual details resulting in *phantom recollection* (i.e., vividly experienced, high-confidence false memories; Brainerd et al., 2001).

Shiffrin, 2017). Additional evidence comes from studies showing that (2.) verbatim traces decay faster than gist traces (e.g., Toglia et al., 1999) and that (3.) gist and verbatim memory can be selectively strengthened by presenting an increasing number of new but gist-preserving items or by repeating study list items, respectively (e.g., Stahl & Klauer, 2009; Stahl & Klauer, 2008). However, my research indicates that these findings may not be beyond the scope of recent single-process and least-commitment models.

The claim that the gist of an item can be processed faster than its surface features is noteworthy because it is typically assumed that, for example, the letters that make up a word have to be processed before its meaning emerges. If, however, the meaning of a word is processed *prior* to its surface features, this would be strong evidence for parallel independent encoding of gist and verbatim memory traces. Two effects are typically cited to support this claim. The *word-superiority effect* is defined as facilitated recognition of letters that are embedded in a word compared to letters presented in isolation or as part of a non-word (Reicher, 1969; Wheeler, 1970). The conclusion that, therefore, meaning is processed in parallel to and faster than the surface features is however challenged by the finding that letter recognition is also facilitated in pronounceable but meaningless pseudowords (Baron & Thurston, 1973). Another effect cited in support of independent gist and verbatim encoding is unconscious semantic priming. For example, a brief masked presentation of a male or female name modulates the speed at which an immediately succeeding name can be classified as male or female (e.g., Draine & Greenwald, 1998). If the sex of the briefly presented name is associated with the same sex as the succeeding name responding is facilitated; conversely, responding is slowed when the names are associated with opposite sexes. The lingering influence of a subsequently unrecognizable word may seem to suggest that, as claimed by fuzzy-trace theory, meaning is preferentially processed and encoded even if surface features (i.e., verbatim information) are not. By this logic unconscious priming effects should be limited to semantic information. This, however, is not the case—surface features also modulate subsequent responding (e.g. Bodner & Dypvik, 2005; Koechlin et al., 1999). Also, it is well established that identifying verbatim information becomes unavailable faster than the gist (see below). Such differential forgetting rates, rather than separate parallel encoding processes, could underlie unconscious semantic priming (e.g., Potter et al., 2002; Potter, 1976). Critically, unconscious semantic priming can be explained by least-committed single-process models of recognition that explicitly take perceptual processes at test into account (Cox & Shiffrin, 2017). If some features of the prime "leak over" into the representation of the probe at test, they will bias the retrieval process—and yield an information dissociation. Note, also, that it is controversial whether semantic priming is possible in the absence of conscious prime processing, as acknowledge by proponents of fuzzy-trace theory (p. 104, Brainerd & Reyna, 2005); for critiques see for example Merikle and Reingold (1998) and Shanks and St. John (1994).

Ample evidence indicates that the details of an experience are forgotten more rapidly than its gist (p.148 ff. Brainerd & Reyna, 2005; e.g., Reyna & Titcomb, 1997). This finding is readily explained by assuming separate verbatim and gist traces that decay at different rates. However, some single-process least-commitment models also predict different forgetting rates. For example, when the gist of an item is instantiated by multiple study list items (as is the case in the commonly used DRM lists, Deese, 1959; Roediger & McDermott, 1995) the gist-repeating study list items strengthen the gist more than the details of each individual item (Hintzman, 1986). It should also be noted that some recent work calls into question the generality of this finding. In contrast to studies using verbal material (e.g. words or sentences; Kintsch et al., 1990; Koriat et al., 2003), details and gist seem to decay at comparable rates for photographs of everyday objects (Andermane & Bowers, 2015). Similarly, pairwise associated person-object-location-triplets appear to be forgotten in an all-or-none fashion (Joensen et al., 2020).

Finally, there is convincing evidence that encoding manipulations can selectively affect gist and verbatim memory. For example, gist memory can be selectively strengthened by increasing the number of new but gist-repeating items on the study list and verbatim memory can be strengthened by repeating study list items (e.g., Stahl & Klauer, 2008; Stahl & Klauer, 2009). In effect, increasing the number of new but gist-repeating study list items increases true and false recognition. On the other hand, if initially the gist of an episode is strongly encoded and recall strategies are prevented, repeated presentation of study list items selectively increases true recognition. Such increases in true recognition without corresponding increases in false recognition challenged many single-process models (p. 112, Brainerd & Reyna, 2005).

In Chapter 4, I report an experiment that tests, whether a single-process model can simultaneously account for the effects of selective influence manipulations of gist and verbatim memory. Contemporary single-process models assume that strengthening episodic memory traces, for example through repetition, causes them to become more distinct, which counteracts increases in false recognition—a process referred to as *differentiation* (e.g., Criss, 2006; p. 170, Criss & Howard, 2015). I demonstrate a theoretical correspondence between the Conjoint Recognition Model (Brainerd et al., 1999; Brainerd et al., 2001), a formal implementation of fuzzy-trace theory, and the Generalized Context Model (Nosofsky, 1986; Nosofsky, 1988; Nosofsky, 2011a)—a least-committed single-process model. Moreover, I show empirically that the Generalized Context Model is able to adequately account for a multidimensional pattern of true and false recognition, which was caused by established selective influence manipulations. These results suggest that gist and verbatim retrieval may be better thought of as independent familiarity increments by partial and exact matches between probes and memory traces in a unitary memory system.

As the above discussion shows, evidence for dual-representations in false recognition is not compelling. The reported dissociations can be parsimoniously explained in terms of information (or retrieval) dissociations and are, thus, within the scope of least-commitment models.

2.4.2.2 False recognition over the short-term

False memories are typically considered a phenomenon of long-term memory and studied with long lists or across relatively long-delays. However, recent studies have established that false recognition and recall can be elicited with very similar methods in short-term memory (e.g., Atkins & Reuter-Lorenz, 2008; Coane et al., 2007; Flegal et al., 2010). These findings have raised questions about the interplay between short- and long-term memory (e.g., Abadie & Camos, 2019). Traditionally, it has been assumed that short- and long-term memory are fundamentally distinct, that is, short-term memory presumably is insulated from long-term memory and powered by different representations. That is to say, the distinction between short- and long-term memory often involves dual-representation assumptions that are similar to but in some aspects fundamentally different from those in long-term memory. However, in recent years unitary memory models have become influential (e.g., Crowder, 1993; Jonides et al., 2008; Nairne, 1990; Nairne, 2002). Accordingly, precommitment explanations for false long-term memory, such as fuzzy-trace theory, have been invoked to explain false short-term memory (Jou et al., 2016; Festini & Reuter-Lorenz, 2013; Flegal & Reuter-Lorenz, 2014; Dimsdale-Zucker et al., 2018; Abadie & Camos, 2019). Others have argued that least-committed single-process models can also account for central short-term memory findings (e.g., Kahana & Sekuler, 2002; Nosofsky et al., 2011; for a review see Nosofsky, 2016). Consistent with unitary memory models, recent work indicates that such models may provide a common basis to jointly explain findings from both short- and longer-term memory (Nosofsky et al., 2020; Nosofsky, Cox, et al., 2014; Schurgin et al., 2019). Accordingly, to extend the account of false recognition form long- to short-term memory, in Chapter 5 I report an experiment that simultaneously employed selective influence manipulations of gist and verbatim memory in a short-term memory task. In support of unitary memory models, I find that the same model accounts for false recognition over the long and short term.

2.5 Global exemplar matching

A common thread running through many of the successful single-process and least-commitment models of learning and memory is that they are instantiations of *global matching models* (Clark & Gronlund, 1996; Kelly et al., 2017; and Osth & Dennis, 2020) or make similar assumptions. At the heart of global matching models is the assumption that each episode leaves a trace (an exemplar) in a unitary memory system—least-committed representations of each episode are retained. Information is retrieved from memory by matching the cue to each trace in parallel. The matching process computes the similarity between the cue and each memory trace and the aggregated cue-trace similarity is an index of the cues familiarity. Additionally, some global matching models also specify a similarity-based cued recall process that returns episodic information (e.g., Hintzman, 1984). The described representational and algorithmic assumptions have allowed global matching models to account for fast responses that, nonetheless are affected by the entire contents in memory—a combination of constraints that motivated dual-process models (pp. 38–39, Clark & Gronlund, 1996). Although some empirical findings have challenged global matching models (Criss & Howard, 2015; Osth & Dennis, 2020), contemporary advancements provide powerful accounts of a wide range of phenomena as diverse as attention, learning, and memory (Cox & Shiffrin, 2017; Logan, 2002; Osth & Dennis, 2020; Schmidt et al., 2016). Even stern proponents of dual-retrieval models acknowledge the power of these models (p. 491, Yonelinas, 2002).

Core features of global matching models are also explicitly or implicitly assumed by several influential models of learning. For example, configural theory of Pavlovian conditioning (Pearce, 2002) assumes that learning results in CS representations that can be later activated (or retrieved) not only by the same CS but also by similar CSs. The CS is matched, in parallel, to all representations in memory and the aggregate association between the activated representations and the US drives responding. At the heart of configural theory is the assumption that compound CSs, which consist of multiple stimuli, are represented configurations that are similar but independent of their constituent stimuli. This assumption, again, mirrors representational assumptions of global matching models (e.g., Clark & Gronlund, 1996; Kelly et al., 2017). Several models of attitude learning and evaluative conditioning also assume parallel similarity-based retrieval (De Houwer, 2018; p. 693 Gawronski & Bodenhausen, 2006; Mitchell et al., 2009; p. 111, Smith & DeCoster, 2000). Much like in global matching models, these assumptions predict that similar CS representations modulate conditioned responding and help to explain generalization and context effects.

Thus, exemplar representations, combined with a global matching retrieval mechanism, have proven to be a widely applicable approach to model diverse cognitive phenomena. Although exemplar models have their origins in memory research, their assumptions mirror those of several influential models of learning, which facilitates alignment with established theoretical approaches. As such, the exemplar representation is a promising candidate for a parsimonious alternative to dual-representation assumptions and a common representational substrate for theories of learning and memory.

2.5.1 Benefits

But what are the benefits of adopting an exemplar representation and a global matching perspective? As I noted in the previous chapter, discovery of representational formats that are consistent with to a broad range of phenomena is interesting because they form a basis for broadly applicable, well-constrained, parsimonious explanations. Thus, a unifying theoretical approach to learning and memory needs to identify a common representational format and defer distinct processing steps to the response stage (also see p. 707, Nosofsky, 1988). A domain-general representational format likely retains stimulus information after minimal preprocessing (Marr, 1982). Exemplar representations are strong candidates for such least-committed representations and, thus, a parsimonious alternative to dual-representation assumptions in learning and memory (also see Jozefowiez, 2018). As a result, this approach may serve as a basis for more integrative theorizing. I believe that this kind of theoretical exchange is mutually beneficial. The traditional focus of learning research on computational level explanations may expose further information dissociations and thereby help to simplify assumptions about the structure of memory [e.g., Houwer (2011); see Chapter 4 and Chapter 5]. Similarly, exemplar representations and the well-formalized and principled mechanisms of similarity-based global matching may serve to specify learning theories and inform the debate between single- and dual-process models (Chapter 3).

Last but not least, as discussed at length above, robust conclusions about the dimensionality of learning and memory require that empirical findings are interpreted in the context of specific theories. Tulving and Bower (1974) noted the importance of precise processing assumptions for theoretical inferences from empirical results:

> *explication of the logic that relates experimental outcomes to statements about properties of memory traces requires specific assumptions about how the stored information is processed when it is retrieved. It is only in the context of a particular process model that inferences can be meaningfully drawn from the experimental data. This restriction holds regardless of what method is used. (pp. 296–297)*

Global matching models make well-tested, mathematically formalized assumptions about representations and retrieval processes. Particularly, the field of evaluative conditioning, where no formalized theories exist, could benefit from exploring a more formalized approach to theorizing.

By virtue of their formalization, they facilitate a detailed analysis of theoretical assumptions and comparison across theories. Model analysis and the exploration of a model's predictions help to build an intuition about the mechanics of a model. What are the crucial assumptions that allow the model to describe a pattern of results? Which assumptions are responsible for empirically unsupported predictions? Such detailed understanding of the model mechanics promotes further theoretical developments. Hintzman (1990) pointedly highlights this benefit to mathematical formalization:

> *To have one's hunches about how a simple combination of processes will behave repeatedly dashed by one's own computer program is a humbling experience that no experimental psychologist should miss. (p. 111)*

A comparison of different formal models may help to pinpoint where predictions diverge and thereby inform empirical tests of these theories. In other cases, a thorough theoretical analysis of the models may reveal non-trivial commonalities or even mathematical equivalences (e.g., Humphreys, Pike, et al., 1989; Jones & Dzhafarov, 2014; Kellen & Klauer, 2019; Kelly et al., 2017). Understanding such commonalities can be informative to all theories involved. In the next chapter, I present a theoretical model analysis that reveals how multidimensionality of true and false recognition can arise in a single-process global matching model. These findings suggest a parsimonious reinterpretation dual-representations in fuzzy-trace theory.

References

Abadie, M., & Camos, V. (2019). False memory at short and long term. *Journal of Experimental Psychology: General*, *148*(8), 1312–1334. https://doi.org/10.1037/xge0000526

Amsel, A. (1967). Partial Reinforcement Effects On Vigor and Persistence: Advances in Frustration Theory Derived from a Variety of Within-Subjects Experiments. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (pp. 1–65). Academic Press. https://doi.org/10.1016/S0079-7421(08)60512-5

Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of a week. *Psychonomic Bulletin & Review*, 1–6. https://doi.org/10.3758/s13423-015-0800-0

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123. https://doi.org/10.1037/h0033773

Arndt, J., & Hirshman, E. (1998). True and False Recognition in MINERVA2: Explanations from a Global Matching Perspective. *Journal of Memory and Language*, *39*(3), 371–391. https://doi.org/10.1006/jmla.1998.2581

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, *36*(1), 74–81. https://doi.org/10.3758/MC.36.1.74

Aust, F., Haaf, J. M., & Stahl, C. (2018). A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000600

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion*, *34*(1), 128–143. https://doi.org/10.1080/02699931.2019.1604322

Baeyens, F., Vansteenwegen, D., & Hermans, D. (2009a). Associative learning requires associations, not propositions. *Behavioral and Brain Sciences*, *32*(02), 198. https://doi.org/10.1017/S0140525X09000867

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, *19*(2), 137–181. https://doi.org/10.1016/0022-2496(79)90016-6

Bar-Anan, Y., & Moran, T. (2018). Simple First: A Skeleton for an Evaluative Learning Model. *Social Psychological Bulletin*, *13*(3), 1–23. https://doi.org/10.5964/spb.v13i3.28761

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, *4*(2), 207–228. https://doi.org/10.1016/0010-0285(73)90012-1

Bodner, G. E., & Dypvik, A. T. (2005). Masked priming of number judgments depends on prime validity and task. *Memory & Cognition*, *33*(1), 29–47. https://doi.org/10.3758/bf03195295

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*(1), 80–99. https://doi.org/10.1037/0033-2909.114.1.80

Brainerd, C. J., Gomes, C. F. A., & Nakamura, K. (2015). Dual Recollection in Episodic Memory. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000084

Brainerd, C. J., Nakamura, K., Chang, M., & Bialer, D. M. (2019). Verbatim editing: A general model of recollection rejection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1776–1790. https://doi.org/10.1037/xlm0000684

Brainerd, C. J., Nakamura, K., & Murtaza, Y. A. (2020). Explaining complementarity in false memory. *Journal of Memory and Language*, *112*, 104105. https://doi.org/10.1016/j.jml.2020.104105

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-Trace Theory and Children's False Memories. *Journal of Experimental Child Psychology*, *71*(2), 81–129. https://doi.org/10.1006/jecp.1998.2464

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory* (Vol. 14). Oxford University Press.

Brainerd, C. J., & Reyna, V. F. (2018). Complementarity in false memory illusions. *Journal of Experimental Psychology: General*, *147*(3), 305–327. https://doi.org/10.1037/xge0000381

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-Recognition Reversal: When Similarity is Distinctive. *Journal of Memory and Language*, *34*(2), 157–185.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*(1), 160–179. https://doi.org/10.1037/0033-295X.106.1.160

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 307–327. https://doi.org/10.1037/0278-7393.27.2.307

Brainerd, C., & Reyna, V. (2002). Fuzzy-Trace Theory and False Memory. *Current Directions in Psychological Science*, *11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old–new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 77–102. https://doi.org/10.1037/0278-7393.26.1.77

Capaldi, E. J. (1967). A Sequential Hypothesis of Instrumental Learning. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (pp. 67–156). Academic Press. https://doi.org/10.1016/S0079-7421(08)60513-7

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*(1), 37–60. https://doi.org/10.3758/BF03210740

Coane, J. H., McBride, D. M., Raulerson, B. A., & Scott, J. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*(1), 62–70. https://doi.org/10.1027/1618-3169.54.1.62

Corneille, O., & Stahl, C. (2019a). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. https://doi.org/10.1037/rev0000076

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*(4), 461–478. https://doi.org/10.1016/j.jml.2006.08.003

Criss, A. H., & Howard, M. W. (2015). Models of episodic memory. In J. R. Busemeyer, Z. J. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (pp. 165–183). Oxford University Press.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, *21*(2), 142–145. https://doi.org/10.3758/BF03202725

De Houwer, J. (2007). A Conceptual and Theoretical Analysis of Evaluative Conditioning. *The Spanish Journal of Psychology*, *10*(02), 230–241. https://doi.org/10.1017/S1138741600006491

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*(1), 1–20. https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin*, *13(3)*, e28046. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853–869. https://doi.org/10.1037/0033-2909.127.6.853

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, *13*(1), 1–21. https://doi.org/10.3758/BF03193807

Dimsdale-Zucker, H. R., Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2018). Serial position-dependent false memory effects. *Memory*, *0*(0), 1–13. https://doi.org/10.1080/09658211.2018.1513039

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*(3), 286–303. https://doi.org/10.1037/0096-3445.127.3.286

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*, *115*(2), 426–446. https://doi.org/10.1037/0033-295X.115.2.426

Dunn, J. C., & Kalish, M. L. (2018). *State-trace analysis*. Springer.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*(1), 91–101. https://doi.org/10.1037/0033-295X.95.1.91

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Festini, S. B., & Reuter-Lorenz, P. A. (2013). The short- and long-term consequences of directed forgetting in a working memory task. *Memory*, *21*(7), 763–777. https://doi.org/10.1080/09658211.2012.754900

Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1331–1338. https://doi.org/10.1037/a0019903

Flegal, K. E., & Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Memory & Cognition*, (42), 1–11. https://doi.org/10.3758/s13421-013-0391-9

Gallo, D. (2006). *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*. Psychology Press.

Gast, A. (2018). A Declarative Memory Model of Evaluative Conditioning. *Social Psychological Bulletin*, *13(3)*, e28590. https://doi.org/10.5964/spb.v13i3.28590

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative Conditioning From the Perspective of the Associative-Propositional Evaluation Model. *Social Psychological Bulletin*, *13*(3), 1–33. https://doi.org/10.5964/spb.v13i3.28024

Gawronski, B., Gast, A., & De Houwer, J. (2014). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition and Emotion*, 1–15. https://doi.org/10.1080/02699931.2014.947919

Heycke, T., Aust, F., Banaji, M. R., Cone, J., Van Dessel, P., Ferguson, M., Hu, X., Jiang, C., Kurdi, B., Rydell, R. J., Spitzer, L., Stahl, C., Vitiello, C., & De Houwer, J. (in prep). *Of Two Minds: A registered replication*.

Heycke, T., Aust, F., & Stahl, C. (2017). Subliminal influence on preferences? A test of evaluative conditioning for brief visual conditioned stimuli using auditory unconditioned stimuli. *Royal Society Open Science*, *4*, 160935. https://doi.org/10.1098/rsos.160935

Heycke, T., & Gawronski, B. (2020). Co-occurrence and relational information in evaluative learning: A multinomial modeling approach. *Journal of Experimental Psychology: General*, *149*(1), 104–124. https://doi.org/10.1037/xge0000620

Heycke, T., Gehrmann, S., Haaf, J. M., & Stahl, C. (2018). Of two minds or one? A registered replication of Rydell et al. (2006). *Cognition and Emotion*, *32*(8), 1708–1727. https://doi.org/10.1080/02699931.2018.1429389

Heycke, T., & Stahl, C. (2020). No evaluative conditioning effects with briefly presented stimuli. *Psychological Research*, *84*(4), 1020–1027. https://doi.org/10.1007/s00426-018-1109-1

Hintzman, D. L. (1990). Human Learning and Memory: Connections and Dissociations. *Annual Review of Psychology*, *41*(1), 109–139. https://doi.org/10.1146/annurev.ps.41.020190.000545

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Hintzman, D. L. (1993). Twenty-Five Years of Learning and Memory: Was the Cognitive Revolution a Mistake? In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 359–391). MIT Press.

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 667–680. https://doi.org/10.1037/0278-7393.18.4.667

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Högden, F., Hütter, M., & Unkelbach, C. (2018). Does evaluative conditioning depend on awareness? Evidence from a continuous flash suppression paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(10), 1641–1657. https://doi.org/10.1037/xlm0000533

Houwer, J. D. (2011). Why the Cognitive Approach in Psychology Would Profit From a Functional Approach and Vice Versa. *Perspectives on Psychological Science*, *6*(2), 202–209. https://doi.org/10.1177/1745691611400238

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208–233. https://doi.org/10.1037/0033-295X.96.2.208

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, *33*(1), 36–67. https://doi.org/10.1016/0022-2496(89)90003-5

Hütter, M., & Rothermund, K. (2020). Automatic processes in evaluative learning. *Cognition and Emotion*, *34*(1), 1–20. https://doi.org/10.1080/02699931.2019.1709315

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*(3), 539–557. https://doi.org/10.1037/a0026477

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General*, *149*(2), 230–248. https://doi.org/10.1037/xge0000648

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*(1), 1–32. https://doi.org/10.1037/a0034190

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The Mind and Brain of Short-Term Memory. *Annual Review of Psychology*, *59*(1), 193–224. https://doi.org/10.1146/annurev.psych.59.103006.093615

Jou, J., Arredondo, M. L., Li, C., Escamilla, E. E., & Zuniga, R. (2016). The effects of increasing semantic-associate list length on the Deese–Roediger–McDermott false recognition memory: Dual false-memory process in retrieval from sub- and supraspan lists. *The Quarterly Journal of Experimental Psychology*, *0*(0), 1–18. https://doi.org/10.1080/17470218.2016.1222446

Jozefowiez, J. (2018). Associative versus predictive processes in Pavlovian conditioning. *Behavioural Processes*, *154*, 21–26. https://doi.org/10.1016/j.beproc.2017.12.016

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*(18), 2177–2192. https://doi.org/10.1016/S0042-6989(02)00118-9

Kellen, D., & Klauer, K. C. (2019). Selecting Amongst Multinomial Models: An Apologia for Normalized Maximum Likelihood. *PsyArXiv*. https://doi.org/10.31234/osf.io/9v2hd

Kelly, M. A., Mewhort, D. J. K., & West, R. L. (2017). The memory tesseract: Mathematical equivalence between composite and separate storage memory models. *Journal of Mathematical Psychology*, *77*, 142–155. https://doi.org/10.1016/j.jmp.2016.10.006

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*(2), 133–159. https://doi.org/10.1016/0749-596X(90)90069-C

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1882–1905. https://doi.org/10.1037/0096-1523.25.6.1882

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What Do We Know About What We Cannot Remember? Accessing the Semantic Attributes of Words That Cannot Be Recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1095–1105. https://doi.org/10.1037/0278-7393.29.6.1095

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C., & Jackson, E. M. (1999). Perceptually Based False Recognition of Novel Objects in Amnesia: Effects of Category Size and Similarity to Category Prototypes. *Cognitive Neuropsychology*, *16*(3-5), 317–341. https://doi.org/10.1080/026432999380816

Kukken, N., Hütter, M., & Holland, R. W. (2020). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating the effects of CS-US pairings and their meaning. *Cognition and Emotion*, *34*(1), 170–187. https://doi.org/10.1080/02699931.2019.1617112

López, F. J., & Shanks, D. R. (2008). Models of Animal Learning and Their Relations to Human Learning. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (1st ed., pp. 589–611). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.026

Lipp, O. V., Mallan, K. M., Libera, M., & Tan, M. (2010). The effects of verbal instruction on affective and expectancy learning. *Behaviour Research and Therapy*, *48*(3), 203–209. https://doi.org/10.1016/j.brat.2009.11.002

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*(2), 376–400. https://doi.org/10.1037/0033-295X.109.2.376

Lovibond, P. F. (2004). Cognitive Processes in Extinction. *Learning & Memory*, *11*(5), 495–500. https://doi.org/10.1101/lm.79604

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3–26. https://doi.org/10.1037/0097-7403.28.1.3

Ly, M., Murray, E., & Yassa, M. A. (2013). Perceptual versus conceptual interference and pattern separation of verbal stimuli in young and older adults. *Hippocampus*, *23*(6), 425–430. https://doi.org/10.1002/hipo.22110

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*(4), 335–384. https://doi.org/10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., Holden, J. E., & Shiffren, R. M. (2004). Modeling the Effects of Repetitions, Similarity, and Normative Word Frequency on Old-New Recognition and Judgments of Frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 319–331. https://doi.org/10.1037/0278-7393.30.2.319

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *275*(942), 483–519. https://doi.org/10.1098/rstb.1976.0090

Marr, D. (1982). *Vision*. W. H. Freeman and Company.

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy*, *1*(2), 57–101. https://doi.org/10.1016/0146-6402(78)90013-9

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model: A Dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 204–218). The Guilford Press.

McLaren, I. P. L., Forrest, C. L. D., McLaren, R. P., Jones, F. W., Aitken, M. R. F., & Mackintosh, N. J. (2014). Associations and propositions: The case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory*, *108*, 185–195. https://doi.org/10.1016/j.nlm.2013.09.014

Merikle, P. M., & Reingold, E. M. (1998). On demonstrating unconscious perception: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General*, *127*(3), 304–310. https://doi.org/10.1037/0096-3445.127.3.304

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* Henry Holt and Co. https://doi.org/10.1037/10039-000

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386. https://doi.org/10.1037/0033-2909.117.3.363

Miller, R. R., & Escobar, M. (2001). Contrasting Acquisition-Focused and Performance-Focused Models of Acquired Behavior. *Current Directions in Psychological Science*, *10*(4), 141–145. https://doi.org/10.1111/1467-8721.00135

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183–198. https://doi.org/10.1017/S0140525X09000855

Moran, T., & Bar-Anan, Y. (2013a). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion*, *27*(4), 743–752. https://doi.org/10.1080/02699931.2012.732040

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2016). The Assimilative Effect of Co-Occurrence on Evaluation Above and Beyond the Effect of Relational Qualifiers. *Social Cognition*, *34*(5), 435–461. https://doi.org/10.1521/soco.2016.34.5.435

Moran, T., Hughes, S., Hussey, I., Vadillo, M. A., Olson, M. A., Aust, F., Bading, K. C., Balas, R., Benedict, T., Corneille, O., Douglas, S. B., Ferguson, M. J., Fritzlen, K. A., Gast, A., Gawronski, B., Giménez-Fernández, T., Hanusz, K., Heycke, T., Högden, F., . . . De Houwer, J. (in press). Incidental Attitude Formation via the Surveillance Task: A Preregistered Replication of Olson and Fazio (2001). *Psychological Science*. https://doi.org/10.31234/osf.io/kavx4

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269. https://doi.org/10.3758/BF03213879

Nairne, J. S. (2002). Remembering Over the Short-Term: The Case Against the Standard Model. *Annual Review of Psychology*, *53*(1), 53–81. https://doi.org/10.1146/annurev.psych.53.100901.135131

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*(8), 285–290. https://doi.org/10.1016/j.tics.2008.04.009

Norman, K. A., Detre, G., & Polyn, S. M. (2008). Computational models of episodic memory. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 189–225). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.011

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (2011a). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Nosofsky, R. M. (2016, January 1). An Exemplar-Retrieval Model of Short-term Memory Search: Linking Categorization and Probe Recognition. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 47–84). Academic Press. https://doi.org/10.1016/bs.plm.2016.03.002

Nosofsky, R. M., Cao, R., Harding, S. M., & Shiffrin, R. M. (2020). Modeling short- and long-term memory contributions to recent event recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000812

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1524–1539. https://doi.org/10.1037/xlm0000015

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315. https://doi.org/10.1037/a0022494

Osth, A. F., & Dennis, S. (2020). Global matching models of recognition memory. *PsyArXiv*. https://doi.org/10.31234/osf.io/mja6c

Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, *114*(1), 188–201. https://doi.org/10.1037/0033-295X.114.1.188

Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, *30*(2), 73–95. https://doi.org/10.3758/BF03192911

Peters, K. R., & Gawronski, B. (2011). Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569. https://doi.org/10.1177/0146167211400423

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of experimental psychology. Learning, memory, and cognition*, *33*(1), 130–144. https://doi.org/10.1037/0278-7393.33.1.130

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, *2*(5), 509. https://doi.org/10.1037/0278-7393.2.5.509

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1163–1175. https://doi.org/10.1037//0096-1523.28.5.1163

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275–280. https://doi.org/10.1037/h0027768

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rescorla, R. A. (2003). Contemporary Study of Pavlovian Conditioning. *The Spanish Journal of Psychology*, *6*(2), 185–195. https://doi.org/10.1017/S1138741600005333

Reyna, V. F., & Brainerd, C. J. (1995a). Fuzzy-trace theory: Some foundational issues. *Learning and Individual differences*, *7*(2), 145–162. https://doi.org/10.1016/1041-6080(95)90028-4

Reyna, V. F., & Brainerd, C. J. (1995b). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1–75. Retrieved July 5, 2016, from http://www.sciencedirect.com/science/article/pii/1041608095900314

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*(2), 95–123. https://doi.org/10.1016/S1041-6080(97)90002-9

Reyna, V. F., & Titcomb, A. L. (1997). Constraints on the suggestibility of eyewitness testimony: A fuzzy-trace theory analysis. *Intersections in basic and applied memory research*, 157–174.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008. https://doi.org/10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of Two Minds: Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes. *Psychological Science*, *17*(11), 954–958. https://doi.org/10.1111/j.1467-9280.2006.01811.x

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schelling, T. C. (1981). *The strategy of conflict* (8. print). Harvard Univ. Pr.

Schmidt, J. R., De Houwer, J., & Rothermund, K. (2016). The Parallel Episodic Processing (PEP) model 2.0: A single computational model of stimulus-response binding, contingency learning, power curves, and mixing costs. *Cognitive Psychology*, *91*, 82–108. https://doi.org/10.1016/j.cogpsych.2016.10.004

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2019). Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. *Nature Human Behaviour*, 325472. https://doi.org/10.1101/325472

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, *60*(3), 291–309. https://doi.org/10.1080/17470210601000581

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367–395. https://doi.org/10.1017/S0140525X00035032

Smith, E. R., & DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review*, *4*(2), 108–131. https://doi.org/10.1207/S15327957PSPR0402_01

Stahl, C., & Aust, F. (2018). Evaluative Conditioning as Memory-Based Judgment. *Social Psychological Bulletin*, *13(3)*, e28589. https://doi.org/10.5964/spb.v13i3.28589

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal Evaluative Conditioning? Above-Chance CS Identification May Be Necessary and Insufficient for Attitude Learning. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000191

Stahl, C., Henze, L., & Aust, F. (2016). False memory for perceptually similar but conceptually distinct line drawings. *PsyArXiv*. https://doi.org/10.31234/osf.io/zr7m8

Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 570–586. https://doi.org/10.1037/0278-7393.34.3.570

Stahl, C., & Klauer, K. C. (2009). Measuring phantom recollection in the simplified conjoint recognition paradigm. *Journal of Memory and Language*, *60*(1), 180–193. https://doi.org/10.1016/j.jml.2008.08.001

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(2), 286–291. https://doi.org/10.1037/a0013255

Stanovich, K. E. (2005). *The robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press.

Stewart, I. (2016). The fruits of a functional approach for psychological science. *International Journal of Psychology*, *51*(1), 15–27. https://doi.org/10.1002/ijop.12184

Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, *23*(3), 165–180. https://doi.org/10.2307/2295722

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review*, *18*(2), 187–209. https://doi.org/10.1177/1088868314527832

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall Accuracy and Illusory Memories: When More is Less. *Memory*, *7*(2), 233–256. https://doi.org/10.1080/741944069

Tulving, E., & Bower, G. H. (1974). The Logic of Memory Representations. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 265–301). Academic Press. https://doi.org/10.1016/S0079-7421(08)60457-0

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*(1), 122–129. https://doi.org/10.1037/h0022014

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A Model of Dual Attitudes. *Psychological Review*, *107*(1), 101. https://doi.org/10.1037/0033-295X.107.1.101

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*(4), 1025–1054. https://doi.org/10.1037/a0020874

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, *133*(5), 800–832. https://doi.org/10.1037/0033-2909.133.5.800

Empirical research

Chapter 3

A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning

Frederik Aust, Julia M. Haaf, & Christoph Stahl

Evaluative conditioning (EC) is a change in liking of neutral conditioned stimuli (CS) following pairings with positive or negative stimuli (unconditioned stimulus, US). A dissociation has been reported between US expectancy and CS evaluation in extinction learning: When CSs are presented alone subsequent to CS-US pairings, participants cease to expect USs but continue to exhibit EC effects. This dissociation is typically interpreted as demonstration that EC is resistant to extinction, and consequently, that EC is driven by a distinct learning process. We tested whether expectancy-liking dissociations are instead caused by different judgment strategies afforded by the dependent measures: CS evaluations are by default integrative judgments—summaries of large portions of the learning history—whereas US expectancy reflects momentary judgments that focus on recent events. In a counterconditioning and two extinction experiments, we eliminated the expectancy-liking dissociation by inducing nondefault momentary evaluative judgments, and demonstrated a reversed dissociation when we additionally induced nondefault integrative expectancy judgments. Our findings corroborated a-priori predictions derived from the formal memory model MINERVA 2. Hence, dissociations between US expectancy and CS evaluation are consistent with a single-process learning model; they reflect different summaries of the learning history.

This chapter is a reprint of a paper accepted for publication at *Journal of Experimental Psycholog: Learning, Memory, and Cognition* on February 28th, 2018.

The final article is available via its DOI: 10.1037/xlm0000600

Evaluative conditioning (EC) is a change in liking of neutral *conditioned stimuli* (CS) following pairings with positive or negative *unconditioned stimuli* (US; De Houwer, 2011; Hofmann et al., 2010). For example, an initially neutral brand logo (the CS) that is repeatedly paired with positive stimuli (the USs) in advertisement settings is later evaluated more positively compared to initial evaluations or unpaired logos. In this sense, EC is considered to be a model of the effects of advertising (Biegler & Vargas, 2013), and of attitude acquisition in general (De Houwer et al., 2001).

Most human associative learning phenomena can be accounted for by a propositional process which presumably requires conscious awareness of the to-be-learned regularities—the CS-US contingencies—to affect behavior (Mitchell et al., 2009). It has been argued, however, that EC violates this principle (Baeyens & De Houwer, 1995; De Houwer et al., 2001). Multiple studies report that EC may occur without conscious awareness of CS-US contingencies (Lovibond & Shanks, 2002; see Sweldens et al., 2014, for a

Individual contributions

Frederik Aust: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing

Julia M. Haaf: Conceptualization, Investigation, Methodology, Resources, Software, Validation, Writing - Original Draft Preparation, Writing - Review & Editing

Stahl Christoph: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing - Original Draft Preparation, Writing - Review & Editing

Contributor Roles Taxonomy (CRediT; Allen et al., 2014; Holcombe et al., 2020)

recent review). Moreover, EC has been claimed to be resistant to extinction and, hence, to occur despite conscious awareness of the absence of CS-US contingencies (Baeyens et al., 1988; Baeyens et al., 2005; Dwyer et al., 2007; Hermans et al., 2002; Vansteenwegen et al., 2006). This dissociation between expectancy and liking cannot be readily explained by the aforementioned propositional learning process; hence, EC was taken to involve distinct processes that differ from those underlying other associative learning phenomena (De Houwer et al., 2001). Consequently, dual-process theories of attitude acquisition, which postulate an additional, automatic, associative learning process (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004; Wilson et al., 2000), have become popular among EC theorists.

Recent research has, however, cast doubt on whether these critical findings hold, and if they do, whether a dual-process account is in fact necessary to explain them. Overall, the evidence for EC without CS-US contingency awareness is weak, with unintentional (incidental) EC perhaps best supported by the data (Sweldens et al., 2014; Corneille & Stahl, 2019b; Stahl, Haaf, et al., 2016; Heycke et al., 2017; Heycke et al., 2018). The present study investigates a recent single-process account of why EC appears to be resistant to extinction (Lipp & Purkis, 2006; Lipp et al., 2010; Lipp et al., 2003). We briefly review the central findings obtained with extinction procedures in EC research and then suggest a parsimonious single-process model for these findings. Next, we present three experiments that tested the model predictions in a counterconditioning and an extinction procedure.

3.1 Resistance to extinction

Until recently, the majority of EC studies supported the interpretation that EC is resistant to extinction (e.g. Baeyens et al., 1988; Baeyens et al., 2005; Hermans et al., 2002). For example, Hermans et al. (2002) report a dissociation between CS evaluation and US expectancy in two experiments. The authors used a common extinction procedure, in which CSs were paired with USs in an acquisition phase and presented alone in a subsequent extinction phase. To assess the effect of the extinction procedure on EC, they compared CS evaluations obtained after acquisition to those obtained after extinction. Hermans et al. (2002) found that EC was unaffected by the extinction phase, whereas US expectancy was extinguished.

The resistance of CS evaluation to extinction stands in contrast to the rapid extinction of conditioned responses observed in Pavlovian conditioning (Lovibond, 2004) and human associative learning more generally (Mitchell et al., 2009). Dissociations between US expectancy and CS evaluation, as those reported by Hermans et al. (2002), pointedly illustrate this contrast and are central to the debate between single- and dual-process learning theorists (e.g., Baeyens et al., 2009b). The latter suggests that, unlike US expectancy, CS evaluation is driven by a distinct learning process that presumably reflects temporo-spatial co-occurrences between CSs and USs (CS-US contiguity) but not the predictive value of a CS (statistical CS-US contingencies; e.g., Sweldens et al., 2014). Hence, the dual-process account posits that a change in the predictive value of a CS, for example due to extinction learning, affects US expectancy but not CS evaluation.

In 2010, a meta-analysis by Hofmann et al. (2010) rekindled this debate: their results indicated that EC is not, strictly speaking, resistant to extinction. They found a substantial reduction of EC in studies that assessed the EC effect both after acquisition and then again after extinction. CS-alone trials reduced the EC effect by 37% (i.e., from d = 0.85 to d = 0.53). This finding is difficult to reconcile with conventional dual-process theories of EC and, thus, motivated new research on the topic.

Gawronski et al. (2014), for example, suspected that extinction of EC may be dependent on characteristics of the study procedure and tested their hypothesis experimentally. They found no extinction when they compared the EC effect between different groups of participants who evaluated CSs only once, either after the acquisition or after the extinction procedure. EC was (partly) extinguished only when participants evaluated CSs twice—after the acquisition *and* after the extinction phase. Extinction was, however, only observed in explicit evaluative ratings but not in affective priming. Based on their results, Gawronski et al. (2014) argued that these changes in CS evaluation do not reflect genuine changes in liking. Instead, they argued that specific judgment-related nuisance processes (e.g., due to repeated CS evaluation) may be responsible for the artifactual extinction of EC in explicit evaluative ratings; the true underlying evaluative representation was assumed to be unaffected by the extinction procedure, as supported by the presumably less obtrusive evaluative priming measure. Gawronski et al. (2014) argued that their finding resolves the contradiction between the extinction effect found by Hofmann et al. (2010) and the resistance to extinction predicted by dual-process theories. They concluded that any theoretical account has to explain how EC is largely resistant to extinction.

Extinction learning is a prominent example of such an expectancy-liking dissociation. A similar dissociation has been reported in counterconditioning procedures, in which CSs are associated with USs of opposing valence in two subsequent parts of the learning procedure (Lipp & Purkis, 2006; Lipp et al., 2010): At the end of this two-part learning procedure, participants exhibited no EC effects, although they continued to expect CSs and USs to co-occur according to the regularity learned in the more recent second part. This dissociative pattern is the opposite of the one observed in extinction procedures, at the end of which participants (continue to) exhibit EC effects but no longer expect the CS to co-occur with USs. Dual-process theories explain this dissociation just as they explain the dissociation in extinction procedures: The associative learning process assumedly is driven by CS-US contiguity and, thus, the predictive value of CSs is irrelevant to their evaluation (Sweldens et al., 2014). Hence, no EC is to be expected because CSs are paired with positive as often as with negative USs.

3.2 The temporal integration hypothesis

Single-process theories cannot explain resistance to extinction, and particularly the dissociation between US expectancy and CS evaluation, by referring to distinctive properties of separate learning systems. Additional assumptions are necessary. Lipp et al. (2010) discussed a set of auxiliary assumptions for the single-process account, which we will refer to as the *temporal integration hypothesis*, to account for the expectancy-liking dissociation (also see Lipp et al., 2003). They argue that US expectancy and CS evaluations reflect different summaries of the same underlying representation. The assumption is that memory stores a unitary representation of the CS-US pairing, and that the learning history is conserved and organized along a temporal dimension or by contextual properties. Moreover, it is assumed that memory for CS-US pairings can be flexibly used to meet the (assumed) task demands. Lipp et al. (2010) argue that, by default, CSs are evaluated under consideration of the entire learning history—participants make *integrative* evaluative judgments. In contrast, predictions or judgments of US expectancy are made by default in reference to recent events—participants make *momentary* expectancy judgments. These opposite default judgment strategies are assumed to be afforded by the tasks. Thus, Lipp et al. (2010) proposed that the expectancy-liking dissociation is not indicative of two independent learning systems; instead, they propose that the dissociation is caused by different default judgment strategies underlying US expectancy versus CS evaluation responses.

The temporal integration hypothesis is inspired by a very similar idea proposed in the field of causal learning. Collins and Shanks (2002) found a dissociation between causal strength judgments and outcome prediction. Participants viewed a series of trials showing imaginary laboratory records that documented butterfly species' reactions to radiation exposure. Radiation caused genetic mutations in half of the butterflies and prevented mutations in the others. Akin to a counterconditioning procedure, these contingencies reversed in the middle of the experiment. Thus, across all trials there was no causal relationship between radiation and mutation for any butterfly species. Similar to US expectancy ratings in EC experiments, intermittent predictions about the occurrence of genetic mutations closely mirrored the changes in contingencies: Participants correctly predicted that radiation would first cause but later prevent mutations, or vice versa. But end-of-study causal strength ratings dissociated from participants' last predictions: In causal strength ratings participants favored neither cause nor prevention. Collins and Shanks (2002) further found that this dissociation was affected by the frequency of these ratings. When participants repeatedly rated causal strength throughout the learning procedure, their end-of-study ratings corresponded to their intermittent predictions.

To explain their findings, Collins and Shanks (2002) argued that participants can flexibly adopt different judgment strategies. For frequent judgments, a momentary strategy is adopted in which ratings reflect only the most recent information (i.e., that has been acquired since the last judgment). In contrast, when judgments are made only at the end of a series of events, an integrative strategy is adopted, in which ratings incorporate information from the entire event series. Rather than being dichotomous, these strategies can be thought of as smaller or larger averaging windows used to aggregate information across time. Matute et al. (2002) explored factors that cause participants to adopt a momentary or integrative judgment strategy. They found that questions targeting the predictive value of a stimulus induced momentary judgments, whereas questions about contiguity and causality induced integrative judgments. Moreover, they were able to manipulate the adopted judgment strategy via postexperimental instructions. In short, this research implies that participants flexibly use the learned information to meet the (assumed) demands of the task set by the experimenter.

Building on the research by Collins and Shanks (2002), Lipp and Purkis (2006) found that dissociations between US expectancy and CS evaluations are similarly affected by the frequency of evaluative ratings. In a counterconditioning and an extinction procedure, participants provided pleasantness ratings either twice (i.e., after each of the two learning phases) or only once at the end. When only one final rating was collected, participants' ratings reflected averages across the entire learning procedure: In the counterconditioning procedure, participants exhibited no EC effect; whereas in the extinction procedure, they exhibited a robust EC effect. In contrast, when participants provided multiple ratings, their CS evaluations reflected only the most recent CS-US contingencies, and the expectancy-liking dissociation was eliminated: In the counterconditioning procedure, participants reported causal relationships between CSs and USs in accord with the contingencies inherent in the respective part of the procedure, and they exhibited EC effects corresponding to these causal judgments. In the extinction procedure, participants reported no causal relationship after the extinction phase, and, correspondingly, they now also failed to exhibited an EC effect. In other words, the final expectations no longer dissociated from end-of-study evaluations—the EC effect was successfully extinguished. Notably, the extinguished EC effect reappeared when participants were asked to evaluate the CSs again in a different context and response format at the end of the study. Lipp et al. (2010) argued that their findings can be explained by the temporal integration hypothesis. They proposed that, when asked repeatedly throughout the learning procedure, participants made momentary judgments that reflected recent trends in CS-US contingencies (i.e., showed EC in counterconditioning but not after extinction). On the other hand, postexperimental pleasantness ratings in a different context and response format were by default integrative judgments that reflected the entire learning history (i.e, showed no EC in counterconditioning but did show EC after extinction).

The temporal integration hypothesis may reconcile expectancy-liking dissociations with single-process theories of EC, but the proposed auxiliary assumptions need to be tested rigorously. Previous research leaves room for alternative explanations of the extinction of EC, and it has not tested the effects of judgment strategies of US expectancy and CS evaluation concurrently. Here we address those shortcomings. We tested two predictions from the hypothesis' core assumptions more stringently and without allowing for alternative accounts in terms of demand effects induced by multiple pleasantness judgments (as proposed, e.g., by Gawronski et al., 2014). Remember that Lipp and Purkis (2006) elicited nondefault momentary CS pleasantness judgments by collecting ratings intermittently during the learning procedure. As argued by Gawronski et al., multiple intermittent CS evaluations could alter the evaluative learning process or bias response behavior by inducing demand characteristics and thereby artificially create momentary judgments. The present studies avoided this potential confound by collecting CS evaluations (as well as, in Experiment 3, expectancy judgments) only after the learning phase and thereby eliminate alternative explanations in terms of demand characteristics.

If previous findings are indeed caused by judgment strategies, then (1) it should be possible to manipulate these strategies for US expectancy and CS evaluation after the learning procedure and without intermittent CS pleasantness judgments. Moreover, if the default judgment strategies for CS pleasantness (Lipp & Purkis, 2006) and US expectancy (Collins & Shanks, 2002; Matute et al., 2002) are malleable, (2) the expectancy-liking dissociation in extinction learning should be reversible if one could elicit the opposite nondefault judgment strategies. A concurrent cross-over manipulation of judgment strategies for both US expectancy and CS pleasantness would predict a double-dissociation pattern. Combined in a single experiment, this constitutes a rigorous test of the temporal integration hypothesis. Confirmation of these double-dissociation predictions, while eliminating alternative accounts, would provide stronger support for the temporal integration hypothesis that goes well beyond that provided by previous studies.

3.3 MINERVA 2: A candidate single-process model

The temporal integration hypothesis does not specify how the learning history is conserved, how temporal organization is achieved, or how the information is summarized to perform judgment tasks. Mitchell et al. (2009) postulated that human associative learning is based on memory for past events. They further suggested that MINERVA 2 (Hintzman, 1988; Hintzman, 1984; Hintzman, 1986), a simple but popular model of episodic memory, may be the simplest model consistent with a memory system supporting their propositional single-process view of human associative learning (p. 187, Mitchell et al., 2009) (see also De Houwer, 1998; Klauer, 2009). In an attempt to fill in the blanks of the temporal integration hypothesis, we followed the suggestion by Mitchell et al. (2009) and adopted the memory architecture formalized in MINERVA 2. We explore the theoretical position that US expectancy and CS evaluation are memory-based judgments that rely on a unitary representation of CS-US pairing episodes. Using a formalized model enables us to make more specific predictions than current process theories of EC.

MINERVA 2 assumes that each CS-US pairing is stored as a trace in a unitary memory system. Episodes are encoded in a feature-based manner. Each memory trace consists of a series of slots, each of which indicates whether a feature is present (or absent) in a given episode. In the present application, subsets of these feature slots are dedicated to CS, US, and context features. When memory is probed (i.e., when a judgment is made), the stimulus and context features of the probe are simultaneously compared to all traces in memory. Each memory trace is activated according to its similarity to the memory probe. The recalled memory content is computed as a weighted average of all memory traces, where similar and strongly activated traces receive a larger weight than dissimilar and weakly activated traces. Hence, the recalled information is a mixture of all memory traces—rather than reflecting one specific past episode.

In line with current theorizing in memory research (e.g., Howard & Kahana, 2002; Zacks et al., 2007), we assume that the unitary memory system holds information about the (temporal) context of all stored episodes. MINERVA 2 is not equipped with a dedicated mechanism to impose a temporal structure on the stored episodes but such an organization can be achieved by assuming that a changing context is encoded in each episode. This conceptualization is consistent with mechanisms proposed in perceptual and memory research, where it is suggested that the continuous flow of information is automatically segmented and structured into discrete events (Zacks et al., 2007). Matute et al. (2011) have similarly invoked the concept of temporal contexts in research on associative learning. They found that participants spontaneously (i.e., without instructions) structure learning procedures by creating temporal contexts. Participants then used these contexts to retrieve associative information to guide their behavior and inform their prediction of future events. Thus, we assumed that the temporal organization of the learning history is retained via (perceived or internally generated) contexts that structure the incoming information into meaningful events. These assumptions allowed us to derive specific predictions for the learning procedures implemented in the present studies.

3.4 The present study

The overarching goal of this research was to test whether a single-process learning account can explain the expectancy-liking dissociation in EC. Building on the work by Collins and Shanks (2002) as well as Lipp and Purkis (2006), we tested the temporal integration hypothesis (Lipp et al., 2010; Lipp et al., 2003), which posits that US expectancy and CS pleasantness judgments are different summaries of a common underlying representation of CS-US pairings. We attempted to modify the default momentary and integrative judgment strategies for US expectancy and CS pleasantness judgments after completion of the learning procedure and without intermittent judgments. Moreover, we aimed at *reversing* the expectancy-liking dissociation in extinction learning by inducing nondefault integrative US expectancy and momentary CS pleasantness judgments.

We conducted one counterconditioning and two extinction experiments (see Table 1 for an overview).[1] For the counterconditioning procedure in Experiment 1, CS-US pairings were presented in two contexts: CSs were paired with positive USs in the first, and with negative USs in the second context, or vice versa. During the learning procedure, participants provided intermittent US expectancy ratings. After learning, participants judged CS pleasantness either without reference to learning contexts (to elicit default integrative judgments) or for a specific context (to elicit momentary judgments). Experiment 2 used an extinction procedure and presented half of the CSs together with USs in the first but alone in the second context. To hold the number of USs constant across contexts, the other half of the CSs was presented alone in the first but with USs in the second

[1] We ran three additional experiments as part of this project, which will be reported elsewhere. The data are available at https://github.com/methexp/rawdata.

context, thereby implementing a concurrent acquisition procedure. Experiment 3 replicated and extended Experiment 2: Participants provided no intermittent judgments but rated US expectancy only after completion of the learning procedure, either for both learning contexts together (to elicit integrative judgments) or for a specific context (to elicit default momentary judgments).

.

3.5 Experiment 1

In parallel with Collins and Shanks (2002) and Lipp and Purkis (2006), we first tested the temporal integration hypothesis and our memory-based judgment simulation of EC with the expectancy-liking dissociation in a counterconditioning procedure. If the assumptions of temporal integration hypothesis hold, single-process theories of EC can account for this dissociation by assuming that end-of-study CS evaluations are integrative judgments, and accordingly, no EC effect is to be expected because the effects of positive and negative CS-US pairings cancel each other out.

We first designed a simulation of a simplified counterconditioning procedure to generate more specific predictions using MINERVA 2 (for details see Appendix A.1). One CS was first paired with a positive and then with a negative US, conversely, a second CS was first paired with a negative and then with a positive US; a third CS was paired with a neutral US. Moreover, we simulated context changes in between the first and second phase of the learning procedure as well as prior to end-of-study CS pleasantness ratings. Thus, we assumed participants would experience the end-of-study rating procedure as different from the learning procedure. To predict US expectancy and CS pleasantness ratings, we reasoned that the CS in question and the current context act as cues to recall previous pairings with USs. If the recalled memory content was positive we predicted an expectation of a positive USs and a positive CS evaluation. We, thus, predicted US expectancy and CS pleasantness ratings based on the same information.

Our simulation predicted a pattern of results consistent with the temporal integration hypothesis, Figure 3.1A. During the learning procedure, the valence of the recalled memory content closely followed the CS-US contingencies. The recalled memory contents acquired the USs' valence but due to the context change the CS-US pairings in the counterconditioning phase quickly reversed the contents' valence. Thus, for the last trial the simulation predicted expectation of the US that had been paired with a given CS in the second context.

More importantly, the same pattern was predicted for end-of-study judgments when the learning contexts were reinstated. For example, when a CS that had first been paired with a positive and then with a negative US was presented in the first context, the valence of the retrieved memory contents was positive. However, when the same CS was presented in the second context, the recalled information was negative. The reinstated context features promoted the activation of memory traces of episodes from the respective context. This contextualized retrieval of CS-US pairings assumedly underlies momentary judgments of US expectancy and CS pleasantness. For the new context—when no learning context was reinstated—our simulation predicted that episodes from both contexts contributed equally to the retrieved memory contents. Positive and negative CS-US pairings effectively cancelled each other out. Thus, the simulation predicted no EC effect in default integrative end-of-study pleasantness ratings.

To conclude, in line with the temporal integration hypothesis, the simulation of the counterconditioning procedure predicted momentary judgments in intermittent US expectancy ratings and both momentary and integrative judgments in end-of-study CS pleasantness ratings, depending on context cues. Hence, our single-process memory model simulation produced an expectancy-liking dissociation, which has been taken as evidence for dual-process theories of EC: Marked US expectancies in the last trial but no EC effect in end-of-study CS pleasantness ratings for the new context. It, nonetheless, predicted EC effects in momentary end-of-study CS pleasantness ratings when learning contexts are reinstated. Therefore, no expectancy-liking dissociation is expected when comparing momentary US expectancy to momentary CS pleasantness ratings.

We designed an experiment to test these predictions. We conducted a counterconditioning experiment with intermittent US expectancy and end-of-study CS pleasantness ratings in different contexts. We showed participants a stream of pictures in which CSs were first paired with positive and later with negative USs, or vice versa. In contrast to Lipp and Purkis (2006) we asked participants to evaluate CSs only after completion of (rather than repeatedly during) the learning procedure. This procedural change ruled out that intermittent CS pleasantness judgments affected the evaluative learning process and artificially induced subsequent momentary judgments (e.g., via conversational logic demands). Participants provided end-of-study CS pleasantness ratings without reference to learning contexts (to elicit default integrative judgments) and for each of the learning contexts (to elicit nondefault momentary judgments). We expected (1) to observe the predicted expectancy-liking dissociation between intermittent US expectancy ratings in the last trial on the one hand and integrative end-of-study CS pleasantness ratings on the other hand, but (2) to eliminate the expectancy-liking dissociation by demonstrating EC effects that mirror US expectancy ratings in momentary end-of-study CS pleasantness ratings.

3.5.1 Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. The simulation code, experimental software[2] and materials, data, and analysis scripts[3] are available at https://osf.io/vnmby/.

3.5.1.1 Participants

We recruited 40 participants from our lab participant database via e-mail for this experiment. Eligible volunteers were 18-60 years old, fluent in German and (according to our database) had not participated in any studies on evaluative conditioning for at least one year. Participants who aborted the experiment were not included in the analyses. The sample size was determined informally based on previous experience with EC experiments. The data of three participants were lost due to a technical error leaving the data of 37 participants for analysis. Participants' mean age was 23.69 years (*SD* = 6.50), 26 were female, 11 studied psychology or media psychology, all participants declared intact color vision, and 9 reported to have had prior

[2]We created all experiments in OpenSesame (Mathôt et al., 2012).

[3]We used R (Version 3.6.3; R Core Team, 2017) and the R-packages *afex* (Version 0.23.0; Singmann et al., 2017), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *emmeans* (Version 1.3.5; Lenth, 2018), and *papaja* (Version 0.1.0.9997; Aust & Barth, 2017) for all analyses and reporting.

knowledge about the CS pictures. We compensated participants with 8€ or course credit.

3.5.1.2 Apparatus and material

We conducted the experiment in five dimly lit and sound-attenuated booths and presented all stimuli on a 17" CRT-monitor.

Because a seemingly random stimulus sequence, a large proportion of filler stimuli, and a low proportion of valent stimuli, have been reported to be conducive to (associative) EC (Jones et al., 2009), our learning procedure consisted of a mixture of critical CS-US pairings and irrelevant filler trials. Critical CS-US pairs consisted of 12 neutral cartoon characters taken from Stahl and Heycke (2016) as CSs and 12 positive or 12 negative low-arousal IAPS pictures as USs (Lang et al., 2008, Table A.1). All positive USs were pictures of animals; all negative USs were pictures of humans. We introduced the confound between US valence and category because we wanted to rule out that intermittent US expectancy judgments affected the evaluative learning process—the confound enabled us to assess US expectancy without referring to US valence.

The filler trials consisted of six neutral CS-US pairs, three CS-CS pairs, three individual CSs, three US-US pairs, as well as three individual USs of intermixed valence, and six blank screens. For filler CSs, we used additional cartoon characters (Stahl & Heycke, 2016) and for filler USs, we used IAPS pictures from two additional US categories. All neutral USs depicted household items, and the intermixed USs depicted natural scenes (see Table A.1). The filler stimuli were included to make contingency learning more demanding, and to obscure the confound between US valence and categories and thereby to further mitigate possible effects of intermittent ratings on evaluative learning.

US expectancy has previously been assessed with predictive ratings (e.g. of the extent to which participants expected a US following the presentation of a CS, p. 224 Hermans et al., 2002; also see Vansteenwegen et al., 2006) or causal questions ("To which extent (0–100%) does [the CS] cause the [US] to appear?", Lipp et al., 2010)(also see Collins & Shanks, 2002; Lipp & Purkis, 2006). In the context of contingency learning, Matute et al. (2002) found that predictive and causal questions elicit comparable integrative judgments but that predictive questions more effectively elicit momentary judgments. Hence, we employed a predictive question: "Next time this creature is presented, what type of picture will it be shown with?" Participants provided probability estimates for animal, human, and object on an eleven-point scale ranging from 0% to 100%.

We collected CS pleasantness ratings on an 19-point scale ranging form *very unpleasant* to *very pleasant*. To assess memory for CS-US pairs, we separately tested recognition memory for US category and US identity for each CS. For US category recognition, we presented individual CSs and participants selected a US category in an 3-alternative forced-choice (3-AFC) task (e.g., "animal", "human", or "object"). For US identity recognition, participants selected one US out of all USs from the correct US category in an 12-AFC. We also performed a funnel debriefing to assess the extent to which participants were aware of the purpose of the study and the hypotheses. The debriefing served to inform the design of future incidental-learning studies; they are irrelevant to the present hypotheses.

3.5.1.3 Procedure and design

After obtaining informed consent, participants filled in demographic information about gender, age, handedness, field of study, and visual impairments. We then instructed participants that we would present a stream of pictures in 2 × 3 blocks and asked them to attend the stream carefully, to detect regularities, and to memorize repeating pairs of pictures (for similar instructions see e.g., Kattner & Green, 2015; Moran & Bar-Anan, 2013b; Richter & Gast, 2017; Zanon et al., 2012). We warned that, during the course of the study, we would test whether they had continuously attended the stream. To distract from the contingency between CS and US valence, we pretended that we were interested in participants' vigilance while they monitored images from surveillance cameras.

Table 3.1: Illustration of our learning procedures in Experiments 1-3

	Learning procedure		Intermittent ratings	End-of-study ratings	
	First context ($\mathscr{A}$)	Second context ($\mathscr{B}$)	US expectancy	CS pleasantness	US expectancy
Experiment 1					
Counterconditioning	CS_1 US_+	CS_1 US_-	CS_1 ($\mathscr{A}$,$\mathscr{B}$)	CS_1 ($\mathscr{C}$,$\mathscr{B}$,$\mathscr{A}$)	–
	CS_2 US_-	CS_2 US_+	CS_2 ($\mathscr{A}$,$\mathscr{B}$)	CS_2 ($\mathscr{C}$,$\mathscr{B}$,$\mathscr{A}$)	–
Experiment 2					
Acquisition	CS_1	CS_1 US_+	CS_1 ($\mathscr{A}$,$\mathscr{B}$)	CS_1 ($\mathscr{C}$,$\mathscr{B}$,$\mathscr{A}$)	–
	CS_2	CS_2 US_-	CS_2 ($\mathscr{A}$,$\mathscr{B}$)	CS_2 ($\mathscr{C}$,$\mathscr{B}$,$\mathscr{A}$)	–
Extinction	CS_3 US_+	CS_3	CS_3 ($\mathscr{A}$,$\mathscr{B}$)	CS_3 ($\mathscr{C}$,$\mathscr{B}$,$\mathscr{A}$)	–
	CS_4 US_-	CS_4	CS_4 ($\mathscr{A}$,$\mathscr{B}$)	CS_4 ($\mathscr{C}$,$\mathscr{B}$,$\mathscr{A}$)	–
Experiment 3					
Acquisition	CS_1	CS_1 US_+	–	CS_1 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)	CS_1 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)
	CS_2	CS_2 US_-	–	CS_2 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)	CS_2 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)
Extinction	CS_3 US_+	CS_3	–	CS_3 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)	CS_3 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)
	CS_4 US_-	CS_4	–	CS_4 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)	CS_4 ($\mathscr{A} \parallel \mathscr{B} \parallel \mathscr{C}$)

Note. Calligraphic font denotes context features; mapping of features to context was counterbalanced in all experiments. In Experiments 1 and 2, participants provided end-of-study ratings for every context (within-subject manipulation) but for only one randomly selected context in Experiment 3 (between-subject manipulation). CS = Conditioned stimulus; US = Unconditioned stimulus

The conditioning procedure consisted of two phases, Table 3.1. In the initial acquisition phase, we paired 6 critical CSs with positive and the remaining 6 critical CSs with negative USs. In the subsequent counterconditioning phase, critical CSs were paired with USs of the opposite valence. CSs were randomly assigned to one of the two US valence orders. Filler CSs that were paired with neutral USs in the first phase were paired with new neutral USs in the second phase.

We created different contexts for the first and second phase to standardize participants' temporal organization of the learning history (Matute et al., 2011) and facilitate later reference to each phase in targeted questions about particular portions of the learning procedure. The context features—background color and CS position—were randomly assigned to the first or second phase for each participant. The background color of the screen was either yellow or blue; CSs were presented either on the left or right side of the screen, with USs on the opposite side.

Both phases consisted of three subblocks, interrupted by self-paced breaks. In each of the subblocks, we presented all critical and filler trials three times. The stimulus sequence entailed no immediate stimulus repetitions but was otherwise random. In each trial, CSs were presented alone for 500 ms and then jointly with USs for another 1000 ms. Each CS was paired with only one US to facilitate accurate memory for pairings. For CS-CS or US-US pairs, one stimulus was randomly chosen to act as CSs; the second stimulus acted as US. There was no delay between trials (Jones et al., 2009). The conditioning procedure consisted of 648 trials (216 with critical CSs), and lasted approximately 20 minutes.

During the learning procedure, we intermittently presented CSs and asked participants to report their current US expectancy: "With what probability would you expect a photograph of a human [animal/object] with this creature?" In each subblock, we randomly selected six of the 18 CS-US pairs (including neutral pairs). Participants made US expectancy judgments on a random trial following the third and final presentation of the selected CS-US pair in each subblock. (i.e., ratings in the first subblock reflected participants' expectations after three CS-US pairings, ratings in the second subblock reflected participants' expectations after six, etc.). In the subblocks of the subsequent counterconditioning phase, we used the same CS-US pair selection as in the acquisition phase. For example, if we selected a CS-US pair for US expectancy ratings in the first subblock of the acquisition phase, we selected the same pair in the first subblock of the counterconditioning phase. Thus, participants reported their US expectancy twice for every CS-US pair, and three subblocks elapsed before the second rating. Each participant provided 36 US expectancy ratings, yielding 3 ratings per experimental condition.

Following the learning procedure, participants provided pleasantness ratings for each CS. Akin to the postexperimental rating condition by Lipp and Purkis (2006), we collected a first rating in a new context. In this new context, we presented CSs in the center of the screen on a black background and asked "How pleasant or unpleasant do you find this creature[, currently]?" That is, there was no reference to learning contexts. We then collected CS pleasantness ratings for the context of the second and then of the first phase. We reinstated the respective context features (background color and CS position) and asked participants "How pleasant did you find this creature during the second [first] half?" Each participant provided 54 pleasantness ratings (including CSs from neutral CS-US pairs), yielding 3 ratings per experimental condition.

Next, we assessed participants' memory for CS-US pairs. We tested pairing memory for the second and then for the first phase. The order served to minimize memory interference and because pairing memory for the counterconditioning phase was of particular interest. After every response, we immediately tested participants' US identity recognition for the same CS-US pair. We probed memory for CS-US pairs in a new random order for each participant and context. Each participant provided 36 US category and US identity recognition responses (including neutral CS-US pairs), yielding 6 responses per experimental condition.

Finally, we administered the funnel debriefing, participants rated the pleasantness of each US category (human, animal, and object) from memory (i.e., the USs were not presented again), and indicated whether they had previously been familiar with the cartoon characters. On average, participants took 56.53 minutes ($SD = 18.02$) to complete the study.

Due to an error in the randomization procedure, we used the same assignment of CSs to US valence orders for all participants, with two consequences. First, the CSs assigned to the US valence orders systematically differed in pleasantness: CSs that were first paired with negative and later with positive USs were more pleasant a priori ($M = 5.52, SD = 0.92$) than CSs first paired with positive and later with negative USs ($M = 4.38, SD = 1.00$).

This confound is unlikely to endanger our conclusions because it works against our predictions of an EC effect in the acquisition context and the absence of an EC effect in the new context. Our remaining predictions largely concern changes in CS pleasantness across contexts within each set of CS-US pairs, for which the confound is irrelevant. Second, CSs were paired with a random US in the acquisition phase, but in the counterconditioning phase some specific CS-US pairs were more likely than others. Mean US pleasantness and arousal were however comparable across conditions and closely matched the means of all USs of the corresponding category (Table A.2). In sum, the error in the randomization procedure is vexing and subpar but unlikely to affect our results or conclusions.

3.5.2 Data analysis

For all analyses, we averaged participants' responses across items. We combined US expectancy ratings for each of the three US categories into a single measure of expectancy of the correct US by subtracting the ratings for incorrect categories from those for the correct category. For example, for CSs paired with pictures of objects, we calculated a US expectancy score $\bar{x}_{\text{expectancy}} = \bar{x}_{\text{object}} - (\bar{x}_{\text{human}} + \bar{x}_{\text{animal}})$ for every participant in every cell of the experimental design.

We performed ANOVAs and base our inference on p values and 95% confidence intervals as well as Bayes factors. For the frequentist analyses we always report Greenhouse-Geisser corrected degrees of freedom. For planned contrasts and post hoc comparisons, we compared least squares means (Lenth, 2018). To infer equivalence between two condition means, we performed two one-sided t tests (TOST; Lakens, 2017; Rogers et al., 1993; Wellek, 2002). In the TOST procedure the analyst defines a region of equivalence around the null value. She compares the mean difference to the upper and lower bound of this region of equivalence using one-sided tests or a 90% confidence interval. The means are deemed equivalent if the difference between them is significantly larger than the lower bound and significantly smaller than the upper bound. For reasons of brevity only the result of the test that yields the larger p value is reported. Thus, in case of a significant TOST the analyst rejects the hypothesis that an effect is of a given size or larger. We adopted symmetric equivalence regions in units of standardized mean differences for within-participant comparisons of

$\Delta \pm 0.3 d_r$ to reject small effects (Lakens, 2017). The α-level for all frequentist analyses was .05; p values were corrected for multiple comparisons where applicable.

For Bayesian ANOVAs we used default multivariate Cauchy priors with a scaling parameter of $r = 0.5$ on the fixed effects (Rouder et al., 2012); for Bayesian t tests we used a default Cauchy prior with a scaling parameter of $r = \sqrt{2}/2$ on the effect size d_z (Rouder et al., 2009). All Bayes factors were estimated to a precision of ±5%. Bayes factors quantify the evidence for an effect relative to the null hypothesis of no effect in the data at hand (e.g, Wagenmakers et al., 2010). We use BF_{10} to denote evidence for an effect relative to the null hypothesis of no effect and BF_{01} to denote evidence for the null hypothesis of no effect relative to an effect. For example, $\mathrm{BF}_{10} > 1$ is evidence for the presence of an effect, whereas $\mathrm{BF}_{01} > 1$ is evidence for the absence of an effect. Bayes factors are readily interpretable as a graded measure of evidence. We will, however, follow the suggestion by Kass and Raftery (1995) to consider $1/3 < \mathrm{BF} < 3$ "not worth more than a bare mention" (p. 777). We do not report our prior beliefs in the hypotheses described here (prior odds; for discussion see Rouder et al., 2012). The interested reader may form their own prior beliefs and use the reported Bayes factors to determine their posterior belief in the hypotheses.

3.5.3 Results

In the following, we focus on the results for US expectancy and CS pleasantness ratings. See Appendix A.2 for analyses of participants' CS-US pairing memory.

3.5.3.1 US expectancy

We analyzed expectancies of the correct US using a 2 (*US valence order*: US+ US− vs. US− US+) × 2 (*Context*: First vs. Second) × 3 (*Pairings*: 3 vs. 6. vs. 9) repeated-measures ANOVA. To facilitate the comparisons between predicted and observed US expectancy as well as between US expectancy and CS pleasantness, Figure 3.1B depicts a difference score between expectancies of positive and negative US.

As expected, participants quickly learned the CS-US contingencies. The number of repetitions of CS-US pairings affected expectancy of the correct

US, $F(1.59, 57.20) = 25.45$, $MSE = 0.22$, $p < .001$, $\hat{\eta}^2_G = .073$, $\mathrm{BF}_{10} = 2.31 \times 10^8$. Follow-up tests indicated that expectancy of the correct US increased from three to six CS-US pairings, $\Delta M = .27$, 95% CI [.19, ∞], $t(72) = 5.56$, $p < .001$, $\mathrm{BF}_{10} = 4.14 \times 10^5$ (one-tailed), but remained unchanged from six to nine CS-US pairings, $\Delta M = .05$, 90% CI [-.04, .15], $t(72) = -2.07$, $p = .042$ (equivalence test adjusted for two comparisons), $\mathrm{BF}_{01} = 2.49$. There was weak evidence that our experimental manipulations had no other effects, all $p \geq .145$, all $\mathrm{BF}_{01} \geq 3.89$. To conclude, participants' expectancy for the correct US built up during and reached a plateau toward the end of each learning phase. At the end of the experiment participants expected CSs to be accompanied by the US that they had last been paired with.

3.5.3.2 CS pleasantness

We analyzed CS pleasantness ratings using a 2 (*US valence order*: US+ US− vs. US− US+) × 3 (*Referenced context*: First vs. Second vs. None) repeated-measures ANOVA.

As predicted, referring to and reinstating learning contexts affected CS pleasantness ratings differently depending on US valence order, $F(1.18, 42.31) = 17.63$, $MSE = 10.32$, $p < .001$, $\hat{\eta}^2_G = .083$, $\mathrm{BF}_{10} = 2.19 \times 10^6$, Figure 3.1B. Follow-up tests provided some evidence that in the new context participants made comparable CS pleasantness ratings for both US valence orders, $\Delta M = -0.12$, 90% CI [-1.15, 0.91], $t(103.46) = -1.41$, $p = .081$ (equivalence test), $\mathrm{BF}_{01} = 5.52$. When we compared participants' ratings for the first and second context, we observed both the predicted increase in perceived pleasantness for CSs that were first paired with negative and then with positive USs, $\Delta M = 2.41$, 95% CI [1.64, ∞], $t(115.91) = 5.15$, $p < .001$, $\mathrm{BF}_{10} = 3.50 \times 10^4$ (one-tailed), and the predicted decrease for CSs that were first paired with positive and then with negative USs, $\Delta M = 2.39$, 95% CI [1.61, ∞], $t(115.91) = 5.10$, $p < .001$, $\mathrm{BF}_{10} = 5.90 \times 10^3$ (one-tailed). Moreover, we found an EC effect for the first context, $\Delta M = 2.19$, 95% CI [1.16, ∞], $t(103.46) = 3.53$, $p < .001$, $\mathrm{BF}_{10} = 31.21$ (one-tailed), and a reversed EC effect for the second context, $\Delta M = 2.62$, 95% CI [1.59, ∞], $t(103.46) = 4.22$, $p < .001$, $\mathrm{BF}_{10} = 153.33$ (one-tailed). Participants' prior knowledge about CSs did not affect these results, $F(1.17, 41.10) = 0.03$, $MSE = 10.61$, $p = .898$, $\hat{\eta}^2_G = .000$, $\mathrm{BF}_{01} = 5.51$. Thus, although we observed no EC effect when we asked participants to report CS pleasantness in a

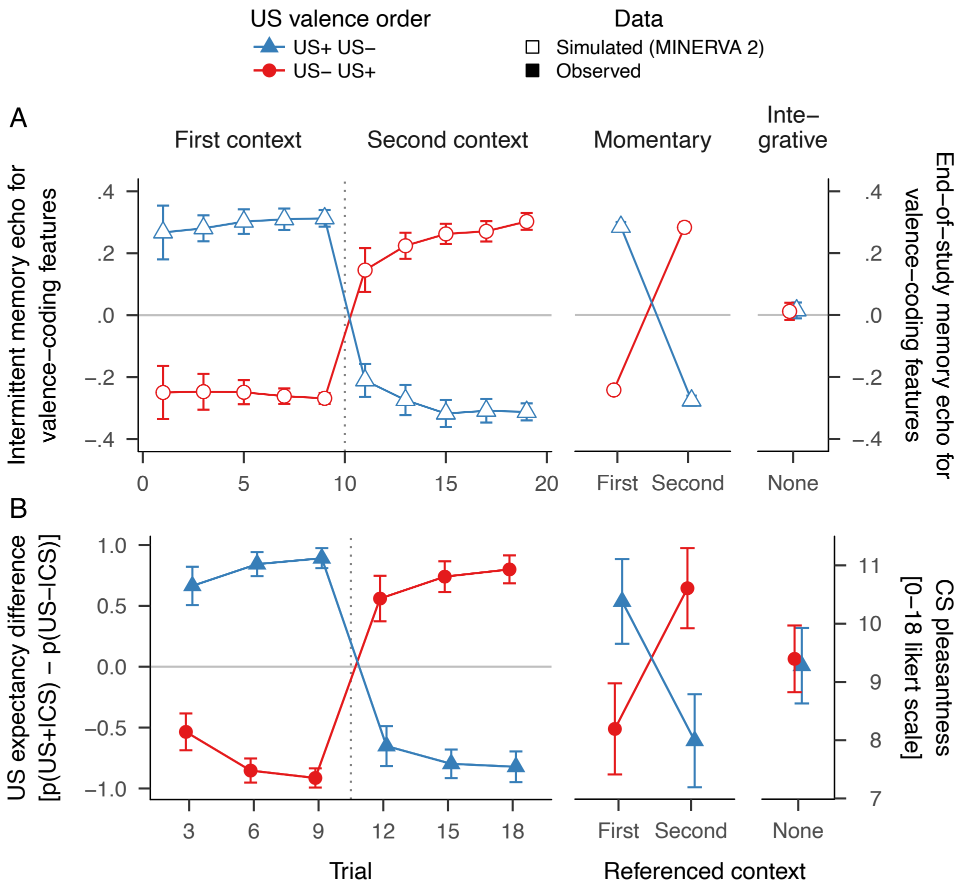

Figure 3.1: Simulated and observed US expectancy and CS pleasantness ratings for Experiment 1. Blue triangles indicate CSs paired with positive USs in the first and negative USs in the second context; red circles indicate CSs paired with negative USs in the first and with positive USs in the second context. **A** Mean normalized memory echo of valence-coding features predicted by MINERVA 2 indicative of the overall valence of the retrieved memory contents. The left plot shows valence retrieved during the learning procedure, the right plot shows the valence retrieved after completion of the learning procedure. Error bars represent 95% confidence intervals. **B** The left plot shows observed differences in mean US expectancy during the learning procedure. Positive values indicate expectancy for positive USs, negative values indicate expectancy for negative USs. The right plot shows observed mean CS pleasantness ratings after completion of the learning procedure. Error bars represent 95% within-subject confidence intervals; CS = Conditioned stimulus, US = Unconditioned stimulus.

new context at the end of the experiment, referring to and reinstating the learning contexts revealed changes in CS pleasantness throughout the learning procedure.

3.5.4 Discussion

The results of our counterconditioning experiment confirm the predictions derived from the temporal integration hypothesis and our simulation. First, we found the predicted expectancy-liking dissociation: Participants reported marked US expectancies throughout and, critically, at the end of the learning procedure—they expected CSs to appear with the most recently paired USs. In contrast, when participants provided CS pleasantness judgments immediately after the completion of the learning procedure and without reference to learning contexts, we found no EC effect. Second, participants made the predicted contextualized CS pleasantness judgments: We observed an EC effect for the initial acquisition context and a reversed EC effect for the counterconditioning context. These momentary CS pleasantness judgments reflected the changes in CS-US contingencies and corresponded to the intermittent US expectancy ratings. Hence, eliciting nondefault momentary evaluative judgments eliminated the expectancy-liking dissociation.

The temporal integration hypothesis posits that repeated judgments affect the adopted judgment strategy but do not affect evaluative learning. Research on contingency learning has shown that nondefault integrative contingency judgments can be elicited after completion of the learning procedure (Collins & Shanks, 2002; Matute et al., 2002). Our findings extend these conclusions to CS pleasantness judgments. In this experiment, participants rated CS pleasantness only after completion of the learning procedure—they made no CS pleasantness judgments during the learning procedure. This approach is an improvement over previous studies in which CS pleasantness was assessed repeatedly during the learning procedure (e.g., Blechert et al., 2008; Lipp & Purkis, 2006; Lipp et al., 2010) because it precludes that intermittent CS pleasantness judgments affected the evaluative learning process.

While our findings corroborate the dissociability of US expectancy and CS liking, they raise questions about the common dual-process interpretation of the expectancy-liking dissociation. The finding that US expectancy extinguishes while EC is resistant to extinction is commonly interpreted as evidence for a second associative learning process . In contrast, MINERVA 2 instantiates a candidate process-model of the single-process learning account (Mitchell et al., 2009). Drawing on the additional assumptions proposed by the temporal integration hypothesis (Lipp et al., 2010), the simulation illustrates that MINERVA 2 can predict the observed expectancy-liking dissociation in counterconditioning. Hence, absence of EC effects despite US expectancy can be explained by a single learning process.

Taken together, our findings support the assumptions of the temporal integration hypothesis that EC yields a single representation of CS-US pairings that informs both US expectancy and CS liking, and that their dissociation is caused by different default judgment strategies.

3.6 Experiment 2

The expectancy-liking dissociation reported in extinction procedures (e.g., Lipp & Purkis, 2006; Hermans et al., 2002) is the reverse of the dissociative pattern in the counterconditioning procedure: At the end of the learning procedure, participants no longer express US expectancies, but still exhibit an EC effect. As in Experiment 1, our reasoning was that inducing nondefault momentary judgments of CS pleasantness, by referring to and reinstating the learning contexts, would reveal extinction of EC effects.

We again began by simulating a simplified acquisition and an extinction procedure using MINERVA 2. The simulation method and assumptions were the same as for Experiment 1. In the extinction procedure, we paired CSs with USs in the first but presented them alone in the second context, Table 3.1. Conversely, in the acquisition procedure, we presented CSs alone in the first, and subsequently paired them with USs in the second context (see De Houwer et al., 2000, for a similar approach). The CS-alone trials in the first context served to equate the number of CS-US pairings and valenced stimuli in the two contexts and, thus, to avoid attentional disengagement or mood effects.

The simulation predicted a pattern of results in line with the temporal integration hypothesis (Figure 3.2A): During the learning procedure, the valence of the retrieved memory contents closely followed the CS-US contingencies. In the acquisition procedure, the recalled memory contents remained neutral during the initial CS-alone trials but quickly acquired the USs' valence during the subsequent CS-US pairing trials. Conversely, in the extinction procedure, the recalled memory contents acquired the USs' valence during CS-US pairing trials but quickly returned to a neutral baseline as a result of the combined context change and CS-alone trials. Thus, for the last trial of the acquisition procedure the simulation predicted expectations of the USs that had last been paired with CSs; in the last trial of the extinction procedure the simulation predicted the absence of a US expectancies.

The same pattern was predicted for end-of-study pleasantness judgments when the learning contexts were reinstated. For example, when a CS in the extinction procedure was presented in the first context—the context in which it was paired with a positive US—the valence of the retrieved memory contents was positive. However, when the same CS was presented in the second context—the context of CS-alone trials—the valence of the retrieved memory contents was neutral. In the new context—when no learning context was reinstated—the valence of the retrieved memory contents was comparable for CSs in the acquisition and extinction procedure. Furthermore, in both procedures the valence of the retrieved memory contents in the new context was comparable to that for the CS-US pairing context[4]. Hence, the simulation predicted comparable EC effects for default integrative end-of-study pleasantness ratings in the acquisition and extinction procedures. It also predicted comparable EC effects for momentary end-of-study pleasantness ratings in the CS-US pairing context and integrative ratings in the new context. Note that these predictions pertain to the current experimental design. As illustrated by the faint symbols in Figure 3.2A, the predictions differ for a design without a concurrent acquisition procedure or neutral CS-US pairs. We will return to this point in the General discussion.

[4]Subsequent exploration identified two procedural factors that, in conjunction with the similarity-based retrieval mechanism of MINERVA 2, contribute to this prediction: (1) In the CS-US pairing context, the valence of the retrieved memory contents decreases as the number of neutral stimulus presentations in that context increases. This is because the retrieval cue for the CS-US pairing trials encompasses context features and, thus, to some degree activates all memory trace from that context (interference). (2) Conversely in the new context, the valence of the retrieved memory contents increases relative to the CS-US pairing context because the interference from neutral stimulus presentations is decreased. The attenuated interference is a consequence of the unique features of the new context, the nonlinear relationship between probe-trace similarity and trace activation, as well as the normalization of the echo contents.

In sum, the simulation predicted momentary judgments in intermittent US expectancy ratings; the momentary or integrative nature of judgments in end-of-study CS pleasantness ratings depended on the choice of context cues. Thus, our single-process memory model simulation predicted the well known expectancy-liking dissociation in extinction: No US expectancy in the last trial of the learning procedure but an EC effect in end-of-study CS pleasantness ratings without reference to the learning contexts. It furthermore predicted that extinction of EC could nonetheless be demonstrated in momentary end-of-study CS pleasantness ratings by referencing and reinstating the context of CS-alone trials. Hence, no expectancy-liking dissociation is expected when comparing momentary US expectancy to momentary CS pleasantness ratings. These predictions are in line with the explanation of the expectancy-liking dissociation proposed by the temporal integration hypothesis.

Based on the methodology of Experiment 1, we designed an experiment to test these predictions. We showed participants a stream of pictures in which CSs were either presented alone and subsequently paired with valent USs (acquisition procedure) or, conversely, paired with USs and subsequently presented alone (extinction procedure; see Table 3.1). We expected (1) to observe the predicted expectancy-liking dissociation between intermittent US expectancy ratings in the last trial of the learning procedure and end-of-study CS pleasantness ratings without reference to the learning contexts, but (2) to eliminate the expectancy-liking dissociation by demonstrating extinction of EC in momentary end-of-study CS pleasantness ratings when learning contexts are referenced and reinstated.

3.6.1 Methods

The experimental method, data analysis plan, and the following hypotheses were preregistered (https://osf.io/vnmby/registrations/): In the extinction procedure, we predicted (1) the expectancy for the US that had been paired with a given CS to extinguish towards the last trial, whereas we predicted (2) an EC effect in end-of-study CSs pleasantness ratings in the new context (i.e. resistance to extinction in integrative CS pleasantness judgments). Furthermore, we predicted a reduced EC effect for end-of-study CS pleasantness ratings in the context of CS-alone trials, compared to (3) the new context and (4) the context of CS-US pairing trials. We, additionally, wanted to test whether (5) EC effects in the new context were comparable to those for the CS-US pairing context. The last hypothesis was introduced to investigate the meta-analytical finding that the extinction of EC in default integrative end-of-study judgments exists but is small (Hofmann et al., 2010). Experimental design, materials, and procedure followed those of Experiment 1 except for the following changes.

3.6.1.1 Participants

To maximize the efficiency and informativeness of our study, we performed a sequential Bayesian analysis while the data were being collected (Rouder, 2014). Thus, the number of participants was not fixed a priori. We set a minimum sample size of $n = 20$ (Schönbrodt et al., 2015) and planned to collect data until they provided strong evidence ($BF_{10} > 10$ or $BF_{01} > 10$) for or against our five hypotheses of primary interest or until we ran out of money (800€). We calculate Bayes factors at the end of each day of data collection.

We recruited 55 new participants. As preregistered we excluded one participant who performed the category recognition task at chance level, that is, they responded correctly to 25% or less of all category recognition questions. We assumed that at-chance category recognition is indicative of inattention because we instructed participants specifically to attend to pairings and detect regularities. We excluded one additional participant with a severe vision impairment, who was allowed to participate to obtain course credit. Thus, we stopped collecting data after 53 valid participants. At this point the data provided strong evidence for hypotheses 1-4. We deviated from our preregistered sampling plan and stopped data collection before the data provided an informative test of hypothesis 5 because it was not relevant to our theoretical predictions. The results of our Bayesian analysis are not affected by the premature termination of the data collection (Rouder, 2014). Participants' mean age was 22.24 years ($SD = 3.34$), 39 were female, 8 studied psychology or media psychology, all participants declared intact color vision, and 19 reported to have prior knowledge about the CS pictures.

3.6.1.2 Material

We adapted the 3-AFC US category recognition response format to the extinction procedure: Because the procedure included CS-alone trials, we added an additional "nothing" response option by which participants could indicate that a CS had not been paired with a US.

3.6.1.3 Procedure and design

For each participant we randomly assigned a positive, neutral, or negative US to each CS; the CS-US pair was then randomly assigned to the acquisition or extinction procedure. CSs in the acquisition procedure were presented alone in the first context and paired with USs in the second context. Conversely, in the extinction procedure, CSs were paired with USs in the first context and presented alone in the second context.

Instructions and assessment of our dependent measures were the same as in Experiment 1. However, we did not assess US identity recognition if CSs had been presented alone in a given context. Because each CS was paired with a US in only one of the two contexts, participants provided 18 US identity recognition responses, yielding 3 per experimental condition.

On average, participants took 51.15 minutes (SD = 7.04) to complete the study.

3.6.2 Results

Preregistered analyses (labeled confirmatory) will be followed by additional (exploratory) analyses. See Appendix A.2 for analyses of participants' CS-US pairing memory.

3.6.2.1 US expectancy

3.6.2.1.1 Confirmatory analyses We analyzed expectancies of the correct US using a 3 (*Valence*: Positive vs. Neutral vs. Negative) × 2 (*Learning procedure*: Acquisition vs. Extinction) × 2 (*Context*: First vs. Second) × 3 (*Pairings*: 3 vs. 6. vs. 9) repeated-measures ANOVA. As in Experiment 1, participants quickly learned the CS-US contingencies, Figure 3.2B. As predicted, we found strong evidence that changes in expectancy of the correct US category across referenced contexts differed between acquisition and extinction procedures, $BF_{10} = 2.85 \times 10^{182}$. We observed this pattern irrespective of the valence of the US, $BF_{01} = 23.05$. Planned contrasts indicated that, averaged across US valences, expectancy for the correct US category after the ninth pairing increased from the first to the second context in the acquisition procedure ($M = 0.85$ 95% HDI [0.76, 0.94], $BF_{10} = 7.29 \times 10^{21}$, one-tailed) but decreased in the extinction procedure, $M = -0.67$ 95% HDI [−0.80, −0.55], $BF_{10} = 5.08 \times 10^{12}$, one-tailed. The data provided no noteworthy evidence as to whether participants had a residual expectancy for the correct US category at the end of the extinction procedure, $M = 0.06$ 95% HDI [0.00, 0.12], $BF_{01} = 1.14$, one-tailed. In sum, following CS-US pairings participants expected the correct US but US expectancy declined rapidly when CSs were subsequently presented alone.

Additionally, learning of CS-US contingencies proceeded faster in the second than the first context ($BF_{10} = 27.50$), regardless of US category valence ($BF_{01} = 50.74$) perhaps due to familiarization with the learning procedure. We found no noteworthy evidence for any other effects of our experimental manipulations, all $BF_{01} \geq 1.87$.

3.6.2.1.2 Exploratory analyses Although participants may have retained some expectancy of the correct US at the end of the extinction procedure, this expectancy was markedly higher in the acquisition than in the extinction procedure, $M = 0.71$ 95% HDI [0.61, 0.82], $BF_{10} = 1.06 \times 10^{16}$ (one-tailed).

3.6.2.2 CS pleasantness

3.6.2.2.1 Confirmatory analyses As a measure of the EC effect, we calculated difference scores between mean evaluative ratings of CSs that were paired with positive and negative USs ($\bar{x}_{EC} = \bar{x}_{US+} - \bar{x}_{US-}$) for every participant in every cell of the experimental design. We analyzed EC effects using a 2 (*Learning procedure*: Acquisition vs. Extinction) × 3 (*Referenced context*: First vs. Second vs. None) repeated-measures ANOVA. As predicted, referring to and reinstating learning contexts affected the EC effect differently depending on the learning procedure, $BF_{10} = 360.74$, Figure 3.2B. Planned contrasts indicated that when participants rated CS pleasantness in the new context, we found strong evidence for an EC effect in the extinction procedure, $M = 2.42$ 95% HDI [1.38, 3.50], $\mathrm{BF}_{10} = 1.79 \times 10^3$ (one-tailed). Moreover, we found evidence that this EC effect was comparable to the EC effect in the acquisition procedure, $M = 0.32$ 95% HDI [0.00, 1.14], $\mathrm{BF}_{01} = 12.30$ (one-tailed). When we compared EC effects for the first and second context, we observed both the predicted increase in the acquisition, $\mathrm{BF}_{10} = 1.79 \times 10^5$ (one-tailed), as well as the predicted decrease in the extinction procedure, $\mathrm{BF}_{10} = 6.77 \times 10^3$ (one-tailed). Critically, in the extinction procedure, the EC effect was reduced when participants rated CSs for the context of CS-alone trials compared to the new context, $\mathrm{BF}_{10} = 74.23$ (one-tailed). We found only relatively weak evidence indicating that our learning procedure may not have extinguished the EC effect completely, $M = 1.17$ 95% HDI [0.19, 2.09], $\mathrm{BF}_{10} = 4.60$. The comparison between the EC effects for the context of CS-US pairing trials and the new context was inconclusive, $\mathrm{BF}_{01} = 2.03$. Similarly, in the acquisition procedure, the EC effect was reduced when participants rated CSs for the context of CS-alone trials compared to the new context, $\mathrm{BF}_{10} = 576.00$ (one-tailed). The comparison between the EC effect for the context of CS-US pairing trials and the new context was again inconclusive, $\mathrm{BF}_{01} = 1.87$. In sum, we found that the EC effect appeared to be resistant to extinction when participants rated CS pleasantness in a new context after completion of the learning procedure; but we found a reduced EC effect when we referenced and reinstated the context in which CS had been presented alone.

3.6.2.2.2 Exploratory analyses We additionally compared the EC effects between learning procedures for each of the contexts. In the first context, participants exhibited a larger EC effect in the extinction than in the acquisition procedure, $M = 2.77$ 95% HDI [1.26, 4.29], $\mathrm{BF}_{10} = 122.86$ (one-sided). In the second context, the comparison was inconclusive, $M = 1.14$ 95% HDI [0.01, 2.37], $\mathrm{BF}_{01} = 1.34$ (one-sided). We found evidence indicating that participants' prior knowledge about CSs did not affect our findings, $\mathrm{BF}_{01} = 8.27$.

3.6.3 Discussion

Our results again confirm the predictions derived from the temporal integration hypothesis and our simulation. First, we replicated the expectancy-

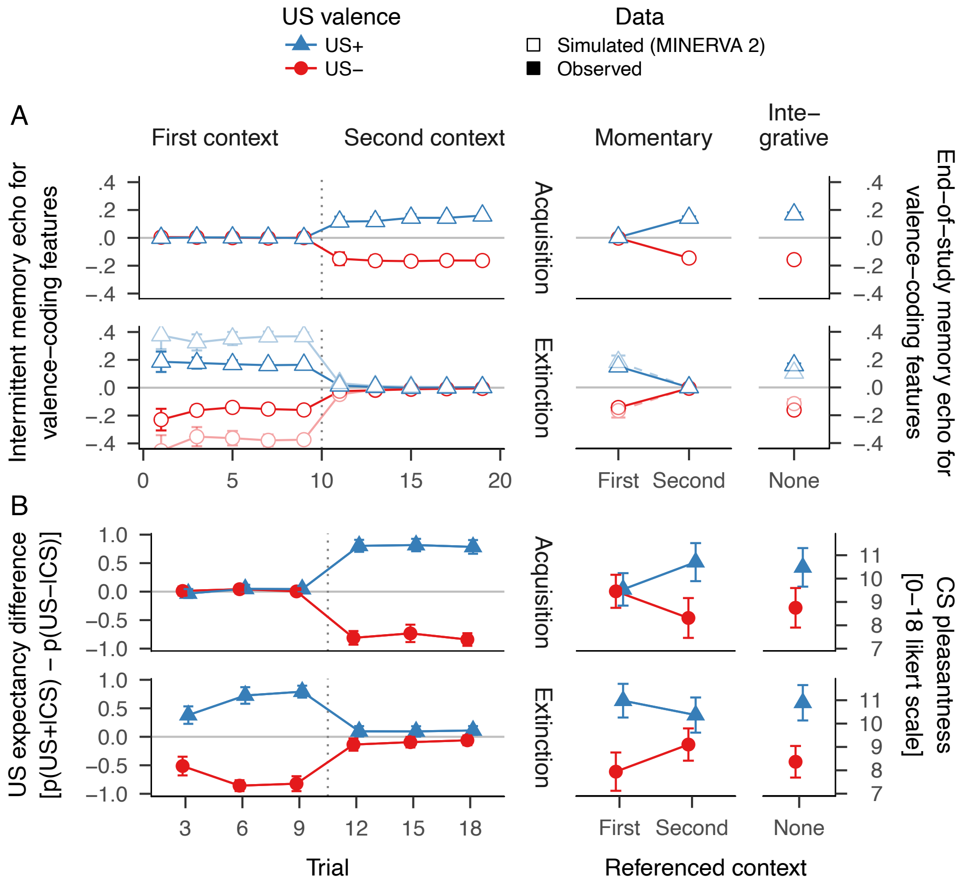

Figure 3.2: Simulated and observed US expectancy and CS pleasantness ratings for Experiment 2. Blue triangles indicate CSs paired with positive USs; red circles indicate CSs paired with negative USs. **A** Mean normalized memory echo of valence-coding features predicted by MINERVA 2 indicative of the overall valence of the retrieved memory contents. The left plot shows valence retrieved during the learning procedure, the right plot shows the valence retrieved after completion of the learning procedure. Faint symbols represent simulated ratings for a variant of our paradigm without the acquisition procedure or neutral CS-US trials. Error bars represent 95% confidence intervals. **B** The left plot shows observed differences in mean US expectancy during the learning procedure for acquisition (top) and extinction (bottom) procedures. Positive values indicate expectancy for positive USs, negative values indicate expectancy for negative USs. The right plot shows observed mean CS pleasantness ratings after completion of the learning procedure for acquisition (top) and extinction (bottom) procedures. Error bars represent 95% within-subject confidence intervals; CS = Conditioned stimulus, US = Unconditioned stimulus.

liking dissociation: In the last trial of the learning procedure, participants reported markedly higher US expectancies in the acquisition than in the extinction procedure. In contrast, when participants provided CS pleasantness judgments in a new context after the completion of the learning procedure (i.e., when learning contexts were not referenced), the EC effects did not differ between the acquisition and extinction procedure. Second, when we referenced and reinstated learning contexts participants again made contextualized CS pleasantness judgments: We observed extinction of the EC effect when participants evaluated CSs in the context of the CS-alone trials. These momentary CS pleasantness ratings, again, reflected the changes in CS-US contingencies and corresponded to the intermittent US expectancy ratings. Thus, as predicted, we elicited momentary CS pleasantness ratings and thereby eliminated the expectancy-liking dissociation. As in Experiment 1, this effect was obtained in the absence of potentially problematic intermittent CS pleasantness ratings. Jointly, our simulation and experimental findings provide further evidence that expectancy-liking dissociations can be explained as the result of different judgment strategies.

In Experiment 1 and 2, we assessed CS pleasantness repeatedly in different learning contexts. The repeated assessment may have introduced demand effects: Based on conversational norms, participants may have assumed that repeated ratings under varying conditions are expected to yield different responses. This also applies to intermittent US expectancy ratings. Moreover, these intermittent US expectancy ratings created a focus on CS-US pairings and US prediction. Although foci on pairings (e.g., Vansteenwegen et al., 2006; Vervliet et al., 2005; Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Hu et al., 2017) and on US prediction (e.g., Kattner & Green, 2015; Kattner, 2014; Zanon et al., 2012) are prevalent in EC research, it may limit the generalizability of our findings, and could be argued to impede automatic associative processes (Olson & Fazio, 2001). We addressed these limitations in Experiment 3.

3.7 Experiment 3

Our previous experiments indicate that the expectancy-liking dissociations in counterconditioning and extinction procedures are caused by different default judgment strategies and can be eliminated by inducing nondefault momentary CS pleasantness judgments. A comprehensive test of the temporal integration hypothesis, however, requires a concurrent manipulation of default judgment strategies for both US expectancy and CS pleasantness. The hypothesis predicts that, when judgment strategies are equated, US expectancy and CS pleasantness ratings should exhibit the same pattern of results. As a corollary, in an extinction procedure a comparison of non-default integrative US expectancy and momentary CS pleasantness judgments should reveal a reversed expectancy-liking dissociation—extinction of EC effects despite continued US expectancy. Experiment 3 was a final test of the temporal integration hypothesis in which we concurrently manipulated the judgment strategies for US expectancy and CS pleasantness ratings.

As for CS pleasantness ratings, we assessed US expectancies only after completion of the learning procedure in a momentary fashion separately for each learning context, as well as in an integrative fashion in a new context. For this procedure, MINERVA 2's predictions match those of Experiment 2. Hence, we expected to observe (1) identical patterns for end-of-study US expectancy and CS pleasantness ratings and (2) a reversed expectancy-liking dissociation (Figure 3.2A).

3.7.1 Methods

The experimental method, data analysis plan, and the following hypotheses were preregistered (https://osf.io/vnmby/registrations/): In the extinction procedure, we predicted (1) persistent US expectancy in end-of-study judgments that referred to both learning contexts (i.e. resistance to extinction in integrative US expectancy judgments). Moreover, we predicted lower US expectancy ratings for the context of CS-alone trials when compared to (2) the context of CS-US pairing trials as well as (3) to both learning contexts. For CS pleasantness ratings, we predicted the same pattern: Despite the extinction procedure, we expected to observe (4) an EC effect in end-of-study judgments in the new context. Moreover, we expected (5) a reduced EC effect for end-of-study CS pleasantness ratings for the context of CS-alone trials compared to the new context and (6) the context of CS-US pairing trials.

Experimental design, materials, and procedure followed those of Experiment 2 except for the following changes.

3.7.1.1 Participants

As in Experiment 2, we performed a sequential Bayesian analysis with minimum sample size of $n = 20$ per between subject condition ($N = 120$). We set out to collect data until they provided strong evidence for or against our six hypotheses of primary interest or until we ran out of money (1920€).

We recruited 273 new participants. As preregistered, we excluded 17 participants who performed the category recognition task at chance level, 57 participants who performed poorly at the identification task during the learning procedure (below $Q_1 - 1.5 \times IQR$, i.e. at least one incorrect response; see Procedure), and one participant who aborted the experiment. Thus, we stopped collecting data after 202 valid participants. At this point the data provided strong evidence for hypotheses 1, 2, and 5. Participants' mean age was 23.61 years ($SD = 6.41$), 146 were female, and 32 studied psychology or media psychology. 7 participants reported vision impairments: five were red-green color blind, one had astigmatism and another had a blind eye. 74 participants reported to have prior knowledge about the CS pictures.

3.7.1.2 Material

In contrast to Experiment 2, we did not collect intermittent US expectancy ratings. Instead, we asked participants to categorize USs ("What do you see right now?") in a 4-AFC task as photographs of either humans, animals, or objects or to indicate that no US was presented. The categorization task served to engage participants during the learning procedure in a manner comparable with our previous experiments. Analogous to CS pleasantness ratings, participants judged US expectancy for each CS after completion of the learning procedure for different contexts. We instructed participants that they would repeat a few trials from the learning procedure. We presented only the CSs and asked "With what probability would you expect a photograph of a human [animal/object] with this creature?" Previous studies have similarly assessed expectancy retrospectively by asking participants to graph the evolution of their US expectancy during the learning procedure (e.g., Raes et al., 2011; Vansteenwegen et al., 2005; Vervliet et al., 2005). To elicit momentary judgments, we noted that these trials were drawn from the first (second) half of the experiment. To elicit integrative judgments, we instructed participants that the trials were "selected randomly from the first and second half of the experiment" with equal probability. We noted that CSs would be shown in the center on neutral background to obscure which half of the experiment the trials were drawn from.

3.7.1.3 Procedure and design

In each of the six subblocks of the learning procedure, we collected US categorization responses for one CS from every US valence (including neutral CS-US pairs) following the third presentation of the CS-US pair. We removed the CS, but the US—if a US had been presented—remained on screen until participants responded. We made this task deliberately easy to avoid drawing too much attention to USs and away from CSs during the learning procedure.

After completion of the learning procedure, participants rated US expectancy and CS pleasantness for each CS. In contrast to Experiment 2, the context for CS pleasantness and US expectancy ratings was manipulated between participants, i.e., participants rated each CSs only for one context. We, thus, collected 3 US expectancy and CS pleasantness ratings per experimental condition and 18 per participant. Additionally, we manipulated the order of US expectancy or CS pleasantness ratings to control for possible order effects (Heycke et al., 2017).

On average, participants took 49.62 minutes ($SD = 12.71$) to complete the study.

3.7.2 Results

Preregistered analyses (labeled confirmatory) will be presented first, followed by additional (exploratory) analyses. In addition to the analyses presented here, we repeated all analyses with a modified set of exclusion criteria: We included participants who made no more than one incorrect response in the intermittent US identification task—some participants reported accidentally clicking the wrong button—but excluded three participants who invariably used the scale mid-point in CS pleasantness ratings (see the online supplemental material). By and large, we found the same results; we indicate noteworthy changes in the exploratory analyses sections. See Appendix A.2 for analyses of participants' CS-US pairing memory.

3.7.2.1 US expectancy

3.7.2.1.1 Confirmatory analyses

We analyzed expectancies of the correct US using a 3 (*Valence*: Positive vs. Neutral vs. Negative) × 2 (*Learning procedure*: Acquisition vs. Extinction) × 3 (*Referenced context*: First vs. Second vs. Both) × 2 (*DV order*: CS pleasantness first vs. US expectancy first) ANOVA with repeated measurements on the first two factors. As predicted, we found strong evidence that the changes in US expectancy across contexts differed between acquisition and extinction procedures, $\mathrm{BF}_{10} = 1.03 \times 10^{33}$, Figure 3.3. We observed this pattern irrespective of US valence ($\mathrm{BF}_{01} = 66.72$) and of whether US expectancy was assessed before or after CS pleasantness, $\mathrm{BF}_{01} = 8.50$. We therefore analyzed all data and averaged across US valences.

As predicted, planned contrasts indicated that expectancy for the correct US category increased from the first to the second learning context in the acquisition procedure ($\mathrm{BF}_{10} = 11.19$, one-tailed) but decreased in the extinction procedure, $\mathrm{BF}_{10} = 6.12 \times 10^4$ (one-tailed). When we referenced both learning contexts, we found strong evidence that participants expected the correct USs despite the previous extinction procedure, $M = 0.37$ 95% HDI $[0.26, 0.48]$, $\mathrm{BF}_{10} = 1.18 \times 10^7$, one-tailed. The comparisons of US expectancy for both contexts versus the second context was inconclusive in both acquisition ($\mathrm{BF}_{10} = 1.38$, one-tailed) and extinction procedures, $\mathrm{BF}_{10} = 1.92$ (one-tailed). There was no noteworthy evidence to suggest that there was any other effect of our manipulations, $\mathrm{BF}_{01} \geq 1.54$. In sum, participants' end-of-study US expectancies corresponded to CS-US contingencies when we referenced and reinstated the learning contexts.

3.7.2.1.2 Exploratory analyses

Because we found no conclusive evidence for or against integrative judgments in the preregistered between-participant comparisons of ratings for the second and the new context, we additionally compared the differences between the acquisition and extinction procedures for all referenced contexts. For the first learning context, participants expressed higher expectancy for the correct US in the extinction than in the acquisition procedure, $M = 0.37$ 95% HDI $[0.25, 0.48]$, $\mathrm{BF}_{10} = 2.97 \times 10^6$ (one-sided). This pattern was reversed in the second context: Participants expressed higher expectancy for the correct US in the acquisition than in the extinction procedure, $M = 0.24$ 95% HDI $[0.14, 0.33]$,

$BF_{10} = 9.85 \times 10^3$ (one-sided). Critically, when we referenced both learning contexts we found some evidence that expectancy for the correct US did not differ between acquisition and extinction procedures, $M = 0.03$ 95% HDI $[0.00, 0.07]$, $BF_{01} = 5.55$ (one-sided). These additional analyses indicate that, like the EC effect, US expectancy appeared to be resistant to extinction when we referenced both learning contexts. Hence, we conclude that we successfully elicited integrative US expectancy judgments.

Compared to the intermittent ratings in Experiments 1 and 2, participants reported expecting USs in the context of CS-alone trials and overall their expectancies were less pronounced. Further analyses suggested that this reflects memory confusions of the learning contexts.

3.7.2.2 CS pleasantness

3.7.2.2.1 Confirmatory analyses We analyzed EC effects using a 2 (*Learning procedure*: Acquisition vs. Extinction) × 3 (*Referenced context*: First vs. Second vs. None) × 2 (*DV order*: CS pleasantness first vs. US expectancy first) ANOVA with repeated measurements on the first factor. As predicted, referring to and reinstating learning contexts affected the EC effect differently depending on the learning procedure, $BF_{10} = 1.45 \times 10^3$, Figure 3.3. This finding was not affected by the order of DVs ($BF_{01} = 5.30$) and, thus, we analyzed all data. End-of-study CS pleasantness ratings in the new context provided some evidence for an EC effect in the extinction conditions, $M = 0.99$ 95% HDI $[0.20, 1.80]$, $BF_{10} = 4.35$ (one-tailed). Moreover, we found evidence, albeit weak, that this EC effect was of comparable magnitude in the extinction and acquisition procedure, $M = 0.60$ 95% HDI $[0.00, 1.52]$, $BF_{01} = 4.64$ (one-tailed). When we compared participants' CS pleasantness ratings for the first and second context, we observed both the predicted increase in the EC effect in the acquisition procedure, $BF_{10} = 19.44$ (one-tailed), as well as the predicted decrease in the extinction procedure, $BF_{10} = 38.71$ (one-tailed). In the extinction procedure, the EC effects for the context of CS-alone trials and the new context were of comparable magnitude , $BF_{01} = 5.10$ (one-tailed). EC in the context of CS-alone trials was not extinguished completely, $M = 1.31$ 95% HDI $[0.34, 2.29]$, $BF_{10} = 7.19$; but we found evidence for partial extinction. The EC effect was clearly larger in the context of CS-US pairing trials than in the new context, $BF_{10} = 88.80$, in line with the meta-analytic finding. Similarly, in the acquisition procedure, the EC effect for the context of CS-US pairing trials was larger than for the new context, $\mathrm{BF}_{10} = 10.95$. The comparison between the EC effect for the context of CS-alone trials and the new context was, however, inconclusive, $\mathrm{BF}_{01} = 1.70$ (one-tailed). We found no noteworthy evidence for any other effects of our manipulations, $\mathrm{BF}_{10} \leq 2.82$. In sum, we found some indication that EC effects were comparable in the acquisition and extinction procedures when participants rated CS pleasantness in the new context after completion of the learning procedure. We also observed the predicted extinction of EC in nondefault momentary CS pleasantness judgments: The EC effect was larger for the context of CS-US pairing trials than for the context of CS-alone trials.

3.7.2.2.2 Exploratory analyses Additionally, in the first learning context participants exhibited a larger EC effect in the extinction than in the acquisition procedure, $M = 2.24$ 95% HDI [0.80, 3.70], $\mathrm{BF}_{10} = 22.69$ (one-sided). This pattern reversed in the second context: Participants exhibited a larger EC effect in the acquisition than in the extinction procedure, $M = 2.30$ 95% HDI [0.82, 3.68], $\mathrm{BF}_{10} = 28.47$ (one-sided). The data were uninformative as to whether participants' prior knowledge about CSs affected these findings, $\mathrm{BF}_{01} = 1.85$.

In the exploratory analysis using the modified set of exclusion criteria ($n = 229$), we found stronger evidence in support of an EC effect in the new context in the extinction procedure, $\mathrm{BF}_{10} = 11.60$ (one-tailed). In this larger sample we also found stronger evidence indicating that the magnitude of this EC effect was comparable in the extinction and the acquisition procedure, $\mathrm{BF}_{01} = 7.03$ (one-tailed).

3.7.3 Discussion

Experiment 3 replicated and extended our previous findings in the absence of both intermittent US expectancy ratings and repeated end-of-study assessment of US expectancy or CS pleasantness across different contexts. We successfully elicited momentary US expectancy judgments by referring to and reinstating the learning contexts that adequately reflected the CS-US contingency changes: In both learning procedures, participants' US expectancy was larger in the context of CS-US pairing trials than in the

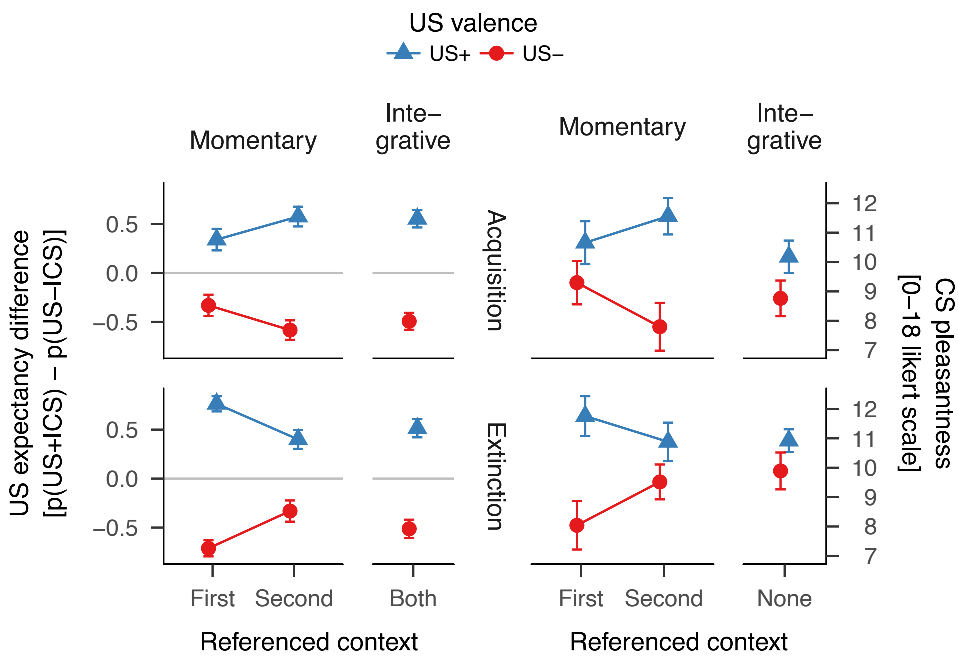

Figure 3.3: US expectancy and CS pleasantness ratings at the end of Experiment 3. The left plot shows observed differences in mean US expectancy for acquisition (top) and extinction (bottom) procedures. Positive values indicate expectancy for positive USs, negative values indicate expectancy for negative USs. The right plot shows observed mean CS pleasantness ratings after completion of the learning procedure for acquisition (top) and extinction (bottom) procedures. Error bars represent 95% within-subject confidence intervals; CS = Conditioned stimulus, US = Unconditioned stimulus.

context of CS-alone trials. Following extinction learning, participants reported residual US expectancies in the context of CS-alone trials, but US expectancy was higher in the acquisition procedure. Critically, US expectancy in the acquisition and extinction procedures was comparable when we referenced both learning contexts, reflecting a nondefault integrative judgment strategy.

As predicted by the temporal integration hypothesis and our simulation, CS pleasantness ratings exhibited the same pattern of results: In both learning procedures, EC effects were larger in the context of CS-US pairing trials than in the context of CS-alone trials. Following extinction learning, participants' ratings exhibited a residual EC effect in the context of CS-alone trials, but the effect was markedly higher in the acquisition procedure. Again, when no learning context was referenced the EC effect was comparable between acquisition and extinction procedures, reflecting the default integrative judgment strategy.

In summary, the extinction procedure showed the well-known expectancy-liking dissociation between default momentary US expectancy ratings and integrative CS pleasantness ratings. Conversely, we found the reversed dissociation between nondefault integrative US expectancy and momentary CS pleasantness ratings. Thus, after equating judgment strategies, US expectancy and CS pleasantness exhibited the same pattern of results. We conclude that expectancy-liking dissociations can be accounted for by differences in default judgment strategies and do not necessitate two distinct learning systems.

3.8 General Discussion

To explain expectancy-liking dissociations in EC (e.g. Baeyens et al., 1988; Baeyens et al., 2005; Hermans et al., 2002) with a single learning process, Lipp et al. (2010) proposed that US expectancy and CS pleasantness ratings afford different default judgment strategies: US expectancy ratings reflect momentary whereas CS pleasantness ratings reflect integrative summaries of the learning history. We tested this temporal integration hypothesis by manipulating participants' judgment strategies after completion of the learning procedure in a counterconditioning and two extinction experiments. Under default conditions (i.e., momentary US expectancy and integrative CS pleasantness judgments), we replicated two expectancy-liking dissociations: Counterconditioning produced no EC effects although participants US expectancies reflected the contingencies in the second part of the learning phase; conversely, extinction produced EC effects in the absence of US expectancies. Our findings corroborate that these dissociations were caused by the difference in strategies. First, we eliminated these dissociations by equating the judgment strategies across measures: CS pleasantness ratings corresponded to the respective US expectancy ratings after we elicited (nondefault) momentary CS pleasantness judgments (i.e., by referring to and reinstating the context of the initial or opposed CS-US pairings in the counterconditioning procedure in Experiment 1, or the context in which CSs had been presented alone as in the extinction procedures of Experiments 2 and 3). Furthermore, we reversed the expectancy-liking dissociation in the extinction paradigm by contrasting (nondefault) integrative US expectancy with (nondefault) momentary CS pleasantness judgments. Results showed extinction of EC but resistance to extinction of US expectancy. Our findings demonstrate that the expectancy-liking dissociations reported in the literature can be produced as the result of different judgment strategies afforded by the dependent measures. Hence, contrary to previous interpretations, expectancy-liking dissociations do not necessitate a second learning process; they can be parsimoniously explained by a single learning process.

Amending and extending Lipp et al. (2010)'s temporal-integration hypothesis, we illustrate how the learning history can be conserved and utilized to perform judgment tasks: Based on previous theorizing (Mitchell et al., 2009), we instantiated learning and retrieval processes by the unitary episodic memory model MINERVA 2 (Hintzman, 1988; Hintzman, 1984; Hintzman, 1986), which enabled us to make specific predictions for both US expectancy and CS evaluations that were subsequently corroborated by our experiments.

3.8.1 Additional findings and limitations

Hofmann et al. (2010) found partial extinction of EC in studies that assessed the EC effect both after the acquisition and again after the extinction phase. At first glance, this finding may seem to contradict the results of our simulation (Figure 3.2A), which predicted comparable CS pleasantness ratings for intermittent ratings at the end of the acquisition procedure, integrative end-of-study ratings, as well as momentary ratings in the CS-US pairing context, However, this discrepancy is due to procedural factors. For learning procedures that elicit momentary CS evaluations via repeated ratings, MINERVA 2 predicts partial extinction. In such cases, postextinction ratings should more strongly reflect recent CS-alone trials. Moreover, the model predictions also depend on other aspects of the experimental designs, such as the presentation of neutral stimuli (e.g., CS-alone trials or neutral CS-US pairs) during the initial acquisition phase in the present studies. This is because the common context causes activation of neutral stimuli, which in turn attenuate CS pleasantness ratings. As illustrated by the faint symbols in Figure 3.2A, MINERVA 2 predicts the partial extinction effect for designs without neutral stimuli during the initial acquisition phase.

In the present studies, we manipulated learning contexts overtly. However, we believe that our findings also pertain to evaluative learning without context manipulations. Lipp and Purkis (2006) found that using paper and pencil rather than a computer for end-of-study valence assessment is sufficient to elicit integrative CS pleasantness judgments without any explicit context manipulation. Similar renewal effects have been found in social impression formation (AAB renewal; Gawronski et al., 2010, Experiment 4). Hence, the standard end-of-study evaluation assessment may act as context change, affect judgment strategies, and produce renewal effects. Moreover, we assume that in the absence of explicitly induced context changes, participants spontaneously generate and use temporal contexts to structure the incoming information and that these contexts can affect behavior (Matute et al., 2011; Zacks et al., 2007). Due to the use of external contexts, contextualization in our study may have been more pronounced but—we assume—not qualitatively different from previous studies. We plan to test these assumptions in future research.

Yet another type of context has been employed in studies on feature-positive learning in EC—another procedure that has produced an expectancy-liking dissociation (Baeyens et al., 1996; Baeyens et al., 1998). In this paradigm, a CS was paired with a negative US only in the presence of (or subsequent to) a feature stimulus; in its absence, the CS was presented alone. EC effects were obtained both when the CS was rated in the presence and absence of the feature stimulus—even when participants correctly reported the stimulus contingencies. For such a procedure, MINERVA 2 predicts a reduced, albeit nonzero, EC effect when the CS is rated in the absence compared to the presence of the feature stimulus. If the nonsignificant reduction, which has so far been obtained only with relatively small samples (Baeyens et al., 1996; Baeyens et al., 1998), proves robust in high-powered studies, it is a finding that challenges our model and may necessitate further refinement.

Comparing our findings to those from studies on social impression formation also reveals a potentially interesting inconsistency (for a review see Gawronski et al., 2018). In their counterconditioning-like paradigm, Gawronski et al. (2010) found that participants' evaluations in a new context reflected the valence of the initially presented information (ABC renewal), whereas the present studies instead found neutral evaluations. Interestingly, Gawronski et al. (2018) also failed to observe those renewal effects in preliminary EC studies (p. 43). Although social impression formation and EC procedures differ in many aspects, we speculate that attentional processes may explain the contradictory results. Our model assumed constant attention to context—we informed participants that the learning procedure consisted of two phases. If, instead, we assume that attention to context increases in the second context—as do Gawronski et al. (2018)—our model predicts ABC renewal. Conversely, when Gawronski et al. (2010) enhanced attention to the first context, ABC renewal was eliminated (Experiment 4). The effects of attention to, and encoding of, context features is closely linked to the above considerations about the different types of context manipulations and deserves further study.

3.8.2 Implications

3.8.2.1 The role of dependent measures

We demonstrate the expectancy-liking dissociation in extinction learning—as well as its reversal—while holding encoding constant and manipulating only the retrieval or judgment process. Somewhat relatedly, Gawronski et al. (2014) interpreted the meta-analytical finding of a small reduction of EC due to extinction (Hofmann et al., 2010) as an artifact of judgment-related nuisance processes. They argued that extinction procedures do not affect the underlying evaluative representations. Our results corroborate the importance of judgment processes in evaluative responses, but they also highlight that similar judgment processes affect expectancy judgments. With this in mind, the expectancy-liking dissociations and the resistance to extinction of EC can similarly be construed as an artifact of judgment-related processes. Resistance to extinction appears to reflect different judgment strategies rather than characteristics of separable learning systems.

More generally, our findings illustrate that conclusions about latent processes require a good understanding, and careful experimental treatment, of the dependent measures. Dissociations taken to imply the operation of different learning processes may alternatively be explained by differences in retrieval or performance processes that bear on the assessed variable. Without a good understanding of the dependent measures (i.e., without an established measurement theory), contrasting these measures runs the risk of comparing apples and oranges. Instead, stronger and more direct tests of dual-process claims can be achieved if the outcomes of experimental manipulations that selectively target the two postulated processes are assessed and compared on a single dependent variable.

3.8.2.2 Explicit versus implicit measures

Another dissociation often discussed in the EC literature is the one between direct measures, such as ratings, and indirect measures, such as evaluative priming. In research on extinction, these dissociations have often been interpreted as evidence for dual-processes theories (e.g., Gawronski et al., 2014; Kattner & Green, 2015). This interpretation rests on the assumption that ratings primarily reflect explicit learning and priming measures reflect implicit learning. Challenging this assumption, EC studies have routinely used direct measures to assess implicit learning (e.g., Olson & Fazio, 2001; Hütter et al., 2012); and effects supporting explicit learning have been obtained on indirect measures (e.g., EC requires awareness; Pleyers et al., 2007; Stahl et al., 2009). A recent review of implicit-explicit dissociations in attitude learning concludes that there is little evidence for the assumption of a link between learning mode and expression mode (Corneille & Stahl, 2019b). Taken together, these findings illustrate that dissociations between direct and indirect measures are often absent; and where obtained, they would merely be consistent with dual-process assumptions but fail to corroborate them.

Gawronski et al. (2014) reported that extinction reduced EC effects on the direct evaluative ratings but did not affect the indirect evaluative priming measure. According to the temporal integration hypothesis, the extinction of EC on the direct measure reflects its context-sensitivity: When repeatedly asked to evaluate the CSs following acquisition and again after extinction, this repetition induces a momentary judgment strategy—participants' latter ratings reflected primarily the information encoded during the extinction phase. Assuming that the resistance to extinction inferred from evaluative priming is not merely an artifact of the measure's inferior reliability and sensitivity, a possible explanation is that judgment strategies can be adapted more readily for explicit ratings than in evaluative priming. In contrast to direct expressions of attitudes, context-sensitive responses require considerable effort in evaluative priming measures (on top of task-specific knowledge and strategies; Klauer & Teige-Mocigemba, 2007; Teige-Mocigemba & Klauer, 2008). Thus, a more targeted manipulation may be necessary to modify the default integrative judgment strategy in evaluative priming. The research reviewed by Gawronski et al. (2018) suggests momentary judgments can be elicited in indirect measures by introducing context manipulations similar to ours. To the degree that such context effects indeed affect indirect measures, our findings should generalize to these measures.

Our aim was not to conclusively rule out the existence of a second learning process. Specific conditions may, for example, allow for an additional implicit misattribution (IM) of unconditioned evaluative responses to the paired CSs (Olson & Fazio, 2001). In contrast to our stimulus-stimulus (S-S) learning account, IM postulates stimulus-response (S-R) learning—links between CSs and evaluative responses. Other properties of CS-US pairings, such as context, should be inconsequential for IM; it therefore cannot readily explain the contextualized EC effects we observed. Our procedure realized some conditions that are taken to promote IM (a seemingly random stream of stimuli and incidental learning of US valence; Hütter & Sweldens, 2013; Jones et al., 2009) but it could be adapted to further bolster misattribution of evaluative responses. Specifically, incidental instructions, together with simultaneous onset of CS and US (Hütter & Sweldens, 2013) may reveal context-insensitive IM.

3.8.2.3 Memory models that account for learning phenomena

Traditionally, learning and memory have often been studied by separate research communities. However, as illustrated by Tolman's notion of memory as latent learning, it has been clear that these phenomena are overlapping as the effects studied in learning research must be mediated by (some form of) memory. The present study is one of several that sketch how to integrate learning phenomena into established memory theorizing. Here we recast EC—traditionally construed as a learning phenomenon—in terms of episodic memory theory. We view evaluative learning as encoding and retrieval of episodic knowledge that may later be used to construct adaptive judgments.

Jamieson et al. (2012) have proposed a related account of associative learning based on an adapted memory model called Minerva-AL. The crucial difference between MINERVA 2 and Minerva-AL resides in the encoding mechanism: Instead of passively encoding episodes, Minerva-AL assumes that CSs evoke predictions about USs and that only discrepancies between predictions and observed events—the prediction error—is stored in memory. Although Minerva-AL makes the same representational assumptions and posits the same retrieval mechanisms as MINERVA 2, the discrepancy-encoding mechanism results in a model that is closely related to classical learning models such as the Rescorla-Wagner model (Rescorla & Wagner, 1972; Miller et al., 1995; Siegel & Allan, 1996).

Minerva-AL was developed to account for phenomena in classical conditioning, which is believed to be a highly intentional learning procedure in which outcome expectations drive responses (Mitchell et al., 2009). Such conditions may encourage and even require continuous predictions and error monitoring. In EC, however, incidental paradigms, which obfuscate CS-US contingencies, are of particular relevance to the single- vs. dual-process debate (Olson & Fazio, 2001; Stahl & Heycke, 2016; Corneille & Stahl, 2019b). It is unclear whether and how participants generate and test predictions about CS-US associations in incidental paradigms; passive encoding of CS-US pairings, as assumed by MINERVA 2, may be a more appropriate assumption here. While MINERVA 2 predicted our findings despite the intentional learning instructions and intermittent US expectancy ratings, extending our approach to other paradigms and effects may necessitate modifications or additional assumptions. Comparing MINERVA 2 and Minerva-AL and exploring their limitations with respect to EC and classical conditioning is an interesting direction for future research.

It has long been known that MINERVA 2 requires additional assumptions to account for some of the critical empirical findings in recognition memory. For example, recall strategies have to be assumed to account for some findings in associative recognition (Clark & Gronlund, 1996). Yet, MINERVA 2, and the class of global-matching models to which it belongs, have been influential. The fact that the model predicted the outcomes of our experiments is an encouraging first indication that MINERVA 2 and related memory models (Clark & Gronlund, 1996; Humphreys, Pike, et al., 1989; Shiffrin & Steyvers, 1997; Kelly et al., 2017) are viable candidates for process models of EC that merit further exploration.

References

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature News*, *508*(7496), 312. https://doi.org/10.1038/508312a

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown* [R package version 0.1.0.9485]. https://github.com/crsh/papaja

Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996). No Evidence for Modulation of Evaluative Flavor–Flavor Associations in Humans. *Learning and Motivation*, *27*(2), 200–241. https://doi.org/10.1006/lmot.1996.0012

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy*, *10*(4), 179–199. https://doi.org/10.1016/0146-6402(88)90014-8

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). *Behaviour Research and Therapy*, *33*(7), 825–831. https://doi.org/10.1016/0005-7967(95)00021-O

Baeyens, F., Díaz, E., & Ruiz, G. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition and Emotion*, *19*(2), 245–268. https://doi.org/10.1080/02699930441000300

Baeyens, F., Hendrickx, H., Crombez, G., & Hermans, D. (1998). Neither Extended Sequential nor Simultaneous Feature Positive Training Result in Modulation of Evaluative Flavor–Flavor Conditioning in Humans. *Appetite*, *31*(2), 185–204. https://doi.org/10.1006/appe.1998.0167

Baeyens, F., Vansteenwegen, D., & Hermans, D. (2009b). Associative learning requires associations, not propositions. *Behavioral and Brain Sciences*, *32*(02), 198. https://doi.org/10.1017/S0140525X09000867

Biegler, P., & Vargas, P. (2013). Ban the Sunset? Nonpropositional Content and Regulation of Pharmaceutical Advertising. *The American Journal of Bioethics*, *13*(5), 3–13. https://doi.org/10.1080/15265161.2013.776127

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation*, *39*(1), 58–70. https://doi.org/10.1016/j.lmot.2007.03.003

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*(1), 37–60. https://doi.org/10.3758/BF03210740

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, *30*(7), 1138–1147. https://doi.org/10.3758/BF03194331

Corneille, O., & Stahl, C. (2019b). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

De Houwer, J. (1998). Leren is eenvoudig, doen is moeilijk: Een nieuwe visie op het onderscheid tussen evaluatieve en Pavloviaanse conditionering [Learning is simple, performance is difficult: A new view on the relation between evaluative and Pavlovian conditioning]. *Gedragstherapie*, *31*, 49–66.

De Houwer, J. (2011). Evaluative conditioning: A review of procedure knowledge and mental process theories. In T. R. Schachtman & S. Reilly (Eds.), *Applications of learning and conditioning*. Oxford University Press.

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture–picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(2), 237–242. https://doi.org/10.1037/0097-7403.26.2.237

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853–869. https://doi.org/10.1037/0033-2909.127.6.853

Dwyer, D. M., Jarratt, F., & Dick, K. (2007). Evaluative conditioning with foods as CSs and body shapes as USs: No evidence for sex differences, extinction, or overshadowing. *Cognition and Emotion*, *21*(2), 281–299. https://doi.org/10.1080/02699930600551592

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, *25*(4), 639–656. https://doi.org/10.1080/02699931.2010.513497

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pitbull: EC effects depend on CS–US relations. *Cognition and Emotion*, *26*(3), 534–540. https://doi.org/10.1080/02699931.2011.588687

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., Gast, A., & De Houwer, J. (2014). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition and Emotion*, 1–15. https://doi.org/10.1080/02699931.2014.947919

Gawronski, B., Rydell, R. J., De Houwer, J., Brannon, S. M., Ye, Y., Vervliet, B., & Hu, X. (2018). Contextualized Attitude Change. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology*. Academic Press. https://doi.org/10.1016/bs.aesp.2017.06.001

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General*, *139*(4), 683–701. https://doi.org/10.1037/a0020315

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In S. P. Shohov (Ed.), *Advances in psychology research* (pp. 17–40). Nova Science Publishers.

Heycke, T., Aust, F., & Stahl, C. (2017). Subliminal influence on preferences? A test of evaluative conditioning for brief visual conditioned stimuli using auditory unconditioned stimuli. *Royal Society Open Science*, *4*, 160935. https://doi.org/10.1098/rsos.160935

Heycke, T., Gehrmann, S., Haaf, J. M., & Stahl, C. (2018). Of two minds or one? A registered replication of Rydell et al. (2006). *Cognition and Emotion*, *32*(8), 1708–1727. https://doi.org/10.1080/02699931.2018.1429389

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Tenzing: Documenting contributorship using CRediT. *MetaArXiv*. https://doi.org/10.31222/osf.io/b6ywe

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional Versus Dual-Process Accounts of Evaluative Conditioning: I. The Effects of Co-Occurrence and Relational Information on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *43*(1), 17–32. https://doi.org/10.1177/0146167216673351

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, *33*(1), 36–67. https://doi.org/10.1016/0022-2496(89)90003-5

Hütter, M., & Sweldens, S. (2013). Implicit misattribution of evaluative responses: Contingency-unaware evaluative conditioning requires simultaneous stimulus presentations. *Journal of Experimental Psychology: General*, *142*(3), 638–643. https://doi.org/10.1037/a0029989

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*(3), 539–557. https://doi.org/10.1037/a0026477

Jamieson, R. K., Crump, M. J. C., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning & Behavior*, *40*(1), 61–82. https://doi.org/10.3758/s13420-011-0046-2

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, *96*(5), 933–948. https://doi.org/10.1037/a0014747

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.2307/2291091

Kattner, F. (2014). Reconsidering the (in)sensitivity of evaluative conditioning to reinforcement density and CS–US contingency. *Learning and Motivation*, *45*, 15–29. https://doi.org/10.1016/j.lmot.2013.09.002

Kattner, F., & Green, C. S. (2015). Cue competition in evaluative conditioning as a function of the learning process. *Acta Psychologica*, *162*, 40–50. https://doi.org/10.1016/j.actpsy.2015.09.013

Kelly, M. A., Mewhort, D. J. K., & West, R. L. (2017). The memory tesseract: Mathematical equivalence between composite and separate storage memory models. *Journal of Mathematical Psychology*, *77*, 142–155. https://doi.org/10.1016/j.jmp.2016.10.006

Klauer, K. C. (2009). Spontaneous evaluations. In F. Strack & J. Förster (Eds.), *Social cognition: The basis of human interaction* (pp. 199–217). Psychology Press.

Klauer, K. C., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology*, *43*(4), 648–655. https://doi.org/10.1016/j.jesp.2006.06.003

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550617697177

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-8). University of Florida, Gainesville, FL.

Lenth, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.1.2]. https://CRAN.R-project.org/package=emmeans

Lipp, O. V., Mallan, K. M., Libera, M., & Tan, M. (2010). The effects of verbal instruction on affective and expectancy learning. *Behaviour Research and Therapy*, *48*(3), 203–209. https://doi.org/10.1016/j.brat.2009.11.002

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation*, *34*(3), 219–239. https://doi.org/10.1016/S0023-9690(03)00011-0

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 431–440. https://doi.org/10.1037/0097-7403.32.4.431

Lovibond, P. F. (2004). Cognitive Processes in Extinction. *Learning & Memory*, *11*(5), 495–500. https://doi.org/10.1101/lm.79604

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3–26. https://doi.org/10.1037/0097-7403.28.1.3

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matute, H., Lipp, O. V., Vadillo, M. A., & Humphreys, M. S. (2011). Temporal contexts: Filling the gap between episodic memory and associative learning. *Journal of Experimental Psychology: General*, *140*(4), 660–673. https://doi.org/10.1037/a0023862

Matute, H., Vegas, S., & De Marez, P.-J. (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 714–725. https://doi.org/10.1037/0278-7393.28.4.714

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386. https://doi.org/10.1037/0033-2909.117.3.363

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183–198. https://doi.org/10.1017/S0140525X09000855

Moran, T., & Bar-Anan, Y. (2013b). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion*, *27*(4), 743–752. https://doi.org/10.1080/02699931.2012.732040

Morey, R. D., & Rouder, J. N. (2015). *Bayesfactor: Computation of bayes factors for common designs* [R package version 0.9.12-2]. https://CRAN.R-project.org/package=BayesFactor

Olson, M. A., & Fazio, R. H. (2001). Implicit Attitude Formation Through Classical Conditioning. *Psychological Science*, *12*(5), 413–417. https://doi.org/10.1111/1467-9280.00376

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of experimental psychology. Learning, memory, and cognition*, *33*(1), 130–144. https://doi.org/10.1037/0278-7393.33.1.130

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Raes, A. K., De Houwer, J., Verschuere, B., & De Raedt, R. (2011). Return of fear after retrospective inferences about the absence of an unconditioned stimulus during extinction. *Behaviour Research and Therapy*, *49*(3), 212–218. https://doi.org/10.1016/j.brat.2010.12.004

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Richter, J., & Gast, A. (2017). Distributed practice can boost evaluative conditioning by increasing memory for the stimulus pairs. *Acta Psychologica*, *179*, 1–13. https://doi.org/10.1016/j.actpsy.2017.06.007

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, *113*(3), 553–565. https://doi.org/10.1037/0033-2909.113.3.553

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008. https://doi.org/10.1037/0022-3514.91.6.995

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential Hypothesis Testing with Bayes Factors: Efficiently Testing Mean Differences. *Psychological Methods*. https://doi.org/10.1037/met0000061

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review*, *3*(3), 314–321. https://doi.org/10.3758/BF03210755

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2017). *Afex: Analysis of factorial experiments* [R package version 0.17-4]. https://github.com/singmann/afex

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal Evaluative Conditioning? Above-Chance CS Identification May Be Necessary and Insufficient for Attitude Learning. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000191

Stahl, C., & Heycke, T. (2016). Evaluative Conditioning with Simultaneous and Sequential Pairings Under Incidental and Intentional Learning Conditions. *Social Cognition*, *34*(5), 382–412. https://doi.org/10.1521/soco.2016.34.5.382

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*(3), 404–420. https://doi.org/10.1037/a0016196

Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review*, *18*(2), 187–209. https://doi.org/10.1177/1088868314527832

Teige-Mocigemba, S., & Klauer, K. C. (2008). 'Automatic' evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology*, *44*(5), 1414–1417. https://doi.org/10.1016/j.jesp.2008.04.004

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 71–79. https://doi.org/10.1037/0097-7403.32.1.71

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquistion context. *Behaviour Research and Therapy*, *43*(3), 323–336. https://doi.org/10.1016/j.brat.2004.01.001

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy*, *43*(3), 357–371. https://doi.org/10.1016/j.brat.2004.02.005

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wellek, S. (2002). *Testing Statistical Hypotheses of Equivalence* (1 edition). Chapman and Hall/CRC.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A Model of Dual Attitudes. *Psychological Review*, *107*(1), 101. https://doi.org/10.1037/0033-295X.107.1.101

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, *43*(3), 155–165. https://doi.org/10.1016/j.lmot.2012.02.003

Chapter 4

An exemplar-familiarity interpretation of verbatim and gist memory in false recognition

FREDERIK AUST, ROSCOE F. J. W. ARAUJO, & CHRISTOPH STAHL

False recognition has been attributed to representations of semantic gist or the summed similarity to all memory traces. Fuzzy-trace theory postulates that each episode leaves independent verbatim and gist memory traces. These independent representations dissociate true and false recognition: Gist traces are conceptual summaries; their retrieval induces (veridical and illusory) familiarity that can cause false recognition. Verbatim traces are detailed reflections of an episode; their retrieval induces vivid remembrance that support true recognition. In contrast to fuzzy-trace theory, global-matching memory models postulate that each episode leaves only one trace in a unitary storage. Dissociations between true and false recognition result from different patterns of probe-trace similarities: False recognition results from deceptive familiarity caused by partial matches with similar (but non-identical) traces. Exact matches are unique to true recognition and make an independent contribution to strength of familiarity.
We demonstrate a formal correspondence between estimates of verbatim and gist memory obtained from the Conjoint-Recognition Model and the contribution of exact and partial matches to the familiarity predicted by the Generalized Context Model. We fit both models to experimental results, which strongly suggests that multiple independent causes contribute to true and false recognition. In line with our theoretical model analysis, both models were able to account for the observed data. Taken together, our theoretical model analysis and model fits suggest that gist and verbatim retrieval may

reflect incremental contributions to familiarity by partial and exact matches between probes and memory traces.

Research into false recognition has informed theoretical questions about episodic memory (Brainerd & Reyna, 2005; Gallo, 2006). The most prominent approach to study false recognition is the Deese-Roediger-McDermott paradigm (DRM; Roediger & McDermott, 1995; Deese, 1959). In the DRM paradigm, participants study lists of words that are semantically related to one critical word (e.g, *chilly*, *hot*, *wet*, *winter*, *freeze*, *heat*, *snow*, and *winter* are all associated with *cold*). If a critical word is not presented, it is referred to as critical non-presented word or, more concisely, as *lure*. False recognition is typically defined as the rate at which lures are endorsed as old as compared to the rate of old-responses to new unrelated words. True recognition is assessed by presenting some critical words during the study phase. When presented again at test, the critical word is typically referred to as *target*. False recognition is, however, not limited to DRM material and can similarly be studied using, for example, categorized word lists (e.g., Buchanan et al., 1999; Pierce et al., 2005), photographs of similar everyday objects (e.g., Stark et al., 2013; Andermane & Bowers, 2015), pictures of abstract novel objects (e.g., Koutstaal et al., 1999; Pidgeon & Morcom, 2014), or merely perceptually similar words (Schacter et al., 1997) or drawings (Stahl, Henze, et al., 2016).

Despite many notable similarities there are systematic differences between true and false memories (Jou & Flores, 2013) and these differences have informed theories of episodic memory. False recognition has been typically

Individual contributions

Frederik Aust: Conceptualization, Data Curation, Formal Analysis, Methodology, Validation, Visualization, Writing - Original Draft Preparation

Roscoe F. J. W. Araujo: Conceptualization, Investigation, Methodology, Project Administration, Resources, Software

Christoph Stahl: Conceptualization, Funding Acquisition, Investigation, Methodology, Resources, Software, Supervision

Contributor Roles Taxonomy (CRediT; Allen et al., 2014; Holcombe et al., 2020)

been attributed to conceptual gist (Reyna & Brainerd, 1995a), semantic associations (Roediger et al., 2001; Underwood, 1965), or inter-item similarities (Arndt & Hirshman, 1998). Additionally, it has been proposed that participants sometimes use meta-cognitive or memory-based strategies to discount deceptively familiar lures (Brainerd et al., 2003; Roediger et al., 2001; Roediger & McDermott, 2000; Schacter et al., 1999). For a comprehensive overview of theoretical accounts we refer readers to Brainerd and Reyna (2005) and Gallo (2006).

Although it is commonly assumed that multiple processes contribute to true and false recognition, the nature of these processes remains contested. A fundamental disagreement is between single- and dual-trace theories. Fuzzy-trace theory is an influential dual-trace account that posits that each episode leaves independent gist and verbatim traces (Reyna & Brainerd, 1995a). While the conceptual gist of an episode is assumed to factor in both true and false recognition, detailed verbatim information is assumed to be unique to true recognition. In contrast, global matching memory models posit that episodes leave only one trace. Here it is assumed that partial matches between probes and memory traces factor in both true and false recognition, but exact matches are unique to true recognition (Arndt & Hirshman, 1998). Despite their conceptual similarity, global matching accounts of false recognition have been discounted based on empirical dissociations between true and false recognition.

In this article, we formalize the conceptual similarity between fuzzy-trace theory and global matching models as instantiated by the Conjoint-Recognition Model (CRM; Brainerd et al., 1999; Brainerd et al., 2001) and the Generalized Context Model (GCM; Nosofsky, 1986; Nosofsky, 1988; Nosofsky, 2011a). The results of this theoretical model analysis are consistent with previous suggestions that gist and verbatim retrieval may reflect incremental contributions of partial and exact matches between probes and traces in episodic memory. We test and confirm the adequacy of this global matching account empirically in an experiment designed to selectively influence the memory-related CRM parameters.

In the remainder of the introduction, we briefly outline how fuzzy-trace theory and global matching models account for dissociations between true and false recognition and examine some of the arguments cited in favor of fuzzy-trace theory's dual-trace assumption. We then motivate our experimental study by reviewing and formally relating CRM parameters to the memory probe familiarity predicted by GCM.

4.1 True and false recognition dissociate

It is widely accepted that true and false recognition are driven by two opponent processes (Reyna & Brainerd, 1995a; Reyna & Lloyd, 1997). This conclusion is drawn from a large body of research showing that true and false recognition (1) are uncorrelated, (2) can be manipulated independently of one another (single dissociations), and (3) exhibit opposite effects (double dissociations; p. 66, Brainerd & Reyna, 2005), such as the opposite developmental trajectories of these processes during childhood during childhood (Brainerd et al., 2008). These findings are thought to be caused by two processes that work in concert in true recognition but in opposition in false recognition. One process is typically conceptualized as a memory signal resulting from semantic, associative, or feature overlap between the memory probe and traces of past episodes (Arndt & Hirshman, 1998; Reyna & Brainerd, 1995a; Roediger et al., 2001). The memory signal becomes stronger as the overlap between probe and memory traces increases and promotes, both, true and false recognition. The second process promotes only true recognition and inhibits false recognition. Theoretical conceptions of this second process are less consistent and include memory-based and meta-cognitive explanations.

Fuzzy-trace theory, an influential account of false recognition, provides a memory-based account of the inhibitory process (Brainerd & Reyna, 1998; Reyna & Brainerd, 1995a; Reyna & Lloyd, 1997). Specifically, the theory postulates that each episode leaves independent gist and verbatim memory traces. Gist traces represent the meaning and relational stimulus information; they are conceptual summaries. Retrieval of gist races induces (true and illusory) familiarity that promotes both true and false recognition. In contrast, verbatim traces are detailed representations of an episode, including perceptual detail and surface features. Retrieval of verbatim traces induces vivid remembrance of past events, which promotes true but prevents false recognition. If a memory probe causes retrieval of a matching verbatim trace, the vivid remembrance provides strong support for a previous encounter with the probe (e.g., the stimulus was part of the study list). If, however, the memory probe causes retrieval of a conflicting verbatim trace, the vivid remembrance can discredit the memory probe as similar to but different from a previous experiences. This process is commonly referred to as *recollection rejection* (Brainerd et al., 2003).

4.2 Dissociations from unitary representations

The success of fuzzy-trace theory and its dual-trace explanation of opponent processes appears to contradict single-trace models. But while the dual-trace assumption is consistent with dissociations between true and false recognition, it is not necessitated by them. Such dissociations may similarly be explained by meta-cognitive response strategies (Dodson & Schacter, 2001; Roediger et al., 2001), dissociations between different information in compound memory traces (p. 56, Clark & Gronlund, 1996; Cowan, 1998), different retrieval operations (e.g., global match and content retrieval, Greve et al., 2010; Hintzman, 1987; Humphreys, Bain, et al., 1989), and possible other causes. Deciding between these alternative explanations requires more detailed considerations of the dissociative evidence. Proponents of fuzzy-trace theory have argued that the available evidence cannot be explained by meta-cognitive strategies (Reyna & Lloyd, 1997). However, dissociations of information in compound memory traces or dissociations between retrieval operations are less easily discounted.

Cowan (1998) early on criticized the idea of independent verbatim and gist traces and cautioned these might be best understood as distinct information in compound memory traces with different relevance to the memory task (p. 149; also see p. 56, Clark & Gronlund, 1996). Cowan noted that in many situations both interpretations yield very similar predictions—especially for verbal material. Reyna and Brainerd (1998) discount this information dissociation account by referencing studies that demonstrate dissociations between memory and reasoning tasks that used verbal material. However, as we will address next, task dissociation can be explained by different retrieval operations or decision rules and are, thus, in principle compatible with compound-trace models.

Whereas fuzzy-trace theory posits that *recollection rejection* relies on retrieval of independently encoded verbatim traces, it can also be construed as cued recall operation (*recall-to-reject*; Hintzman et al., 1992; Humphreys, Bain, et al., 1989; Malmberg, 2008). In short, a *matching* operation may compare a memory probe to the contents of memory. Matching yields a mnemonic strength signal that summarizes the overall closeness of the match—similar traces increment the strength signal more than dissimilar traces. This strength signal elicits feelings of familiarity and drives simple recognition decisions. As the strength signal contains no details about the past events, a second *retrieval* operation is required for reproductive tasks, such as cued recall. Retrieval yields a (more or less) faithful representation of the past as returned by memory. Hintzman (1987) demonstrated that matching and retrieval may yield independent (single dissociation) or even negatively related performance on recognition and cued recall tasks (double dissociation). In false recognition paradigms, a lure may be strategically used as a recall cue and the retrieved details may expose the lure as similar but new. Hence, the assumption of recollection rejection does not presuppose dual memory traces.

The broader argument here is that, depending on the task, the contents of memory may be cued and summarized in different ways. Such operations may yield responses that dissociate not because they rely on different memory systems or independent traces but because of the qualitatively different operations or decision rules applied to a "common representational substrate" (p. 707, Nosofsky, 1988, on dissociations between categorization and recognition tasks). This view follows Marr's principle of least commitment, which states that an adaptive flexible system should operate on the available information at the last possible moment (i.e., at the response stage) to avoid having to undo previous processing (pp. 485-486, Marr, 1976).

To summarize, the previous discussion illustrates that empirical single and double dissociations of true and false recognition are consistent with but do not necessitate the assumption of independent memory traces for gist and verbatim information. Such dissociations may be caused by dissociations of information in compound traces, or retrieval operations, or both. Distinguishing between single- and dual-trace representations requires experimental control of the retrieval operations (i.e. recollection rejection) and a specification of the single-trace representation. Global matching memory models are obvious candidates for the latter.

4.3 Global matching memory models

In contrast to fuzzy-trace theory, compound-trace global matching memory models postulate that each episode leaves only one memory trace (e.g., Clark & Gronlund, 1996; Hintzman, 1986; Shiffrin & Steyvers, 1997). These models attribute dissociations between true and false recognition to the information in memory traces. Each trace compounds item and context information. At test, the probe is matched against the contents of memory. The resulting strength signal reflects a summary of the similarity between the probe all traces of past episodes. Dissociations in recognition performance due to context information are ubiquitous, but item information alone may dissociate memory performance. For example, Arndt and Hirshman (1998) argued that in the global matching model MINERVA 2 (Hintzman, 1986) dissociations between true and false recognition can result from item-related probe-trace similarities: False recognition results from deceptive familiarity caused by partial matches between lures and similar but different memory traces. Exact matches, by definition, are unique to true recognition and should, thus, make an independent contribution to the familiarity of targets. Accordingly, MINERVA 2 correctly predicted that (1) increasing the number of similar items on a study list increases false recognition more than true recognition, (2) increasing presentation duration of the study list increases true but not false recognition (single dissociation), and (3) increasing the similarity between target and lure can increase false recognition but decrease true recognition (double dissociation).

In terms of fuzzy-trace theory, global matching models posit that gist activation corresponds to partial matches of features shared between lures and studied items, whereas verbatim activation reflects an increment of the match signal caused by features that distinguish lures from studied items—or lures from targets, respectively. This characterization of gist and verbatim activation is theoretically interesting because it suggest a mutually beneficial unifying relationship between fuzzy-trace theory and global matching models. Fuzzy-trace theory has stimulated ample empirical and theoretical work on the interplay of gist and verbatim memory. For example, a recently proposed framework specifies how task demands moderate the interaction of gist and verbatim information (Brainerd et al., 2019). The mathematical model and its parent theory, however, offer no account of how encoding and retrieval processes operate.[1] Conversely, global matching models provide detailed explanations of how information is encoded and how inter-item similarities factor into matching and retrieval processes. But the elaboration of the task-dependent interaction of these processes could benefit from the theoretical developments inspired by fuzzy-trace theory. Moreover, establishing a relationship with fuzzy-trace theory may broaden the applicability of global matching models research on reasoning (Reyna & Brainerd, 1995b; for an example see Dougherty et al., 1999).

The plausibility of the global matching account and its construal of gist and verbatim memory, however, presupposes descriptive adequacy, which in turn depends on the assumed inter-item similarities. Simulations, such as those reported by Arndt and Hirshman (1998), typically make convenient ad hoc assumptions about the inter-item similarities (also see Shiffrin & Steyvers, 1997). This is somewhat unsatisfying for two reasons. First, the similarity structure of the material is at the heart of the assumed matching (and retrieval) process and, thus, an important source of constraint. Simulating inter-item similarities adds auxiliary assumptions and free parameters, which can artificially increase a model's flexibility. Within plausible limits, inter-item similarities are effectively estimated to maximize the fit between simulation and observation. Second, Johns and Jones (2010) recently demonstrated that, in the case of lexical semantics, many simulated similarity distributions deviate from those of empirical estimates. Compared to simulated similarities the empirical distributions are more heavily skewed towards fewer high-similarity word pairs, which constitutes a violation of the simulation assumptions. Immediately relevant to the current discussion, Johns and Jones (2010) chose an experiment on false recognition to showcase how inappropriate assumptions about inter-item similarities can affect simulation results. When they constrained MINERVA 2 by using empirically estimated similarities the fit between simulated and observed data was substantially reduced; what is worse, the simulation diverged *qualitatively* from the observed data. The results by Johns and Jones (2010) raise doubts about the MINERVA 2 account of false recognition and call for more rigorous assessment of the global matching account.

Instead of MINERVA 2, we focus on the Generalized Context Model (GCM; Nosofsky, 1986; Nosofsky, 1988; Nosofsky, 2011a)—a different but related

[1] More generally, threshold models, such as the mathematical specification of fuzzy-trace theory, have rarely inspired the development of more detailed models of encoding and retrieval operations (p. 341, Malmberg, 2008).

global matching model. GCM is a well-tested, successful global matching model of recognition in long- and short-term memory (Nosofsky, Cox, et al., 2014; Nosofsky et al., 2020; also see Nosofsky, Cao, Cox, et al., 2014), as well as categorization, and has previously been applied to false recognition paradigms (Zaki & Nosofsky, 2001; see also van Vugt et al., 2013). Moreover, the model takes empirical similarity estimates into account and its formalism obviates the need for simulations, which allows us to fit it to observed data using conventional methods. In short, GCM is a seasoned broad-scoped representative of global matching models that is well-suited to assess the adequacy of the false recognition account. Next, we review and compare the mathematical specification of fuzzy-trace theory and the GCM before we test the descriptive adequacy of the global matching account.

4.4 Theoretical model analysis

In the following, we briefly review the Conjoint-Recognition Model (CRM; Brainerd & Reyna, 1998; Brainerd et al., 1999; Brainerd et al., 2001), a formalization of fuzzy-trace theory for recognition memory, as well as the GCM. We then show how exact and partial matches in make independent contributions to the matching signal and how they relate to gist and verbatim memory in CRM.

4.4.1 Conjoint-Recognition model

As we have discussed, fuzzy-trace theory is a dual-trace theory that posits that each episode leaves independent gist and verbatim memory traces. Gist traces represent the meaning and relational stimulus information, whereas verbatim traces are detailed representations that include perceptual detail and surface features. The independence of gist and verbatim trace is central to the explanatory principles of fuzzy-trace theory (p. 84 ff., Brainerd & Reyna, 2005). Specifically, it is assumed that gist and verbatim traces are encoded in parallel and retrieved independently. For example, remembering your last visit to a microbrewery (gist trace) implies nothing about the probability that you also remember details such as the fruity mango savor of the pale ale you had (verbatim trace) and vice versa. The subjective and behavioral consequences of independent retrieval, however, are asymmetric: The more informative verbatim trace presumably always trumps the gist trace. Because verbatim retrieval of your pale ale experience encompasses the corresponding gist of the episode (a visit to a microbrewery) the possible retrieval of the gist trace has no effect. In case the retrieved gist and verbatim traces are in conflict, it is assumed that the decision is always determined by the more informative verbatim trace (e.g. "I did not drive my car home; I remember stopping at a cash machine to pay the cab driver")—an instance of recollection rejection.

To measure the contribution of gist and verbatim traces to observed behavior, fuzzy-trace theory has been formalized in a measurement model, the CRM (Brainerd et al., 1999; Brainerd et al., 2001). CRM is a multinomial processing-tree (MPT) model that estimates the conditional probabilities of successful gist G and verbatim V retrieval as well as the probability of guessing old b when both retrieval processes fail (see Figure 4.1). CRM is a measurement model in the sense that it specifies the processes that contribute to the observed behavior and how they affect observed responses, but it has nothing to say about how each of these process operates. In this sense, CRM is closer to the computational than the representational and algorithmic level of analysis (Marr, 1982). We view CRM as a useful measurement tool rather than a detailed theory of recognition memory.

CRM incorporates a recollection rejection process $V^{(L)}$ (i.e., verbatim retrieval given a lure) and allows for targets and lures to differ with respect to their effectiveness as cues for gist and verbatim retrieval. The latter consideration is implemented by means of order-constraints on the model parameters. With reference to the encoding specificity principle (Tulving & Thomson, 1973), it is assumed that, compared to targets, lures are more effective retrieval cues for gist traces (i.e., $G^{(L)} > G^{(T)}$) but, conversely, that targets are more effective retrieval cues for verbatim traces, $V^{(L)} < V^{(T)}$. Guessing is assumed to affect all probe types the same, i.e. $b^{(T)} = b^{(L)} = b^{(N)}$. Due to the large number of model parameters relative to observed independent responses, CRM cannot be estimated in a standard false recognition paradigm. To increase the number of observations, it is necessary to either manipulate how participants map *old* and *new* responses onto the different probe types (Brainerd et al., 1999), to introduce a third *new but similar* response option (Stahl & Klauer, 2008), or to make simplifying assumptions.

Wherever possible, CRM assumptions have been tested and model param-

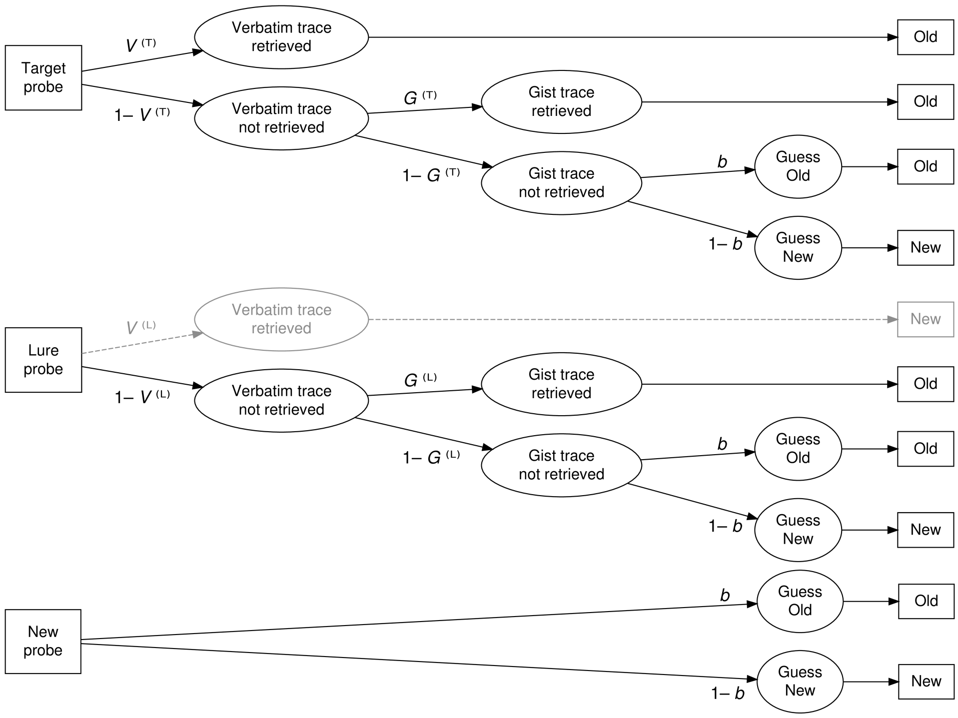

Figure 4.1: Visualization of the Simplified Conjoint Recognition model for old-new-recognition. Rectangles on the left represent memory probe types; rectangles on the right represent observed responses. Ovals represent latent cognitive states, which are sequentially traversed along the branches of the processing tree. The model parameters V, G, and b represent the conditional probabilities of state transitions for targets (T), lures (L), or new distractors (N), respectively. In the current application, we assume negligible retrieval of verbatim traces for lure probes (i.e., recollection rejection, $V^{(L)} = 0$) and comparable effectiveness of lures and targets for gist retrieval, $G^{(L)} = G^{(T)}$.

eters have been validated in selective-influence studies (e.g., Brainerd et al., 1999; Brainerd et al., 2001; Stahl & Klauer, 2008; Stahl & Klauer, 2009). Selective influence is the gold standard for testing whether a model parameter reflects a circumscribed psychological construct. The rationale is that experimental manipulations which can be assumed to target one of the psychological constructs should (1) selectively influence the corresponding model parameter but (2) none of the other parameters. If only the targeted model parameter responds to a selective influence manipulation, the parameter is thought to be a valid indicator of the psychological construct. For example, repeating targets on the study list (verbatim repetition) should selectively strengthen verbatim memory whereas increasing the number of study list items related to a lure (gist repetition) should selectively strengthen gist memory.

Despite its intuitive appeal successful selective influence does not necessitate the assumption of multiple latent causes (Dunn & Kirsner, 1988; Newell & Dunn, 2008). Because a rigorous test of the descriptive adequacy of global matching models should involve data that demonstrably involve multiple latent causes, we adopted the more rigorous approach to map out multiple latent dimensions in the bivariate state-trace space of true and false recognition (for details see Selective influence experiment). An appealing property of state-trace analysis is that it requires minimal assumptions, most notably that the observed responses are monotonically related to their assumed single latent cause. For example, we would have to assume that true and false recognition are both monotonically related to the memory strength signal generated by the matching process. While this assumption holds in old-new-recognition, it is violated in the common approaches used to estimate the full CRM, namely manipulating the mapping of response options to probe types or using an old-similar-new response format. Hence, establishing multiple latent causes via state-trace analysis requires simplifications of the full CRM.

4.4.1.1 Simplifying assumptions

We designed an experiment that allowed us to introduce simplifying assumptions to CRM which render it applicable to old-new-recognition responses. Specifically, our study lists were random sequences of photographs of everyday objects from various categories (e.g., apples, bagels, and beach balls); in all but one condition we presented multiple exemplars from each category. Photographs were randomly selected to serve as study list items, targets, lures, or new memory probes. This design justifies two simplifying assumptions without distorting the core properties of CRM. First, we assumed negligible a rate of recollection rejection, i.e. $V^{(L)} = 0$. This assumption is tenable theoretically as well as empirically. Theoretically, a lure can be rejected via recollection only if verbatim traces are retrieved for all studied category instances and is therefore unlikely to be successful or convincing when multiple category exemplars have been studied as part of a long randomly intermixed study list. Moreover, empirical estimates of recollection rejection, as measured by $V^{(L)}$, typically cannot be distinguished from 0 (Stahl & Klauer, 2008; Stahl & Klauer, 2009). Second, we assumed that targets and lures are equally effective retrieval cues for gist traces, i.e. $G^{(L)} = G^{(T)}$. Again this assumptions is tenable theoretically as well as empirically. For stimulus material, such as DRM lists, where inter-item associations are low on the study list but each item is strongly associated with the critical word, it is clear that lures cue retrieval of a list's gist more effectively than studied items. It is not clear why the same should be the case, when all items are exemplars from a common category and lures are selected at random. Accordingly, several previous studies report indistinguishable empirical estimates of $G^{(L)}$ and $G^{(T)}$ (Stahl & Klauer, 2008; Stahl & Klauer, 2009).

It is worth noting that any violations of our assumptions have predictable effects. Non-negligible contributions of recollection rejection biases estimates of $G^{(L)}$ downwards. Manipulations that selectively strengthen verbatim retrieval and improve recollection rejection would, therefore, reduce $G^{(L)}$ estimates. Better gist retrieval for lures compared to targets biases estimates $V^{(T)}$ downwards. Consequently, selectively strengthening gist retrieval would reduce $V^{(T)}$ estimates. In sum, a selective influence manipulation that has the opposite effect on V and G estimates indicates a violation of our simplifying assumptions.

The simplified model posits that true recognition results from either verbatim or gist retrieval and guessing, whereas false recognition results from gist retrieval and guessing; old-responses to unrelated new items reflect only guessing. These assumptions are reflected in the following model equations of the probabilities for old-responses to targets, lures, and new unrelated probes:

$$p(\text{“Old”} \mid \text{T}) = V + (1-V)G + (1-V)(1-G)b \tag{4.1}$$
$$p(\text{“Old”} \mid \text{L}) = G + (1-G)b \tag{4.2}$$
$$p(\text{“Old”} \mid \text{N}) = b. \tag{4.3}$$

4.4.2 Generalized context model

GCM is a single-trace theory that postulates that each episode leaves one memory trace (Nosofsky, 1986; Nosofsky, 1988; Nosofsky, 2011a). Each trace compounds item and context information. When a memory probe is presented, the similarity $\eta(x_i, y_j)$ between the ith probe x and each memory trace y is determined. The weighted sum of all J similarities is the memory strength signal, or familiarity f, which informs the recognition judgment,

$$f_i = \sum_{j=1}^{J} m_j \eta(x_i, y_j)$$

Each similarity is weighted by the memory strength m_j of the trace, which is typically assumed to decay with lag as a power-function. Hence, a large number of moderately similar memory trace may yield a strength signal that is comparable to that of a few highly similar traces. For example, you may be inclined to affirm that you tried the coffee lager at the microbrewery, either because you did, or because you tried a variety of lagers, coffee stouts, and coffee porters.

To reach a decision, the strength signal is compared to a response criterion k,

$$p(\text{“Old”} \mid x_i) = \frac{f_i^\gamma}{f_i^\gamma + k^\gamma}$$

where $k \geq 0$. The response-scaling parameter γ modulates the extremity of the predicted probabilities for any given familiarity and response criterion (McKinley & Nosofsky, 1995; Nosofsky, 2011a; Nosofsky & Zaki, 2002). Psychologically response scaling can be thought of as a response caution parameter (Nosofsky & Palmeri, 1997). That is, γ can be transformed to an error tolerance α for accidentally choosing the response, which is less likely given the probe familiarity (Navarro, 2007). As response caution increases error tolerance decreases and the response probability is increasingly biased towards 0 if $k > f_i$ or 1 if $k < f_i$. Note that alternative interpretations of γ have been suggested (e.g., Ashby & Maddox, 1993; Navarro, 2007).

In contrast to CRM, GCM specifies how memory probes are matched to the contents of memory. Each item is represented in a K-dimensional similarity space with similar items positioned close by and dissimilar items far apart. The dimensions of this space depend on the stimulus material and can represent elementary visual properties such as color, size, and line width, or higher-level semantic dimensions. The similarity space of the material and each items position are estimated by multidimensional scaling (MDS) of, for example, pairwise similarity ratings. In accordance with Shepard's law of generalization (1987), the psychological similarity between the probe and a memory trace is assumed to decrease exponentially as the scaled distance d in similarity space increases,

$$\eta(x_i, y_j) = \exp(-c \cdot d(x_i, y_j)^p)$$

where $c \geq 0$ and reflects an individual's sensitivity to dissimilarity. Increasing c decreases psychological similarity between probe and trace. The parameter p must be positive and determines the shape of function that relates distances to similarity. $p = 1$ yields an exponential similarity gradient, whereas $p = 2$ yields a Gaussian similarity gradient.

Finally, it is assumed that participants selectively attend stimulus dimensions to maximize task performance. In GCM, selective attention is formalized as a distortion of the similarity space by expanding or compressing the K dimensions and thereby modulating the distances,

$$d(x_i, y_j) = \Big[\sum_{k=1}^{K} w_k |x_{ik} - y_{jk}|^r\Big]^{1/r}$$

where w_k represents attention weights that scale the kth dimension. The distance is formalized as Minkowski distance, for which the parameter r must be positive and determines the form of the distance metric. $r = 1$ yields city block distances, whereas $r = 2$ yields euclidean distances. In most applications, p and r set to fixed values based on a prior assumptions, that is, they are typically not free parameters (Nosofsky, 2011a).

As noted previously, GCM is a successful model of recognition in long- and short-term memory (Donkin & Nosofsky, 2012b; Nosofsky & Palmeri, 2014; Nosofsky, Cox, et al., 2014; Nosofsky, Cao, Cox, et al., 2014; Nosofsky, 2011b), as well as categorization, and has previously been applied to false recognition paradigms (Zaki & Nosofsky, 2001; see also van Vugt et al., 2013). Due to the comparably detailed specification of the matching process, GCM is typically considered a process model, that is closer to the representational and algorithmic than the computational level of analysis (Marr, 1982).

4.4.3 Decision rule as a bridge

Next, we formally derive a prediction about the correspondence between CRM parameters and GCM familiarities. These predictions suggests a theoretical bridge between single-trace and dual-trace representations, that is, a bridge between the memory strength signal predicted by GCM (and possibly global matching models more generally) and gist and verbatim retrieval. Note that GCM relates to participants' responses to individual items, whereas CRM relates to aggregate responses to the three probe types.[2] The following formulas apply to individual items, but we expect them to hold in aggregate at the person level.

Owing to their different specificity, bridging CRM and GCM requires some abstraction. We, therefore, relate the models at the decision stage. Given that gist and verbatim retrieval proceed independently and in parallel, the CRM model equations formalize how the available information is used to reach a decision. Solving these equations for the model parameters yields,

[2] It is technically possible to extend the CRM modelling framework to estimate item-specific effects on model parameters. This would allow to empirically examine the formal correspondence at the level of individual responses. In the subsequent application of CRM we decided not to estimate item-specific effects for theoretical and practical reasons. As we have discussed, fuzzy-trace theory is a computational-level explanation that is silent with respect to the details of the retrieval process. Hence, item-specific effects on model parameters would estimated without regard for stimulus properties and, critically, the interactions between stimuli. Without any constraint item-specific effects would only serve a descriptive purpose without informing the model comparison. The lack of constraint also poses a practical challenge. Unless the number of items is small relative to the number of participants, estimates of item-specific effects (and their interactions) are bound to be imprecise and, thus, uninformative, due to the large number of possible combinations.

$$V = \frac{p(\text{“Old”} \mid \text{T}) - p(\text{“Old”} \mid \text{L})}{1 - p(\text{“Old”} \mid \text{L})} \qquad G = \frac{p(\text{“Old”} \mid \text{L}) - p(\text{“Old”} \mid \text{N})}{1 - p(\text{“Old”} \mid \text{N})} \qquad b = p(\text{“Old”} \mid \text{N})$$

For GCM, we ignore, for the moment, the process that produces the memory strength signal, which feeds into the decision rule. We assume three signal strengths for each probe type,

$$p(\text{“Old”} \mid \text{T}) = \frac{f^{(\text{T})^\gamma}}{f^{(\text{T})^\gamma} + k^\gamma} \qquad p(\text{“Old”} \mid \text{L}) = \frac{f^{(\text{L})^\gamma}}{f^{(\text{L})^\gamma} + k^\gamma} \qquad p(\text{“Old”} \mid \text{N}) = \frac{f^{(\text{N})^\gamma}}{f^{(\text{N})^\gamma} + k^\gamma}$$

where $f^{(\text{T})} \geq f^{(\text{L})} \geq f^{(\text{N})}$. By substitution the CRM parameters can now be re-expressed in terms of familiarities,

$$V = \frac{f^{(\text{T})^\gamma} - f^{(\text{L})^\gamma}}{f^{(\text{T})^\gamma} + k^\gamma} \tag{4.4}$$

$$G = \frac{f^{(\text{L})^\gamma} - f^{(\text{N})^\gamma}}{f^{(\text{L})^\gamma} + k^\gamma} \tag{4.5}$$

$$b = \frac{f^{(\text{N})^\gamma}}{f^{(\text{N})^\gamma} + k^\gamma}. \tag{4.6}$$

This result is consistent with the previously suggested global matching account of dissociations between true and false recognition (Arndt & Hirshman, 1998). Gist-based false recognition (as estimated by G) reflects the increased familiarity of lures ($f^{(\text{L})}$) compared to unrelated new probes, $f^{(\text{N})}$. Exact matches between probes and memory traces, by definition, are unique to true recognition. Therefore, the increased familiarity of targets $f^{(\text{T})}$ compared to lures $f^{(\text{L})}$ (as estimated by V) contributes independently to true recognition.

Conversely, this result suggests an alternative interpretation for CRM parameters. Estimates of gist activation (G) reflect the partial increment in lure familiarity attributable to partial matches. Similarly, estimates of verbatim activation (V) reflect the partial increment in target familiarity attributable to matching features that distinguish lures from studied items—or lures from targets, respectively.

Because our analysis is limited to the decision rule the above relationship between CRM and GCM is in some sense trivial. However, with some additional assumptions, the above equations can be further examined at the level of inter-item similarities, see Appendix B.1 for details. The following simplified expressions only apply to the special case where $\gamma = 1$ and $m_j = m_{j'}$, but highlight the separable contributions of exact and partial matches. We assume that the study list is composed of sublists of related items and that any item can serve as memory probe, as is the case in study lists composed of exemplars from various categories (e.g., Buchanan et al., 1999; Pierce et al., 2005; Stark et al., 2013; Andermane & Bowers, 2015). Note that these are the same conditions that justify our simplifying assumption that $G^{(L)} = G^{(T)}$ in the CRM. Under these conditions the numerator in Equation (4.5) expands to

$$f^{(L)} - f^{(N)} \approx \sum_{p=1}^{P} \underbrace{\eta(L, r_p)}_{\text{partial match}} - \eta(N, r_p) \tag{4.7}$$

where L is a lure, N a new unrelated distractor, and r_p are traces of study list items from the same sublist as the lure. The summed similarities between lure L and related traces r_p represents the contribution of partial matches, which is corrected for the summed similarity of the same traces in response to a new unrelated distractor N.

Similarly, the numerator in Equation (4.4) can be specified as

$$f^{(T)} - f^{(L)} \approx O[\underbrace{\eta(T, T^*)}_{\text{exact match}} - \eta(L, T^*)] \tag{4.8}$$

where O is the number of identical T^* traces from repeated presentations during study. The summed similarity between target T and its corresponding traces T^* represents the contribution of exact matches, which is corrected for the summed similarity of the same traces in response to a lure L. The correction yields an estimate of the incremental contribution of exact matches over partial matches. When encoding is noise-free, $\eta(T, T^*) = \eta(T, T) = 1$ and further simplifies to

$$f^{(T)} - f^{(L)} \approx O[1 - \eta(L, T^*)]$$

where $1 - \eta(L, T^*)$ is the dissimilarity between L and T^*. This result has two implications: First, because every episode leaves a new trace in memory the corrected familiarity of targets should increases linearly with the number of target presentations. This constraint can, however, be relaxed by assuming that repeated study increases the memory strength of targets. Second, the above shows that traces of related study list items (i.e., items from the same sublist, r_p) are irrelevant to the numerator of V—they do however factor in the denominator. As Equations (4.7) and (4.8) show, in terms of GCM, G estimates the incremental familiarity of lures due to features shared by sublist items, whereas V estimates the incremental familiarity due to features that distinguish between lures and target traces.

The above predictions contradict fuzzy-trace theory and previously reported results with respect to the effect of the selective influence manipulations. Fuzzy-trace theory predicts that repeating a target on the study list should increase estimates of V but not G. According to Equations (4.5) and (4.7), additional target traces, like any other related traces, increase $f^{(L)}$, which in turn increases estimates of G. Hence, GCM predicts that target repetitions increase true as well as false recognition.

More recent memory models posit that memory traces become differentiated and, thus, less similar to related items with repeated study (Criss, 2006; Criss & Koop, 2015; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998). In GCM, differentiation may be thought of as increased dissimilarity sensitivity c for target traces T^*. Such sharpening of target traces could prevent an increase in G and false recognition through a reduction of $\eta(L, T^*)$. In other words, differentiation would allow true recognition to be selectively influenced, see Appendix B.1.1. Thus, it is to be expect that differentiation may be necessary to account for the selective influence of target repetitions.

Fuzzy-trace theory further predicts that increasing the number of related items on the study list increases estimates of G but not V. GCM predicts that additional related items increase both lure familiarity $f^{(L)}$ and target familiarity $f^{(T)}$. According to Equation (4.4) the increase in $f^{(T)}$ may yield reduced estimates of V because the relative contribution of exact matches to target familiarity decreases relative to the contribution of partial matches. The extent of the reduction in V depends on the relative magnitude of exact and partial match contributions to $f^{(T)}$ and may additionally be modulated by response scaling, i.e., allowing for $\gamma > 1$. Hence, we speculate that constraining $\gamma = 1$ may limit GCM's ability to account for selective influence manipulations.

The plausibility of the derived expressions is conditional on GCM's ability to describe true and false recognition as well as the less constrained CRM. Assessing the hypothesized correspondence between CRM and GCM, thus, requires an empirical test of the descriptive adequacy of GCM for true and false recognition. As the previous discussion indicates, a selective influence study of CRM parameters, which allows to distinguish between multiple latent causes of true and false recognition, constitutes a challenging empirical test for GCM.

4.5 Interim summary

Empirical data strongly suggest that multiple latent causes contribute to true and false recognition but the nature of these causes continues to be debated. An important disagreement is between single- and dual-trace theories. Fuzzy-trace theory assumes that independent storage and retrieval of verbatim or gist traces for each episode are responsible for dissociations between true or false recognition. Global matching models, on the other hand, assume that each episode leaves a single trace in a unitary storage. At test, exact and partial matches between probes and memory traces make separable contributions to the mnemonic strength signal and recognition responses. There is considerable conceptual overlap between the distinctions of partial versus exact matches and of gist versus verbatim traces, and thus, between the posited latent dimensions: Gist may be conceptualized as a partial match between a probe and traces in memory, whereas verbatim information may be thought of as exact matches. If there is, indeed, a close correspondence between these concepts, global matching model should be able to account for the effects of selective influence manipulations on the gist and verbatim retrieval. Such a finding would suggest that V and G parameters in CRM may estimate the contributions of exact and partial matches to probe familiarity rather than independent retrieval of verbatim and gist traces. We ran an experiment to test this possibility.

4.6 Selective influence experiment

We designed a selective influence experiment that conforms to the assumptions of our model specifications. The presented study list consisted of sublists of photographs of exemplars from everyday object categories. For each participant exemplars were randomly selected to serve as target, lure, or new unrelated probe. To selectively influence gist retrieval we manipulated the number of category exemplars on the study list (*gist repetition*) and to selectively influence verbatim retrieval (*verbatim repetition*) we presented targets repeatedly during study (p. 111-112, Brainerd & Reyna, 2005). Both manipulations have been used in previous selective influence studies (e.g., Stahl & Klauer, 2008; Stahl & Klauer, 2009). To verify the selective influence of our manipulation on gist and verbatim retrieval, we estimated gist and verbatim retrieval using our simplified CRM. Further, to establish that our manipulations create a challenging test bed for the GCM, we performed a state-trace analysis to test whether our results necessitate multiple latent causes (Dunn & Kirsner, 1988; Newell & Dunn, 2008). Taking into account empirical estimates of inter-item similarities, we then fit GCM to the observed responses to test the model's descriptive adequacy (Johns & Jones, 2010). To foreshadow the results, our selective influence manipulation was successful and GCM produced acceptable fits to the observed data.

4.6.1 Method

We performed an old/new-recognition experiment with selective influence manipulations targeting G and V parameters of the CRM.

4.6.1.1 Participants

Sixty students of the University of Cologne, sampled from our lab database, participated in the experiment in exchange for 7€ or course credit (see Table 4.1 for sample demographics). All participants provided informed consent.

4.6.1.2 Material

We used categorized photographs of everyday objects from the *Massive Memory* database (Brady et al., 2008). The database contains 240 categories,

Table 4.1: Summary of participant demographics.

		Age						
Category size	n	Mean	SD	Min	Max	German	Female	Psychology
4 vs. 8	30	25.33	7.33	19	46	93.33	73.33	80.00
1 vs. 11	30	26.90	9.20	18	50	93.33	83.33	86.67

consisting of 16 exemplars each, and provides estimates of within-category inter-items similarities derived from MDS (Hout et al., 2014). Photographs from different categories served as primacy and recency buffers.

4.6.1.3 Procedure

The experimental procedure consisted of study, retention, and test phases. During the study and test phase, participants worked on computers in individual booths and instructions were displayed on the computer screen. We instructed participants to memorize a series of photographs as we would later test their memory. Study lists consisted of 1280 items selected from a subset of 160 categories and was enclosed by eight primacy and eight recency buffer items. The subset of 160 categories as well as the exemplars from each category were selected at random for each participant. We varied the number of target presentations (once vs. five times) to selectively influence verbatim retrieval as assessed by the V parameter in CRM. We further varied the category size, i.e., the number of exemplars from each category, to selectively influence gist retrieval as assessed by the G parameter in CRM. Both manipulations were within-participants. The strength of the category size manipulation differed between two groups of participants: One half studied either 4 or 8 exemplars per category, the other half studied either 1 or 11 exemplars. Hence, the study list consisted of the same number of categories and items in both groups. The subset of 160 categories was equally distributed among the within-participant conditions at random. Each photograph was presented for 1 s with an inter-stimulus interval of 200 ms totaling 27 minutes.

Following a 25 minute retention interval, in which participants worked on a paper-pencil intelligence test, we administered the recognition test. We presented one exemplar from all 240 categories in an old-new-response format. The exemplar from each category served one of three probe types: target, lure, or new unrelated distractor. A target probe was an exemplar that was included on the study list, whereas a lure was a randomly selected exemplar from a study-list category that had not been presented during the study phase. New unrelated distractors were exemplars from the 80 categories that were not used to compose the study phase. We used only one exemplar per category during the recognition test. Thus, we collected 20 responses to targets and lures in each of the four within-participant conditions and 80 responses to new unrelated distractors . New unrelated distractor categories can not be matched to a target presentation or category size condition because we presented no exemplars from these categories. The probe type manipulation was, thus, not fully crossed with the selective influence manipulations.

To summarize, the experiment realized a 2 *probe type* (target, lure) × 2 *target presentations* (once, five times) × 2 *category size* (small, large) × 2 *category size strength* (4 vs. 8, 1 vs. 11) design. We varied the strength of the category size manipulation between participants. New unrelated distractors constituted an additional probe type condition that was not crossed with the other within-participant factors.

4.6.1.4 Data analysis

We report Bayes factors as relative measures of evidence for our statistical and cognitive models (e.g, Wagenmakers et al., 2010). The Bayes factor may be equally interpreted as a models prior predictive accuracy or as relative likelihood of the observed data given the model and the specified prior distributions on its parameters. Hence, Bayes factors are directly interpretable as graded measures of evidence.

We used R (Version 3.6.3; R Core Team, 2017) and the R-packages *bridgesampling* (Version 0.6.0; Gronau & Singmann, 2018), and *papaja* (Version 0.1.0.9997; Aust & Barth, 2017) for all our analyses.

4.6.1.4.1 State-trace analysis The fact that experimental manipulations selectively influence model parameters does not rule out that the observed data can be explained by a single latent cause (Dunn & Kirsner, 1988; Newell & Dunn, 2008). To test whether our results rule out a single latent cause, we performed state-trace analyses (Dunn & Kirsner, 1988). State-trace analysis allows testing hypotheses about the latent dimensionality of the observed data with minimal assumptions. Applied to true and false recognition (i.e., old-responses to targets and lures) the logic goes as follows: Assume that true and false recognition are mediated by a single latent dimension, such as a unidimensional mnemonic strength signal. Further assume that the probability of endorsing targets and lures as old is a monotonic function of this latent dimension. Then it follows that true and false recognition must be monotonically related, i.e., the rank order of condition means should be the same. The rank order of conditions for true and false recognition can be visualized in a bivariate state-trace plot—a scatter plot of condition means. Any violation of monotonicity necessitates the assumption of at least one additional, functionally independent, latent dimension.

State-trace analysis requires specification of three analytic factors. The *state factor* consists of two levels that make up the axes of the bivariate state-trace plot—here true and false recognition. The *dimension factor* maps to an experimental manipulation that interacts with the state factor to affect the latent dimensionality, e.g. by creating a single dissociation. According to fuzzy-trace theory, the number of target presentations should only affect true recognition. Hence, we defined the number of target presentations as dimension factor. The *trace factor* constitutes an auxiliary experimental manipulation. Because two data points are non-diagnostic with respect to monotonicity the trace factor is used to generate additional data points. The trace factor is assumed to have a monotonic effect on both states and in both dimensions. We defined the category size as trace factor because CRM and GCM agree that true and false recognition should increase with category size irrespective of the number of target presentations.

Monotonic regression and null hypothesis significance testing (NHST) can be used to test whether true and false recognition are non-monotonically related (e.g., Kalish et al., 2016). This approach, however, is problematic for two reasons. Models of non-monotonic relationships are clearly more flexible than models of monotonic relationships because no rank order constraints apply. It is not clear that NHST adequately penalizes non-monotonic models for their greater flexibility. Secondly, existing NHST approaches do not readily provide evidence for the equality of rank orders and are therefore not well-suited to support the hypothesis of a single latent dimension. We, therefore, performed Bayesian inference by model comparison using Bayes factors (Prince et al., 2012; Davis-Stober et al., 2016). Bayes factors naturally take model flexibility into account and quantify evidence in favor of either hypothesis.

The Bayesian approach further allows to incorporate additional assumptions based on prior knowledge. Specifically, based on prior experimental evidence the monotonic effect of the trace factor (category size) can be enforced as an additional order constraint on the condition means in both the monotonic and non-monotonic model (Davis-Stober et al., 2016). The additional trace constraint focuses hypothesis testing on the model characteristics of interest and increases the diagnosticity of the model comparison. We verified the consistency of our data with the trace assumption by comparing a trace to a non-trace model, which did not enforce a monotonic effect of the trace factor. We analyzed true and false recognition rates on the unlimited probit scale, placed uniform priors on all condition rank orders (Davis-Stober et al., 2016), and analyzed the two groups of participants separately.

Because neither monotonicity nor non-monotonicity in state-trace space are necessarily preserved when responses are averaged across participants we analysed participants' data separately (Prince et al., 2012). Individual consideration of participants implies the question whether conclusions hold for all participants. For example, we may ask "Do participants' responses unanimously violate monotonicity and hence support multiple latent dimensions?" Davis-Stober et al. (2016) recently proposed a non-parametric aggregated Bayes factor (ABF) to test this hypothesis. In short, averaging of non-monotonic outcome spaces can yield some but not *every* monotonic shape. If the averaged data exhibit a rank order that cannot result from individual non-monotonic rank orders, we can conclude that at least one participant's responses favor a single latent dimension. Unfortunately, this test of homogeneity is asymmetric. If, conversely, the aggregated data exhibit a rank order that is consistent with individual non-monotonic rank orders, we cannot infer that the data of all participants unanimously favor multiple latent dimensions. It does, however, provide a minimal test of the unanimity assumption required to compute the group Bayes factor (GBF) that synthesizes the evidence by multiplying Bayes factors for individual participants (Prince et al., 2012; Davis-Stober et al., 2016; also see Klaassen et al., 2018). We additionally report the averaged evidence, that is the geometric mean of the group Bayes factor (gGBF), which does not rely on the unanimity assumption (p. 2283; Klaassen et al., 2018). gGBF provides an estimate of the evidence expected for newly sampled participants.

4.6.1.4.2 Cognitive modelling We implemented CRM and GCM as Bayesian hierarchical models and estimated the joint posterior distribution of parameters by No-U-Turn sampling. To ensure convergence to the stationary distribution we monitored sampler diagnostics and set a criterion of $\hat{R} < 1.01$. To compare models using Bayes factors we estimated their marginal likelihoods using Warp-III bridge sampling (Gronau et al., 2017; Gronau & Singmann, 2018). We quantified the estimation error from 25 repeated sampling runs and report 2.5% and 97.5% qunatiles when the estimation error exceeds 5%. Note that comparisons of non-nested models was not possible because CRM was fit to responses aggregated across items (Binomial likelihood) whereas GCM was fit to responses to individual items (Bernoulli likelihood). For prior settings and details on quantitative model fit assessment see Appendix B.2.

4.6.1.4.2.1 Conjoint recognition model We implemented the hierarchical CRM using the latent-trait approach with random participant effects (Klauer, 2010). We assumed that the probability of guessing "old" was unaffected by our category size strength manipulations and, thus, constrained b to be equal across groups. To evaluate the success of our selective influence manipulations we additionally specified order-constraint model variants by reparameterization. Priors for the reparameterized model were latent-trait approximations to the uninformative priors derived by Heck and Wagenmakers (2016). Order- and equality-constraints were enforced at the participant level where possible.

In the application of MPT models it is customary to test whether the model predictions deviate from the observed category frequencies. We visually compared posterior predictive distributions of each model to the observed data (Chapter 6, Gelman et al., 2015). We additionally quantified model fit by computing posterior predictive p values from summary T statistics (Equations 17 and 18, Klauer, 2010) for participant means, condition means, and the variance-covariance structure.

4.6.1.4.2.2 Generalized context model The model structure of our hierarchical GCM implementation closely followed that of the latent-trait CRM. Because GCM parameters are constrained to be positive, we transformed latent parameters using the exponential function—instead of the probit transformation used to obtain probabilities in CRM. Hence, we implemented GCM as a hierarchical log-normal model.[3]

As noted above, in the calculation of the psychological similarity, some parameters are typically set to fixed values based on prior assumptions (Nosofsky, 2011a). The parameter r determines the distance metric. Following the standard assumption for integral stimulus dimensions, we set $r = 2$ yielding the Euclidean distance (e.g., Nosofsky, 2011a). We further assumed that our 1 s presentation duration was sufficient to yield largely accurate, distinguishable stimulus representations. We, therefore, specified an exponential similarity gradient by setting $p = 1$ (Nosofsky, 2011a).

Besides these standard assumptions, we introduced two additional simplifying assumptions. Ideally, one MDS solution of the entire stimulus space would be derived to estimate the dimensionality of the psychological similarity space K and the pairwise distances between stimuli $d(x_i, y_j)$. However, deriving this solution for the large stimulus set at hand is impractical (see Nosofsky et al., 2018, for an impressive effort to derive a solution for 360 photographs of rocks from 30 subtypes). Hout et al. (2014) provide similarity ratings for exemplar pairs within each category of our stimulus material—albeit not for stimulus pairs from different categories. Hence, we estimated stimulus dimensionality and distances from category-wise MDS solutions. Based on visual inspections of plots of Kruskal's Stress, explained variance, and mean correlations of distances we assumed that all similarity spaces can be approximated by three component dimensions, i.e., $K = 3$. Without distance estimates for stimuli from different categories the model cannot predict old-responses to new unrelated distractors, i.e., exemplars from new categories. We therefore estimated an additional auxiliary parameter d_{inter}, common to all participants, as a stand-in for any $d(x_i, y_j)$ where x_i and y_j are exemplars from different categories. Owing to the category-wise MDS the derived component dimensions are not com-

[3] We also implemented GCM as a hierarchical Gamma model (Bartlema et al., 2014) and compared the results for the most important model variants. The results were highly similar. Because the log-normal model structure more readily allows for the incorporation of parameter correlations and is computationally more tractable, we decided to adopt this approach.

parable across categories. Practical considerations aside, it seemed implausible that one set of attentional weights is optimal for all categories. We, therefore, made the simplifying assumption that participants attended all dimensions equally, i.e. $w_k = 1$. The assumption that participants optimally distribute attention across the component dimensions is at the core of GCM (e.g., Nosofsky, 1991). Although some recent studies indicate that in some applications attention weights play only a minor role (e.g., Nosofsky et al., 2011), setting $w_k = 1$ may impair the models ability to fit the data. Finally, we assumed that the long retention interval as well as primacy and recency buffers largely mitigate differential effects of forgetting on memory traces and set $m_j = 1$. The resulting model has three parameters per participant: a dissimilarity sensitivity c, a response criterion k, and the response-scaling parameter γ.

Our theoretical model analysis suggests that GCM may need to allow for traces to become differentiated through repeated study to account for selective influence manipulations on V. Whereas differentiation is the consequence of mechanistic assumptions in more recent memory models (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997), differentiation is most readily incorporated into GCM in a more descriptive fashion. Specifically, we allowed for increased dissimilarity sensitivity $c' = c\delta$ for target traces T^* when targets were presented five times, see Appendix B.1.1. Although in most previous applications c was held constant for all exemplars, recent work suggests that it may vary with experimental factors, such as study-test lag (Nosofsky et al., 2011; Nosofsky & Gold, 2016; Nosofsky & Gold, 2016). Finally, we considered the possibility that repeating targets on the study list could also increase memory strength, that is, we added an additional parameter $m' = m\beta$ for target traces T^* when targets were presented five times (cf. Nosofsky et al., 2011; Nosofsky & Gold, 2016). The most complex model, thus, had five free parameters per participant in addition to d_{inter}, which was common to all participants.

Because we suspected that freely varying c and m with the number of target presentations would grant GCM an undue flexibility, we explored another model variant in which we constrained the parameters to be equal, i.e. $\lambda = m = c$ and $\lambda' = m' = c'$. When c and m are estimated independently for a number of study-test lags, the parameters tend to be highly correlated (Nosofsky et al., 2011; Nosofsky & Gold, 2016). It is an established finding that the decrease in memory strength with lag follows a power function—

the "power-law of memory strength" (Donkin & Nosofsky, 2012a). A close inspection of the c estimates varying with study-test lag reveals that their decrease is also well described by a power-function. What is more, when c and m are put on a common scale, their power-functions are almost indistinguishable, Figure 4.2. Because the scale of m is essentially arbitrary, we reasoned that an equality constraint may be tenable. Incidentally, setting memory strength equal to dissimilarity sensitivity can be interpreted as scaling the exponential similarity gradient such that it conforms to the probability density function of the exponential distribution. This has the psychologically plausible implication that an increase in c reduces the activation of dissimilar or moderately similar memory traces and simultaneously increases activation for very similar or identical memory traces [cf. Criss (2006); also see Figure 4.7].

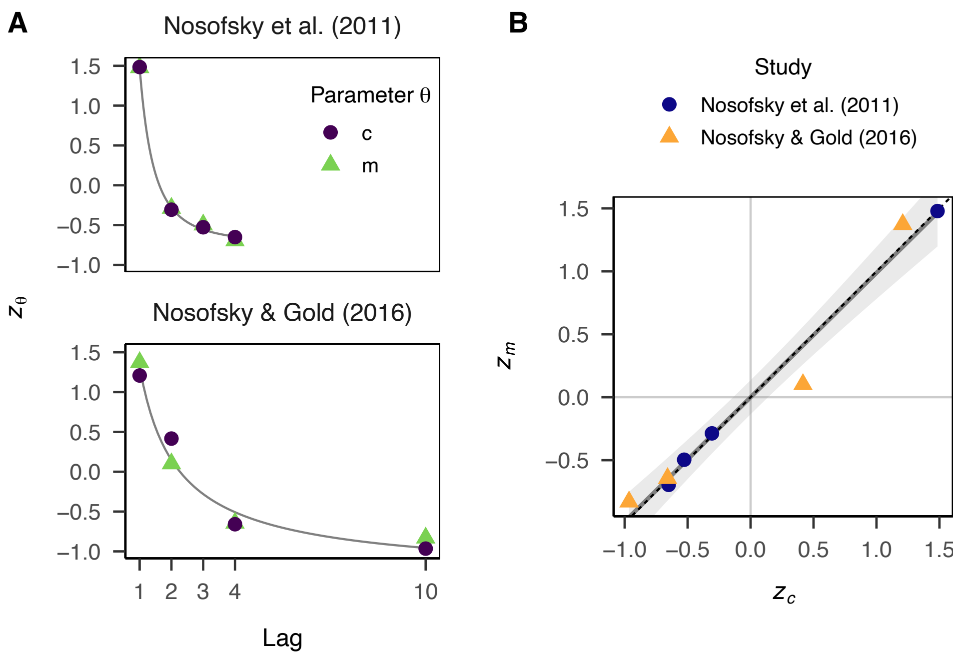

Figure 4.2: Comparison of dissimilarity sensitivity c and memory strenght m estimates reported by Nosofsky et al. (2011) and Nosofsky and Gold (2016). Parameter estimates are z-standardized to ensure comparable scaling. Both parameters decrease with study-test lag according to a common power function (**A**). Consequently, the estimates are almost identical, $R^2 = .98$, 90% CI $[0.89, 0.99]$, $F(1,6) = 239.04$, $p < .001$ (**B**). The solid line represents the estimated regression line with confidence interval; the dashed line represents the main diagonal with intercept $b_0 = 0$ and slope $b_1 = 1$.

4.6.2 Results

We first present the results of our state-trace analysis before proceeding with the results from applying the cognitive models.

4.6.2.1 State-trace analysis

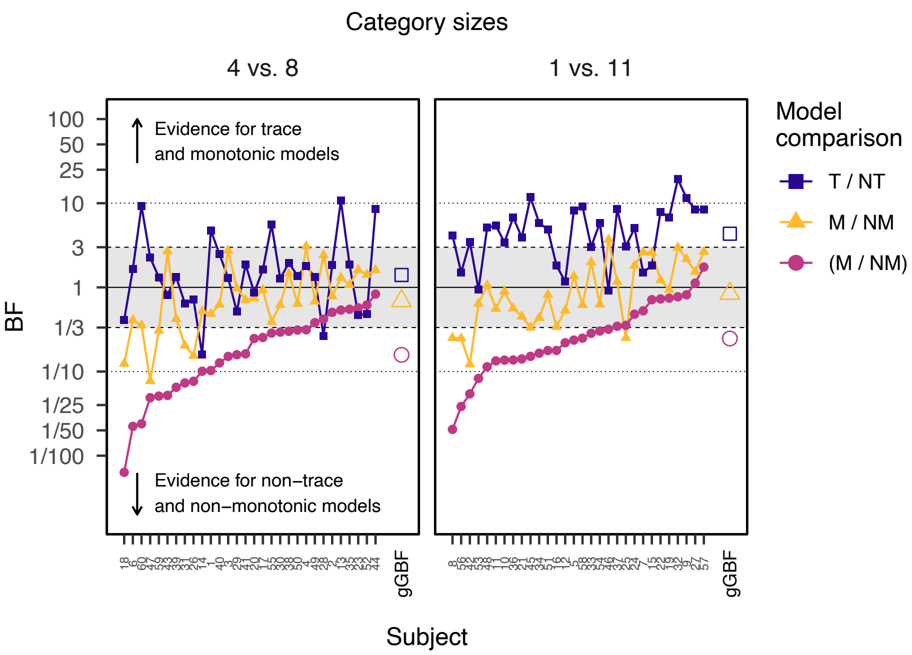

Figure 4.3: Bayes factors for individual participants in favor of the trace model (T) representing the assumption that true and false recognition always increase as studied category size increases in both dimensions relative to the non-trace model (NT), which posits a violation of the trace assumption in at least one dimension (T vs. NT) or in favor of a monotonic model of true and false recognition (M) relative to a non-monotonic model (NM) with (M vs. NM | T) and without incorporating the trace assumption (M vs. NM). Higher values indicate support for trace and monotonic models. Participants are ordered according to their evidential value for the monotonic relative to the non-monotonic model given the trace assumption. Big unfilled points represent the geometric mean of the group Bayes factor (gGBF).

The results of the Bayesian state-trace analysis of the probit-transformed proportions of participants' old-responses are shown in Figures 4.3. While the data of most participants supported the non-monotonic model and hence multiple latent dimensions the evidence was modest. To make the model comparisons more informative we incorporated the trace assumption into the monotonic and non-monotonic models. The assumption is that increasing the number of exemplars in a category increases true and false recognition regardless of whether the target was presented once or five times. In line with our prior expectations, this additional assumption was almost unanimously supported in the group that received the strong category size manipulation, which contrasted 1 to 11 exemplars per category, $\mathrm{gGBF}^{(1\text{ vs. }11)}_{\mathrm{T/NT}} = 4.34$. In the group that received the weak category size manipulation, which contrasted 4 to 8 exemplars per category, the evidence was ambiguous, $\mathrm{gGBF}^{(4\text{ vs. }8)}_{\mathrm{T/NT}} = 1.40$. However, when combined across participants the evidence overwhelmingly favored the trace over the non-trace model $\mathrm{GBF}^{(4\text{ vs. }8)}_{\mathrm{T/NT}} = 26,152.92$- and $\mathrm{GBF}^{(1\text{ vs. }11)}_{\mathrm{T/NT}} = 1.34 \times 10^{19}$-to-1, respectively. We additionally tested whether the groups' aggregated data indicated that any participant's responses violated the trace assumption. For the strong category size manipulation the aggregated data provided evidence against individual violations $\mathrm{ABF}^{(1\text{ vs. }11)}_{\mathrm{T/U}} = 14.01$. For the weak category size manipulation the model comparison was inconclusive, $\mathrm{ABF}^{(4\text{ vs. }8)}_{\mathrm{U/T}} = 1.98$. Taken together, these results support the trace assumption.

Incorporation of the trace assumption increased the evidence for the non-monotonic model for all except one participant yielding almost unanimous support for the non-monotonic model and hence multiple latent dimensions. Across both category size strength conditions only two participants' responses favored the monotonic model, albeit weakly, $\mathrm{gGBF}^{(4\text{ vs. }8)}_{\mathrm{(NM/M)\mid T}} = 6.33$ and $\mathrm{gGBF}^{(1\text{ vs. }11)}_{\mathrm{(NM/M)\mid T}} = 4.05$. Taken together the evidence of all participants' overwhelmingly favored the non-monotonic model $\mathrm{GBF}^{(4\text{ vs. }8)}_{\mathrm{(NM/M)\mid T}} = 1.12 \times 10^{24}$- and $\mathrm{GBF}^{(1\text{ vs. }11)}_{\mathrm{(NM/M)\mid T}} = 1.67 \times 10^{18}$-to-1, respectively. We again analysed the aggregated data to test whether any participants' responses violated non-monotonicity. For the strong category size manipulation the aggregate data provided evidence against individual violations $\mathrm{ABF}^{(1\text{ vs. }11)}_{\mathrm{(NM\mid T)/U}} = 14.02$. For the weak category size manipulation the model comparison was inconclusive, $\mathrm{ABF}^{(4\text{ vs. }8)}_{\mathrm{U/(NM\mid T)}} = 1.97$[4]. In sum, these results suggest that multiple latent dimensions underlie the observed pattern of true and false recognition. Hence, our data provide a challenging test for

[4]Note that the aggregate Bayes factors for the trace and the non-monotonic constraints are very similar because the models make largely the same predictions. The convex hull of the the trace constraint encompasses the convex hull of the non-monotonic constraint and consists of only 2.47% unique extremal vertices.

Table 4.2: Median posterior parameter estimates of the population probability medians $\widetilde{\mu}^{(\theta)}$ (95% HDI in parentheses) for the minimally constrained Conjoint Recognition model $\mathcal{M}_0^{\mathrm{CRM}}$.

Target presented	Category size	V	G
Once	4	0.33 [0.25, 0.41]	0.12 [0.08, 0.16]
	8	0.32 [0.24, 0.40]	0.22 [0.17, 0.26]
Five times	4	0.91 [0.87, 0.94]	0.11 [0.07, 0.15]
	8	0.87 [0.80, 0.92]	0.24 [0.19, 0.28]
Once	1	0.28 [0.20, 0.36]	0.04 [0.01, 0.07]
	11	0.32 [0.21, 0.42]	0.30 [0.24, 0.36]
Five times	1	0.89 [0.83, 0.94]	0.09 [0.05, 0.12]
	11	0.89 [0.84, 0.94]	0.27 [0.21, 0.33]

Note. We constrained the populations means of the parameter b for guessing "old" to be equal across groups, $\widetilde{\mu}^{(b)} = .02$ 95% HDI [.01, .03].

the single-trace GCM.

4.6.2.2 Cognitive models

4.6.2.2.1 Conjoint-recognition model We first fit the simplified conjoint recognition model to the data. As a baseline we fit an unconstrained model $\mathcal{M}_0^{\mathrm{CRM}}$ with $\mu^{(V)}$, V, and $\mu^{(G)}$ parameters free to vary across groups and conditions. We compared this baseline model to constrained models to assess the effect of our selective influence manipulations. First, we constrained V to increase from one than five target presentation within participants but to be unaffected by category size, $\mathcal{M}_V^{\mathrm{CRM}}$. We specified an additional model, which constrained $\mu^{(V)}$ of the corresponding target presentation conditions to be equal across groups, $\mathcal{M}_{V\times}^{\mathrm{CRM}}$. Second, we constrained G to be unaffected by the number of target presentations, but to increase with category size within participants, $\mathcal{M}_G^{\mathrm{CRM}}$. Again, we specified an additional model, which constrained $\mu^{(G)}$ to also increase with category size across groups, $\mathcal{M}_{G\times}^{\mathrm{CRM}}$. Finally, to assess the joint success of our selective influence manipulations we combined both constraints, $\mathcal{M}_{V\times G\times}^{\mathrm{CRM}}$.

The CRM with guessing parameters $\mu^{(b)}$ constrained to be equal across groups was able to describe the data, suggesting that varying the strength of the category size manipulation had no noticeable effect on participants' guessing behavior (Figures 4.4 and B.2).

The estimates of the group-level model parameters suggested that our selective influence manipulations had the predicted effects, Table 4.2. First, presenting targets repeatedly increased estimates of average verbatim memory whereas increasing the category size did not. This conclusion was confirmed by our model comparisons. Relative to the unconstrained model, the data supported the group-wise selective influence of increasing the number of target presentations on V by $\mathrm{BF}_{V/0} = 4.70 \times 10^{12}$ $[9.04 \times 10^{11},\ 4.87 \times 10^{13}]$-to-1. Additionally constraining $\mu^{(V)}$ to be equal across groups for one and five target presentations was supported $\mathrm{BF}_{V\times/V} = 45.10$ [15.42, 121.68]-to-1. Conversely, increasing the category size increased estimates of average gist memory whereas presenting targets repeatedly did not. Again, this conclusion was confirmed by our model comparison Relative to the unconstrained model, the data supported the group-wise selective influence of increasing increasing the category size on G by $\mathrm{BF}_{G/0} = 7.56 \times 10^{7}$ $[7.10 \times 10^{6},\ 9.88 \times 10^{8}]$-to-1. Additionally constraining $\mu^{(G)}$ to increase with category size across groups was supported $\mathrm{BF}_{G\times/G} = 2.66$ [0.24, 26.97]-to-1. Jointly, all constraints were supported by the data $\mathrm{BF}_{V\times G\times/0} = 7.02 \times 10^{21}$ $[1.11 \times 10^{21},\ 4.87 \times 10^{22}]$-to-1 relative to the unconstrained model and $\mathrm{BF}_{V\times G\times/V\times} = 3.31 \times 10^{7}$ $[1.38 \times 10^{7},\ 9.03 \times 10^{7}]$-to-1 or $\mathrm{BF}_{V\times G\times/G\times} = 3.49 \times 10^{13}$ $[5.19 \times 10^{12},\ 1.48 \times 10^{14}]$-to-1 relative to the models constraining either V and $\mu^{(V)}$ or G and $\mu^{(G)}$, respectively. Visual inspection (Figure 4.4) and quantitative assessment of posterior predictions indicated that all models describe the data adequately, Figure B.2.

Of all selective influence effects, only the the increase of $\mu^{(G)}$ with category size across groups ($\mathscr{M}^{\mathrm{CRM}}_{G\times}$) was not clearly favored over the the model that enforced the order constraint only for G within participants. The estimates of the unconstrained model suggest that this may reflect an interaction with the number of target presentations. The estimates of $\mu^{(G)}$, as predicted, increase with category size when targets were presented once. When targets were presented five times, however, estimates of $\mu^{(G)}$ were more similar for small (1 or 4) and large categories (8 or 11). However, given that the estimated evidence covered a range from inconclusive to moderately strong we would caution not to over-interpret this descriptive difference.

Taken together these results suggest that the multiple latent dimensions identified by our state-trace analysis are consistent with independent ver-

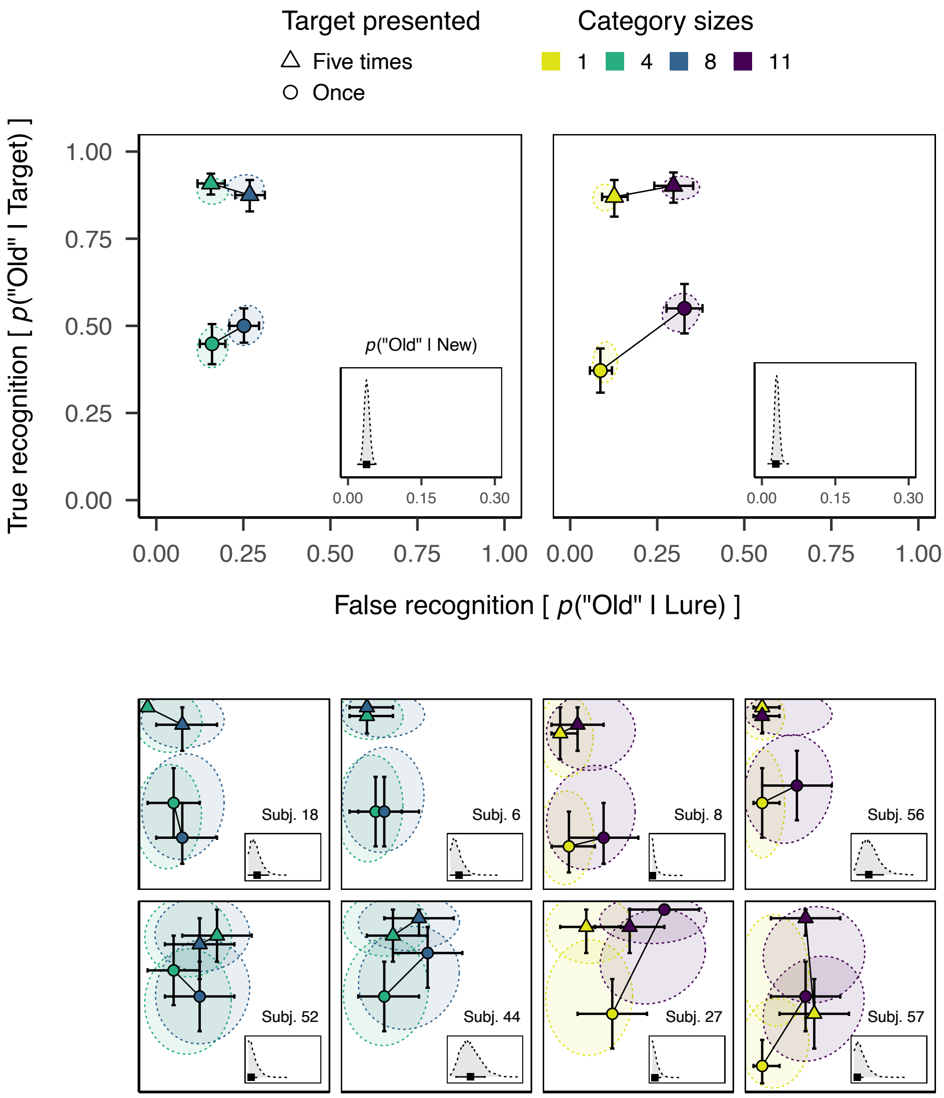

Figure 4.4: State-trace plot of averaged (top) and individual responses (bottom) and posterior predictions of $\mathscr{M}^{\text{CRM}}_{V\times G\times}$. From each group we show the two participants with strongest support for the non-monotonic (top row) and monotonic models (bottom row, see Figure 4.3). Points represent average observed rates of "Old"-responses; error bars indicate 95% bootstrap confidence intervals based on 10,000 bootstrap samples. Ellipses represent multivariate normal-approximations to 95% credible regions posterior predictions. The inset shows the proportion of "Old"-responses to unrelated new probes; kernel density estimates represent the posterior predictions.

batim and gist representations.

4.6.2.2.2 Generalized context model Global matching models posit that the multiple latent dimensions are the result of separable contributions of exact and partially matched memory traces. To assess the adequacy of this alternative account, we next fit the GCM to the data. As a baseline, we fit an unconstrained model $\mathcal{M}_0^{\mathrm{GCM}}$ with parameters $\mu^{(c)}$, $\mu^{(c')}$, $\mu^{(m')}$, $\mu^{(k)}$, and $\mu^{(\gamma)}$ free to vary across groups. We compared this baseline model to three constrained models. First, we constrained the population means of all parameters to be equal across groups $\mathcal{M}_\times^{\mathrm{GCM}}$ to test whether the strength of the category size manipulation affects retrieval or decision processes rather than being fully attributable to the contents of participants' memory. Based on our theoretical model analysis, we additionally tested whether response-scaling and differentiation are necessary to account for the selective influence manipulations. We constrained $\gamma = 1$, effectively removing response scaling from the model ($\mathcal{M}_\gamma^{\mathrm{GCM}}$) and we constrained $\delta = 1$ ($\mathcal{M}_\delta^{\mathrm{GCM}}$) or $\beta = 1$ ($\mathcal{M}_\beta^{\mathrm{GCM}}$). Finally, we compared these models to the model that constrained memory strength and dissimilarity sensitivity to be equal, $\mathcal{M}_\lambda^{\mathrm{GCM}}$.

The estimates of the group-level model parameters suggested that the strength of the category size manipulation had little effect on the model parameters, Table 4.3. Indeed constraining the population means of all parameters to be equal across groups was supported by the data $\mathrm{BF}_{\times/0} = 6.28 \times 10^5$ $[1.32 \times 10^4,\ 2.07 \times 10^7]$-to-1. In line with our theoretical model analysis, differentiation was favored $\mathrm{BF}_{0/\delta} = 1.06 \times 10^{41}$ $[2.37 \times 10^{40},\ 5.57 \times 10^{41}]$-to-1 relative to the model without it. Inspection of the posterior predictions confirmed differentiation was necessary to account for the selective influence of repeated target presentations on true recognition. Repeated target presentations also affected memory strength; the data favored an increase in memory strength for targets that were presented five times $\mathrm{BF}_{0/\beta} = 1.29 \times 10^{26}$ $[6.96 \times 10^{24},\ 1.74 \times 10^{28}]$-to-1 relative to the model that simply added new traces to memory. Inspection of the posterior predictions revealed that the models that did not allow for increase in memory strength were unable to predict the high frequency of "Old"-responses to targets that had been presented five times. When assuming an increase in memory strength with repeated target presentations, the removal of response scaling from the model was favored $\mathrm{BF}_{\gamma/0} = 577.08$ $[18.95,\ 6{,}706.38]$-to-1.

Table 4.3: Median posterior parameter estimates of the population means in natural scale $\exp(\mu^{(\theta)})$ (95% HDI in parentheses) for the unconstrained Generalized Context Model ($\mathcal{M}_0^{\text{GCM}}$) and the constraint where $\lambda = c = m$, $\mathcal{M}_\lambda^{\text{GCM}}$.

	Category size	
Parameter	4 vs. 8	1 vs. 11
$\mathcal{M}_0^{GCM}$		
c	1.25 [1.07, 1.43]	1.24 [1.09, 1.38]
c'	8.95 [3.77, 16.46]	4.61 [2.48, 7.77]
m'	5.61 [2.82, 9.72]	5.59 [2.94, 9.04]
k	5.84 [3.55, 9.44]	6.39 [3.66, 10.43]
γ	0.88 [0.71, 1.07]	0.89 [0.73, 1.06]
$\mathcal{M}_\lambda^{GCM}$		
λ	1.28 [1.11, 1.46]	1.21 [1.07, 1.36]
λ'	8.17 [4.21, 13.77]	4.69 [2.61, 7.88]
k	4.00 [3.10, 5.05]	3.52 [2.66, 4.53]
γ	0.81 [0.69, 0.93]	0.94 [0.79, 1.12]

Note. We additionally estimated a common average inter-category distance for both groups, $d_{\text{inter}} = 8.08$ 95% HDI [7.44, 8.81] and $d_{\text{inter}} = 8.07$ 95% HDI [7.39, 8.76] respectively, see Figure B.3.

As evident from Figure 4.5, the favored model $\mathscr{M}_{\times}^{\text{GCM}}$, which assumes the same group-level parameters across groups, exhibits some quantitative misfit, but clearly reproduces a pattern of results that is consistent with two latent dimensions. In the group with the stronger category size manipulation (1 vs. 11 exemplars) the model predicts too much false recognition when eleven exemplars were presented and targets were presented five times. Also when more than one category exemplar was presented, the model predicts too little true and too much false recognition. However, keeping in mind the constraint exerted by the inter-item similarities as well as our simplifying assumptions, overall the fit of the GCM to the multidimensional outcome space is satisfactory both for individuals and at the group level.

The winning model, however, was $\mathscr{M}_{\lambda}^{\text{GCM}}$, which constrained memory strength and dissimilarity sensitivity to be equal. It was favored $\text{BF}_{\lambda/0} = 1.15 \times 10^3$ $[29.87,\ 9{,}603.22]$-to-1 relative to the unconstrained model. Again, constraining the population means of all parameters to be equal across groups was supported by the data $\text{BF}_{\lambda\times/\lambda} = 8.98 \times 10^3$ $[4{,}762.81,\ 26{,}185.69]$-to-1. The predictions of the model were almost indistinguishable from those of the unconstrained model.

Finally, we assessed the correspondence between CRM and GCM by inspecting individual participants' V, G, and b estimates. We compared estimates from CRM to estimates calculated from GCM's posterior predicted mean response probabilities per conditions, Figure 4.6. In accordance with the previous results there were some divergences. For example, due to the over-prediction of "Old"-responses to lures there was a tendency for GCM produce larger G estimates than CRM. In the group with the weaker category size manipulation (4 vs. 8 exemplars) GCM estimates of G were more variable than those from CRM. For V, the estimates produced by GCM tend to be slightly lower compared to those from CRM, because GCM predicts to few old-response to repeatedly presented targets. However all things considered, the estimates of verbatim and gist activation derived from CRM and GCM are well correlated.

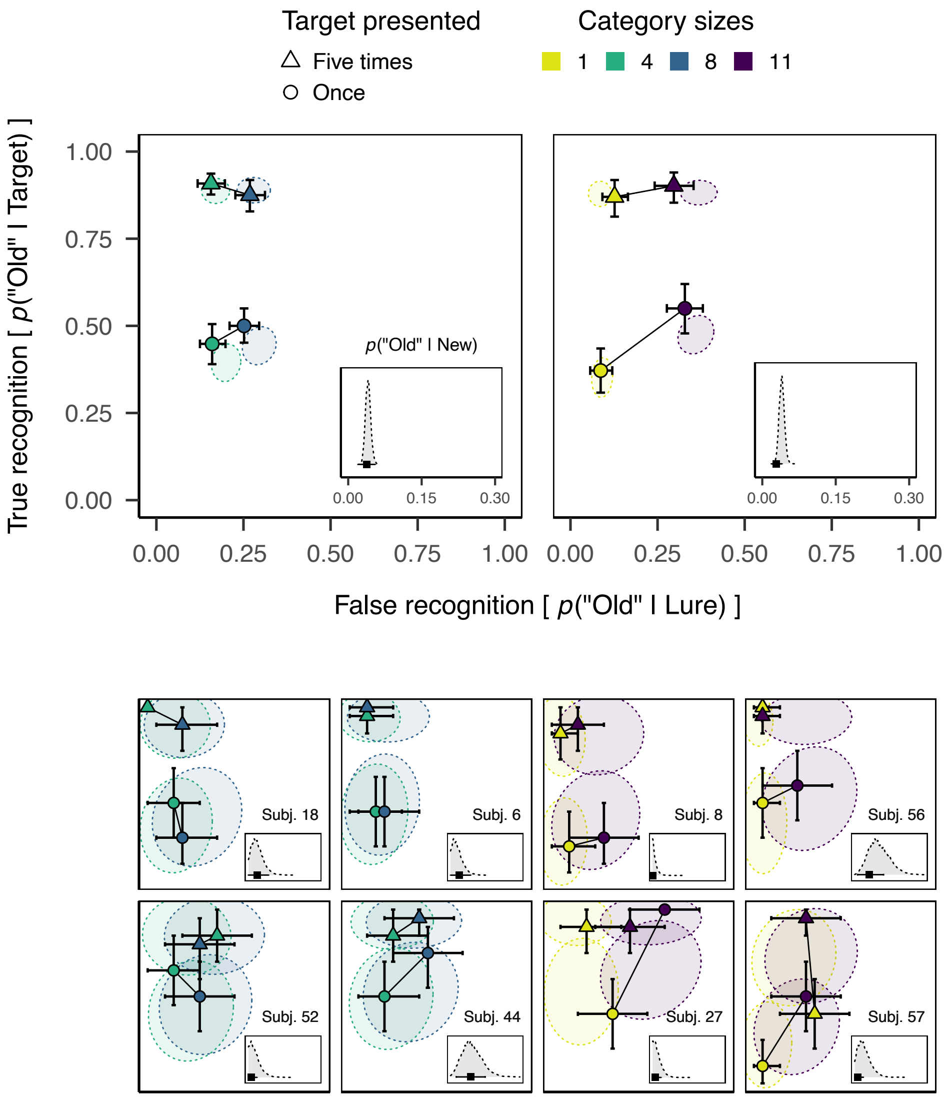

Figure 4.5: State-trace plot of averaged (top) and individual responses (bottom) and posterior predictions of $\mathscr{M}_{\times}^{\text{GCM}}$. From each group we show the two participants with strongest support for the non-monotonic (top row) and monotonic models (bottom row, see Figure 4.3). Points represent average observed rates of "Old"-responses; error bars indicate 95% bootstrap confidence intervals based on 10,000 bootstrap samples. Ellipses represent multivariate normal-approximations to 95% credible regions posterior predictions. The inset shows the proportion of "Old"-responses to unrelated new probes; kernel density estimates represent the posterior predictions.

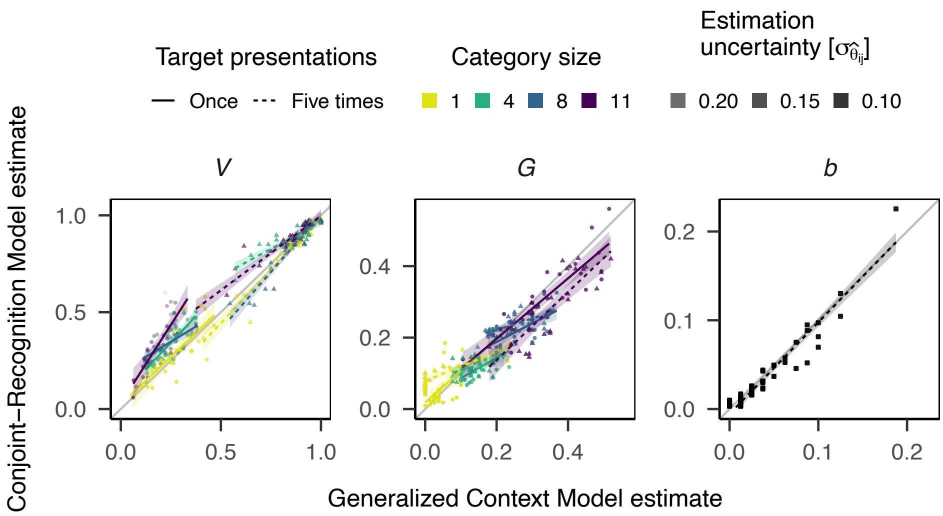

Figure 4.6: Scatter plot of verbatim retrieval (V), gist retrieval (G), and guessing probabilities (b) as estimated from the Conjoint-Recognition Model (CRM, $\mathscr{M}_0^{\text{CRM}}$) and calculated from the posterior predictions of the Generalized Context Model (GCM, $\mathscr{M}_0^{\text{GCM}}$). The estimation uncertainty represents the standard deviation of each parameters posterior distribution from the GCM. Lines show weighted linear regression predictions of CRM parameters from GCM estimates using estimation uncertainty as observation weights. Ribbons represent the 95% confidence interval of expected values.

4.7 Discussion

We have examined the formal relationship between two influential single- and dual-trace accounts of true and false recognition—fuzzy-trace theory and global matching memory models. Our results suggest that what fuzzy-trace theory conceptualizes as independent gist and verbatim traces may in fact reflect incremental contributions of partial and exact matches between probes and memory traces. We then tested the global matching account empirically in a study designed to selectively influence CRM parameters. State-trace analysis of our selective influence manipulations confirmed that multiple latent dimensions contributed to the observed true and false recognition responses. CRM fits confirmed that these latent dimensions were consistent with independent gist and verbatim traces. However, in line with our theoretical model analysis, the single-trace GCM also produced satisfactory fits to the observed responses. Our empirical results, thus, support the possibility that independent gist and verbatim activation may reflect partial and exact matches between probes and memory traces. Put differently, gist traces may reflect features shared between lures and study list items, whereas verbatim traces may reflect those features (combinations) unique to study list items.

Our findings demonstrate single dissociations between true and false recognition do not necessitate dual trace assumptions or independent retrieval processes. Critically, GCM accounted for our results without making ad hoc assumptions about the distribution of inter-item similarities. This potentially puts GCM at odds with MINERVA 2, which has been found to be unable to account for true and false recognition in the DRM paradigm when the number of associates is manipulated (Johns & Jones, 2010). The different outcomes for GCM and MINERVA 2 may be theoretically informative, but at this time we can only speculate about the cause of the discrepancy. Given the close relationship between stimulus representations in both models (Kelly et al., 2017), the diverging findings may be attributable to differences in encoding and retrieval mechanics. Differentiation and increases in memory strength are an obvious candidate causes. Another reason for the diverging results may be the stimulus material. Kelly et al. (2017) modeled responses to DRM lists, whereas we used categorical photographs of everyday objects. It is possible that the semantic similarity structure of DRM lists is a more challenging constraint for global matching models. Future research should examine whether our findings generalize to different stimulus materials.

The close correspondence between CRM and GCM we observed encourages further exploration. For example, we have discussed how global matching models may account for double dissociations between true and false recognition. If the opponent process of recollection rejection is conceptualized as cued recall, manipulations that improve recall of targets, such as target priming, may increase true but decrease false recognition. However, even without assuming recollection rejection, similar dissociations may, in principle, result from the differentiation of memory traces. Another mechanism that may yield double dissociations is a change in dissimilarity sensitivity across groups. For example, Reyna and Kiernan (1994) observed higher true but lower false recognition in older compared to younger participants. Such a pattern could result from greater dissimilarity sensitivity in older participants combined with a more stricter response criterion (Nosofsky & Zaki, 1998). Further exploration of these alternative accounts for key findings in false recognition are an interesting avenue for future research.

Finally, a key postulate of fuzzy-trace theory is that verbatim traces are forgotten at a faster rate than gist traces (p. 84, Brainerd & Reyna, 2005). In terms of global matching models, these different temporal dynamics of exact and partial matches are to be expected. In false recognition paradigms, memory probes match many memory traces partially, but only a few traces exactly. In other words, partial matches or gist representations are more redundant and thus less directly affected by forgetting. Further, because the activation of memory traces decrease exponentially as dissimilarity between probe and trace increases, the contribution of exact matches to the summed similarity will be disproportionately affected by forgetting. A particularly influential finding on the forgetting rates of gist and verbatim traces are so-called false recognition sleeper effects: As time passes, true recognition decreases whereas false recognition increases. Brainerd and Reyna (2005) note

> *This effect is only possible under the assumption that verbatim and gist traces have opposite effects on false-memory reports, and it is therefore a highly diagnostic result with respect to that assumption. (p. 151)*

But recent research, suggesting that not only memory strength, but also dissimilarity sensitivity decreases over time (Nosofsky et al., 2011; Nosofsky & Gold, 2016), offers an alternative account for this sleeper effect. For simplicity, consider our model that constraints memory strength and dissimilarity sensitivity to be equal. As time passes, the similarity gradient flattens out and approaches a uniform distribution, Figure 4.7. This causes a substantial decrease in activation for exact matches and hence in V. Depending on the similarity between lure and memory trace, activation for partial matches may decrease if the lure is highly similar, but activation may also non-negligibly increase if the lure is moderately similar. Thus, GCM and potentially other global matching models may offer an alternative account for false recognition sleeper effects. Take together, it appears that the global matching account of false recognition can accommodate the key principles of fuzzy-trace theory.

4.7.1 Implications for Fuzzy-Trace Theory

Fuzzy-trace theory is an influential theory with great heuristic value which inspired many accurate predictions about true and false recognition. The single-trace conceptualization we propose takes nothing away from these achievements. Similarly, our results should not be interpreted as a threat to the utility of CRM as a measurement model. As our theoretical model analysis and empirical assessment suggest CRM model parameters appear to closely correspond to the incremental contributions of partial and exact matches between probes and memory traces. While GCM, and other global matching models, may provide a more complete account of the matching process, CRM is a more readily applicable measurement model because it does not require information about inter-item similarities and it can be fit to observed responses using well-developed easy-to-use estimation procedures. As such, CRM is a useful measurement tool regardless of the validity of the dual-trace assumption.

Our suggested reconceptualization of gist and verbatim activation is largely consistent with their functional interpretation put forth when the model was first introduced (Brainerd et al., 1999). In the original formulation, CRM parameters were functionally labelled as reflecting similarity ($S = G$) and identity judgments ($I = V$). In global-matching terms, gist activation reflects partial matches between probes and memory traces and

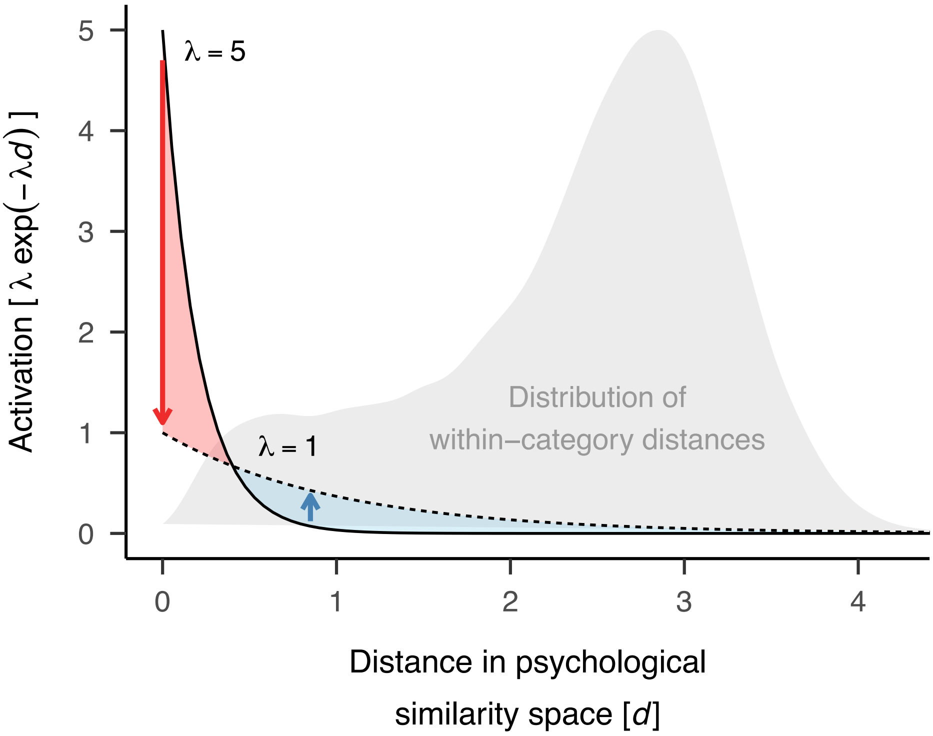

Figure 4.7: Change in similarity gradient as memory strength and dissimilarity sensitivity decrease in our model $\mathscr{M}_{\lambda}^{\mathrm{GCM}}$, where $\lambda = c = m$. The red shaded areas indicate a decrease in activation as λ decreases, the blue shaded areas indicate an increase in activation as λ decreases. For comparison, the grey filled curve represents the empirical distribution of within-category distances for the color photographs used in our experiment.

elicits similarity judgments, whereas verbatim activation reflects exact matches and elicits identity judgments.

Critically, the global matching interpretation of gist and verbatim memory may help to clarify some theoretical and empirical questions. A conceptual difficulty of separate verbatim and gist traces is the potential ambiguity of an episodes gist. For any given episode, gist could be extracted at different levels of abstraction. In our initial example of a visit to a microbrewery, the gist could simply be about a night out with friends, having beers with friends, or a beer tasting event at the local microbrewery. If gist is extracted immutably at encoding, it is not clear how one of the many potential gists is selected.

Similarly, the global matching interpretation of gist and verbatim activation readily accounts for perceptual false recognition. According to fuzzy-trace theory, gist traces represent the semantic, conceptual, and associative components of an episode, whereas verbatim traces represent perceptual and contextual details. This distinction is difficult to reconcile with the empirical finding of false recognition of lures that share only perceptual features with the studied material. Several studies have found false recognition for orthographically or phonologically similar but semantically unrelated lures (e.g., *hate, mate, late*, *fate*; Schacter et al., 1997; Brainerd et al., 1995; Budson et al., 2003; Ly et al., 2013; Shiffrin et al., 1995). A recent study suggests that perceptual false recognition is also possible for perceptually similar but conceptually unrelated drawings (Stahl, Henze, et al., 2016). To account for these findings, fuzzy-trace theory needs to soften its distinction between gist and verbatim representations and allow for the representation of a perceptual gist.

The global-matching view resolves both these challenges. Viewing gist activation as a partial match between the probe and memory traces obviates the need to settle on any gist interpretation for a given episode and readily accounts for perceptual false recognition. In this sense, what we have discussed as a dichotomy between exact and partial matches, of course, constitutes a continuum that implies a similar gist-verbatim-continuum. Thus, we agree with Cowan (1998), who, in his commentary on the paper that first introduced fuzzy-trace theory, cautioned that

> *We may use the distinction between verbatim and gist information as a useful simplification that allows a better understanding of memory,*

> *but let us resist the assumption that the data require a sharp division between the two in principle. (p. 149)*

4.7.2 Future research

Before any strong conclusions can be drawn, more work is needed to explore the generality of our findings. For example, in light of previous reports (Johns & Jones, 2010), it should to be tested whether GCM can account for data from the DRM paradigm. If GCM continues to perform adequately, the global matching perspective suggests novel empirical tests and paves the way for theoretical exchange with related models of memory and areas of research. For example, GCM has been extended to predict response times in addition to rates of old-responses (Donkin & Nosofsky, 2012a; Nosofsky & Palmeri, 2014). Such joint prediction of response rates and latencies may be used to test the global matching account in future applications. CRM currently makes no predictions about response latencies.

In our experiment, repeated presentation of study list items increased true recognition but had no effect on false recognition. In contrast, other studies report that repeated presentations decreased false recognition (pp. 111-112, Brainerd & Reyna, 2005). As we have demonstrated, GCM requires a differentiation extension to account for the selective influence on true recognition. In differentiation models (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997), differentiation emerges mechanistically as a consequence of encoding noise and corrective trace updating. According to these mechanics, differentiation may increase or decrease false recognition depending on target-lure similarity as well as the proportion of study list items that are strengthened via repeated study (Criss, 2006, also see Figure 4.7). These factors may explain the inconsistencies in the reported effects of repeated study on false recognition and suggest an informative test of the global matching account of false recognition. Again, such detailed predictions do not follow from CRM or fuzzy-trace theory.

Relatedly, the necessity of a differentiation process may motivate more principled extensions of GCM. We approximated differentiation as greater item-specific dissimilarity sensitivity (and an increase in memory strength) because GCM assumes noise-free encoding. A close relative of GCM, the noisy exemplar model (NEMo; Kahana & Sekuler, 2002), allows for encoding noise and may thus provide a starting point to develop a differentiation mechanism. The model assumes that encoding noise takes the form of multivariate Gaussian noise on the stimulus representation in multidimensional space. Differentiation could, for example, be implemented by a reduction of the variance of the noise distribution as a function of repeated study.

4.7.3 Conclusion

We have formalized the previously suggested correspondence between gist and verbatim activation postulated by fuzzy-trace theory and partial or exact matches between probes and memory traces postulated global matching models. Our results confirm the correspondence between these mnemonic components of true and false recognition. Taken together, the formal model correspondence and the results of our model fitting exercise suggest that verbatim and gist retrieval estimates can be viewed as measures of partial and exact memory matches. This reinterpretation of gist and verbatim activation is more than a reformulation of fuzzy-trace theory. It suggests new directions for empirical as well as theoretical research.

References

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature News*, *508*(7496), 312. https://doi.org/10.1038/508312a

Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of a week. *Psychonomic Bulletin & Review*, 1–6. https://doi.org/10.3758/s13423-015-0800-0

Arndt, J., & Hirshman, E. (1998). True and False Recognition in MINERVA2: Explanations from a Global Matching Perspective. *Journal of Memory and Language*, *39*(3), 371–391. https://doi.org/10.1006/jmla.1998.2581

Ashby, F. G., & Maddox, W. T. (1993). Relations between Prototype, Exemplar, and Decision Bound Models of Categorization. *Journal of Mathematical Psychology*, *37*(3), 372–400. https://doi.org/10.1006/jmps.1993.1023

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown* [R package version 0.1.0.9485]. https://github.com/crsh/papaja

Bartlema, A., Lee, M., Wetzels, R., & Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: Case studies in selective attention and representation in category learning. *Journal of Mathematical Psychology*, *59*, 132–150. https://doi.org/10.1016/j.jmp.2013.12.002

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brainerd, C. J., Nakamura, K., Chang, M., & Bialer, D. M. (2019). Verbatim editing: A general model of recollection rejection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1776–1790. https://doi.org/10.1037/xlm0000684

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-Trace Theory and Children's False Memories. *Journal of Experimental Child Psychology*, *71*(2), 81–129. https://doi.org/10.1006/jecp.1998.2464

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory* (Vol. 14). Oxford University Press.

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*(3), 343–382. https://doi.org/10.1037/0033-2909.134.3.343

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-Recognition Reversal: When Similarity is Distinctive. *Journal of Memory and Language*, *34*(2), 157–185.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*(1), 160–179. https://doi.org/10.1037/0033-295X.106.1.160

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*(4), 762–784. https://doi.org/10.1037/0033-295X.110.4.762

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 307–327. https://doi.org/10.1037/0278-7393.27.2.307

Buchanan, L., Brown, N. R., Cabeza, R., & Maitson, C. (1999). False Memories and Semantic Lexicon Arrangement. *Brain and Language*, *68*, 172–177. https://doi.org/10.1006/brln.1999.2072

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition*, *51*(3), 251–261. https://doi.org/10.1016/S0278-2626(03)00030-7

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*(1), 37–60. https://doi.org/10.3758/BF03210740

Cowan, N. (1998). Children's Memories According to Fuzzy-Trace Theory: An Endorsement of the Theory's Purpose and Some Suggestions to Improve Its Application. *Journal of Experimental Child Psychology*, *71*(2), 144–154. https://doi.org/10.1006/jecp.1998.2466

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*(4), 461–478. https://doi.org/10.1016/j.jml.2006.08.003

Criss, A. H., & Koop, G. J. (2015). Differentiation in episodic memory. In J. G. W. Raaijmakers, A. H. Criss, R. Goldstone, R. M. Nosofsky, & M. Steyvers (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin* (pp. 112–125). Psychology Press.

Davis-Stober, C. P., Morey, R. D., Gretton, M., & Heathcote, A. (2016). Bayes factors for state-trace analysis. *Journal of Mathematical Psychology*, *72*, 116–129. https://doi.org/10.1016/j.jmp.2015.08.004

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dodson, C. S., & Schacter, D. L. (2001). If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*(1), 155–161. https://doi.org/10.3758/BF03196152

Donkin, C., & Nosofsky, R. M. (2012a). A Power-Law Model of Psychological Memory Strength in Short- and Long-Term Recognition. *Psychological Science*, *23*(6), 625–634. https://doi.org/10.1177/0956797611430961

Donkin, C., & Nosofsky, R. M. (2012b). The structure of short-term memory scanning: An investigation using response time distribution models. *Psychonomic Bulletin & Review*, *19*(3), 363–394. https://doi.org/10.3758/s13423-012-0236-8

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. https://doi.org/10.1037/0033-295X.106.1.180

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*(1), 91–101. https://doi.org/10.1037/0033-295X.95.1.91

Gallo, D. (2006). *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*. Psychology Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015). *Bayesian Data Analysis*. CRC press.

Greve, A., Donaldson, D. I., & van Rossum, M. C. W. (2010). A single-trace dual-process model of episodic memory: A novel computational account of familiarity and recollection. *Hippocampus*, *20*(2), 235–251. https://doi.org/10.1002/hipo.20606

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*(Supplement C), 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Gronau, Q. F., & Singmann, H. (2018). *Bridgesampling: Bridge sampling for marginal likelihoods and bayes factors* [R package version 0.6-0]. https://CRAN.R-project.org/package=bridgesampling

Heck, D. W., & Wagenmakers, E.-J. (2016). Adjusted priors for Bayes factors involving reparameterized order constraints. *Journal of Mathematical Psychology*, *73*, 110–116. https://doi.org/10.1016/j.jmp.2016.05.004

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1987). Recognition and recall in MINERVA 2: Analysis of the 'recognition-failure' paradigm. *Modelling cognition* (pp. 215–229). John Wiley & Sons.

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 667–680. https://doi.org/10.1037/0278-7393.18.4.667

Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Tenzing: Documenting contributorship using CRediT. *MetaArXiv*. https://doi.org/10.31222/osf.io/b6ywe

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A Multidimensional Scaling Database with Similarity Ratings for 240 Object Categories from the Massive Memory Picture Database. *PLoS ONE*, *9*(11), e112644. https://doi.org/10.1371/journal.pone.0112644

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208–233. https://doi.org/10.1037/0033-295X.96.2.208

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*(5), 662–672. https://doi.org/10.3758/PBR.17.5.662

Jou, J., & Flores, S. (2013). How are false memories distinguishable from true memories in the Deese–Roediger–McDermott paradigm? A review of the findings. *Psychological Research*, *77*(6), 671–686. https://doi.org/10.1007/s00426-012-0472-6

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*(18), 2177–2192. https://doi.org/10.1016/S0042-6989(02)00118-9

Kalish, M. L., Dunn, J. C., Burdakov, O. P., & Sysoev, O. (2016). A statistical test of the equality of latent orders. *Journal of Mathematical Psychology*, *70*, 1–11. https://doi.org/10.1016/j.jmp.2015.10.004

Kelly, M. A., Mewhort, D. J. K., & West, R. L. (2017). The memory tesseract: Mathematical equivalence between composite and separate storage memory models. *Journal of Mathematical Psychology*, *77*, 142–155. https://doi.org/10.1016/j.jmp.2016.10.006

Klaassen, F., Zedelius, C. M., Veling, H., Aarts, H., & Hoijtink, H. (2018). All for one or some for all? Evaluating informative hypotheses using multiple N = 1 studies. *Behavior Research Methods*, *50*(6), 2276–2291. https://doi.org/10.3758/s13428-017-0992-5

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C., & Jackson, E. M. (1999). Perceptually Based False Recognition of Novel Objects in Amnesia: Effects of Category Size and Similarity to Category Prototypes. *Cognitive Neuropsychology*, *16*(3-5), 317–341. https://doi.org/10.1080/026432999380816

Ly, M., Murray, E., & Yassa, M. A. (2013). Perceptual versus conceptual interference and pattern separation of verbal stimuli in young and older adults. *Hippocampus*, *23*(6), 425–430. https://doi.org/10.1002/hipo.22110

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*(4), 335–384. https://doi.org/10.1016/j.cogpsych.2008.02.004

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *275*(942), 483–519. https://doi.org/10.1098/rstb.1976.0090

Marr, D. (1982). *Vision*. W. H. Freeman and Company.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*(4), 724–760. https://doi.org/10.1037/0033-295X.105.4.734-760

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 128–148. https://doi.org/10.1037/0096-1523.21.1.128

Navarro, D. J. (2007). On the interaction between exemplar-based concepts and a response scaling process. *Journal of Mathematical Psychology*, *51*(2), 85–98. https://doi.org/10.1016/j.jmp.2006.11.003

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*(8), 285–290. https://doi.org/10.1016/j.tics.2008.04.009

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27. https://doi.org/10.1037/0096-1523.17.1.3

Nosofsky, R. M. (2011a). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Nosofsky, R. M. (2011b). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Nosofsky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, *75*, 97–129. https://doi.org/10.1016/j.cogpsych.2014.08.003

Nosofsky, R. M., Cao, R., Harding, S. M., & Shiffrin, R. M. (2020). Modeling short- and long-term memory contributions to recent event recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000812

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1524–1539. https://doi.org/10.1037/xlm0000015

Nosofsky, R. M., & Gold, J. (2016). Memory strength versus memory variability in visual change detection. *Attention, Perception, & Psychophysics*, *78*(1), 78–93. https://doi.org/10.3758/s13414-015-0992-4

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315. https://doi.org/10.1037/a0022494

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Palmeri, T. J. (2014). An Exemplar-Based Random-Walk Model of Categorization and Recognition. In J. Busemeyer, J. Townsend, Z. Wang, & A. Eidels (Eds.), *Mathematical and Computational Models of Cognition*. Oxford University Press.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*(2), 530–556. https://doi.org/10.3758/s13428-017-0884-8

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations Between Categorization and Recognition in Amnesic and Normal Individuals: An Exemplar-Based Interpretation. *Psychological Science*, *9*(4), 247–255. https://doi.org/10.1111/1467-9280.00051

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psy-

chology: Learning, Memory, and Cognition, *28*(5), 924–940. https://doi.org/10.1037/0278-7393.28.5.924

Pidgeon, L. M., & Morcom, A. M. (2014). Age-related increases in false recognition: The role of perceptual and conceptual similarity. *Frontiers in Aging Neuroscience*, *6*, 283. https://doi.org/10.3389/fnagi.2014.00283

Pierce, B. H., Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2005). The modality effect in false recognition: Evidence for test-based monitoring. *Memory & Cognition*, *33*(8), 1407–1413. https://doi.org/10.3758/BF03193373

Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, *17*(1), 78–99. https://doi.org/10.1037/a0025809

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Reyna, V. F., & Brainerd, C. J. (1995a). Fuzzy-trace theory: Some foundational issues. *Learning and Individual differences*, *7*(2), 145–162. https://doi.org/10.1016/1041-6080(95)90028-4

Reyna, V. F., & Brainerd, C. J. (1998). Fuzzy-Trace Theory and False Memory: New Frontiers. *Journal of Experimental Child Psychology*, *71*(2), 194–209. https://doi.org/10.1006/jecp.1998.2472

Reyna, V. F., & Brainerd, C. J. (1995b). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1–75. Retrieved July 5, 2016, from http://www.sciencedirect.com/science/article/pii/1041608095900314

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*(2), 178–191. https://doi.org/10.1037/0012-1649.30.2.178

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*(2), 95–123. https://doi.org/10.1016/S1041-6080(97)90002-9

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., & McDermott, K. B. (2000). Tricks of Memory. *Current Directions in Psychological Science*, *9*(4), 123–127. https://doi.org/10.1111/1467-8721.00075

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing False Recognition in Younger and Older Adults: The Distinctiveness Heuristic. *Journal of Memory and Language*, *40*(1), 1–24. https://doi.org/10.1006/jmla.1998.2611

Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. *Neuropsychology*, *11*(3), 331. https://doi.org/10.1037/0894-4105.11.3.331

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of experimental psychology. Learning, memory, and cognition*, *21*(2), 267–287. https://doi.org/10.1037/0278-7393.21.2.267

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Stahl, C., Henze, L., & Aust, F. (2016). False memory for perceptually similar but conceptually distinct line drawings. *PsyArXiv*. https://doi.org/10.31234/osf.io/zr7m8

Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 570–586. https://doi.org/10.1037/0278-7393.34.3.570

Stahl, C., & Klauer, K. C. (2009). Measuring phantom recollection in the simplified conjoint recognition paradigm. *Journal of Memory and Language*, *60*(1), 180–193. https://doi.org/10.1016/j.jml.2008.08.001

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*(1), 122–129. https://doi.org/10.1037/h0022014

van Vugt, M. K., Sekuler, R., Wilson, H. R., & Kahana, M. J. (2013). An electrophysiological signature of summed similarity in visual working memory. *Journal of Experimental Psychology: General*, *142*(2), 412–425. https://doi.org/10.1037/a0029759

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 1022–1041. https://doi.org/10.1037/0278-7393.27.4.1022

Chapter 5

An exemplar-familiarity model of false recognition over the short term

FREDERIK AUST & CHRISTOPH STAHL

False recognition can be elicited over the short and long term by presenting lists of semantically related words. This suggests that semantic and categorical relations, traditionally considered to be represented in long-term memory, also affect short-term memory. Hence, false recognition in short-term memory poses a problem to models positing that short-term memory is insulated from long-term memory. Because of their similarity short-term false recognition has been explained by established dual-process theories of long-term memory, most notably Fuzzy-Trace theory. Here we test an alternative exemplar-familiarity account, the Generalized Context Model (GCM). We conducted an experiment, with photographic and verbal material, and combined selective influence manipulations of gist and verbatim memory. We observed false recognition with all materials and found that our manipulations selectively affected estimates of gist and verbatim activation of the Conjoint-Recognition model. Finally, we show that GCM is able to account for the observed true and false recognition rates. We conclude that GCM is a serious alternative to dual-process accounts of false recognition in short-term memory.

False memories are typically considered a phenomenon of long-term memory and studied with long lists or across relatively long retention intervals. However, recent studies have established that false recognition and recall can similarly be elicited over with much shorter lists and brief retention intervals (e.g., Coane et al., 2007; Atkins & Reuter-Lorenz, 2008; Flegal et al., 2010). These findings suggest a close relationship between short- and long-term memory (e.g., Abadie & Camos, 2019) and support the idea of a unitary memory in which the same operating principles govern remembering over the short- and long-term (e.g., Brown et al., 2007; Crowder, 1993; Jonides et al., 2008; Nairne, 1990; Nairne, 2002). In this spirit, here we used the Generalized Context Model (GCM; Nosofsky, 1986; Nosofsky, 1988; Nosofsky, 2011a), a process model of categorization and long-term recognition memory, to model short-term false recognition. To challenge the model, we simultaneously strengthened individual study list items via repeated presentation and varied the number of related study list items. By and large, the model was able to account for the effects of our manipulations across three different materials.

5.1 False recognition over the short term

Short-term false memories can be elicited with a variety of procedures. Most studies have adapted the classic Deese-Roediger-McDermott paradigm (DRM; Deese, 1959; Roediger & McDermott, 1995) and presented lists of words that are strongly associated with one word that is not presented during study—the *lure*. As in long-term memory, these converging associations yield substantial rates of false memories in old-new recognition (e.g., Atkins & Reuter-Lorenz, 2008; Flegal et al., 2010; Festini & Reuter-Lorenz, 2013), serial recognition (Macé & Caza, 2011; Tse et al., 2011), recall (Atkins & Reuter-Lorenz, 2008; Dimsdale-Zucker et al., 2018), and serial recall paradigms (Tehan, 2010; Tse et al., 2011). Additionally, converging associates slow response times to lures (Coane et al., 2007; Jou et al., 2016). Moreover, false memories in short-term memory can be elicited by visual or verbal presentation of word lists (Macé & Caza, 2011;

Individual contributions

Frederik Aust: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing - Original Draft Preparation

Christoph Stahl: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision

Contributor Roles Taxonomy (CRediT; Allen et al., 2014; Holcombe et al., 2020)

Olszewska et al., 2015), with simultaneous (e.g., Atkins & Reuter-Lorenz, 2008; Flegal et al., 2010) or sequential list presentation (e.g., Flegal & Reuter-Lorenz, 2014; Olszewska et al., 2015; Jou et al., 2016), but also with visual material, such as photographs of or stylized faces (van Vugt et al., 2013; Iidaka et al., 2014). These studies show that false memories can be elicited with similar methods over the short- and long-term.

Besides demonstrating the robustness and generality of false memories over the short term, these studies indicate that the effect is guided by similar principles as apply over the long term. For example, false recognition increases with the length of DRM lists (Coane et al., 2007; Jou et al., 2016), decreases when participants are directed to forget converging associates (Festini & Reuter-Lorenz, 2013), and increases when study and test modality mismatch (an encoding-specificity effect; Olszewska et al., 2015). Moreover, direct comparisons of false memories over the short and long term suggest that false recognition rates, as well as the subjective phenomenology (as measured by confidence ratings and remember/know judgments), are similar (Flegal et al., 2010; Flegal & Reuter-Lorenz, 2014). These findings indicate that similar operating principles underlie short- and long-term false recognition (see also p. 309, Nosofsky et al., 2011; Cowan, 1999) and support the stronger assumption that the same processes govern short- and long-term memory (e.g., Brown et al., 2007; Crowder, 1993; Jonides et al., 2008; Nairne, 1990; Nairne, 2002).

While there is growing support for a unitary verbal short- and long-term memory representation, the case is less clear for visual short- and long-term memory (for a review see Brady et al., 2011). Two important characteristics that presumably are unique to short-term memory are its high fidelity of the representation (e.g., Zhang & Luck, 2008), a fundamentally limited capacity (Cowan, 2001; Ma et al., 2014). Recently, these properties of visual short-term memory have been extensively studied with a psychophysical adjustment method referred to as continuous report (Bays et al., 2009; Wilken & Ma, 2004; Zhang & Luck, 2008). Participants view color patches and, at test, are asked to report these colors on a circular color wheel. The deviation of participants responses from the true value is recorded and the distribution of deviations across trials characterizes the fidelity of participants memory representations. Because the distribution is symmetric, Gaussian-like with fat tails, some have argued that it is a mixture that can be decomposed into to types of responses: (1) On some trials the to-be-reported color is represented in memory and participants report the contents of memory to the best of their ability. These responses provide insight into the fidelity of the underlying representations. (2) Owing to its limited capacity, on other trials the to-be-reported color is not represented in memory—the memory "slots" are filled with other items. Hence, participants guess randomly and responses are distributed uniformly along the color wheel. In this and other paradigms, direct comparisons of short- and long-term memory have revealed a higher fidelity of short-term memory (e.g., Biderman et al., 2019; Schurgin & Flombaum, 2015; Schurgin & Flombaum, 2018). The lower fidelity of long-term memory, as well as the slot-like architecture that is assumed to underlie guessing in short-term memory, indicate structural differences that separate short- and long-term memory.

Recently, it has been argued that guessing in continuous report paradigms and the fidelity differences between short- and long-term memory can be explained by combining well-established functional characteristics of perception and memory (Miner et al., 2019; Schurgin et al., 2019). The universal law of generalization (Shepard, 1987) posits that perceived similarity decreases exponentially as stimuli become more different in, for example, size or color. When combined with the standard assumption that long-term memory is well characterized in terms of signal detection theory, the fat-tailed response-error distribution in the continuous report paradigm can be explained by memory strength and the similarity structure of the stimulus material (Schurgin et al., 2019). Thus, no guessing or slot-like memory structure need to be assumed. Accordingly, Miner et al. (2019) found that the state-trace space of guessing and fidelity estimates is consistent with a common underlying cause and that this common cause is well characterized as memory-strength. Under this view, the different fidelities of short- and long-term memories are a natural consequence of different memory strengths—as memory strength decreases, fidelity decreases. It follows that comparisons of memory fidelity across the short and long term must equate for memory strength. Miner et al. (2019) show that the fidelity of long-term memory can be as high as that of short-term memory if, for example, long-term memory items are strengthened by repetition. The increase in fidelity (or precision) of long-term memory is a phenomenon referred to as *differentiation* in the literature on long-term memory (Criss, 2006; Criss & Koop, 2015). Taken together, this work provides a compelling unitary storage account for a broad range of continuous report data.

The continuous report paradigm assesses short-term memory for an isolated elementary stimulus dimension, such as color. Evidence on false recognition in verbal short-term memory, however, suggests similar principles in short- and long-term memory with respect to semantic or categorical relations between stimuli. Extending this work to visual short-term memory requires more complex stimulus material. Recent investigations using photographs of categorically and semantically related stimuli have yielded conflicting findings. Quinlan and Cohen (2016) found no increase in short-term memory capacity when stimuli were members of a common category compared to unrelated stimuli. That is, participants appeared to be unable to efficiently chunk the information by exploiting category membership as organizing principle. This implies that visual short-term memory may be insulated from long-term memory and lead Quinlan and Cohen (2016) to conclude that short-term memory is "precategorical." In contrast, O'Donnell et al. (2018) did find increased capacity when stimuli were semantically related. In long-term memory, it is well established that semantic and categorical relations among visual stimuli affect memory performance and, in particular, can increase both true and false recognition (see Chapter 4). Thus, under a unitary memory view lists composed of multiple category exemplars should also increase true and false recognition in short-term memory. In the current study, we therefore used both verbal and photographic stimuli to assess the influence of semantic and category-level information on verbal and visual short-term memory.

5.2 Models of short-term false recognition

False recognition in short-term memory appears to follow similar operating principles as long-term false recognition. Hence, explanations for false long-term memory have been invoked to explain false memory over the short-term—namely fuzzy-trace theory (Reyna & Brainerd, 1995a) and the activation monitoring framework (McDermott & Watson, 2001; Roediger & McDermott, 2000; Roediger et al., 2001). Fuzzy-trace theory assumes that two encoding processes operate in parallel to create independent representations of an episodes gist and a verbatim details. At test, gist and verbatim traces presumably are retrieved independently. This dual-representation assumption fundamentally contradicts single-process models, such as global matching models of long-term memory (Clark &

Gronlund, 1996; Kelly et al., 2017; and Osth & Dennis, 2020). For example the Generalized Context Model (GCM; Nosofsky, 1986; Nosofsky, 1988; Nosofsky, 2011a), or its extension the Exemplar-based Random Walk Model (EBRW; Nosofsky & Palmeri, 1997; Nosofsky & Palmeri, 2014), assume that every episode leaves a trace in a unitary memory. At test, the probe is matched against all memory traces in parallel and the summed similarity of all traces is an index of memory strength and subjective familiarity. GCM and its variants can account for several key findings from short-term memory (e.g., Kahana & Sekuler, 2002; Nosofsky et al., 2011; for a review see Nosofsky, 2016). Consistent with unitary memory models, recent work indicates that such models may provide a common basis to jointly explain findings from both short- and longer-term memory (Nosofsky et al., 2020; Nosofsky, Cox, et al., 2014; Schurgin et al., 2019).

In Chapter 4, we presented a formal correspondence between the Conjoint Recognition Model (CRM; Brainerd et al., 1999; Brainerd et al., 2001), a formal implementation of fuzzy-trace theory, and the GCM. Moreover, we showed empirically that GCM is able to adequately account for true and false recognition in a selective influence study of CRM parameters in visual long-term memory. Selective influence manipulations aim at affecting a circumscribed psychological construct, while leaving other involved constructs unaffected. Specifically, repeating targets on the study list (verbatim repetition) should selectively strengthen verbatim memory whereas increasing the number of study list items related to a lure (gist repetition) should selectively strengthen gist memory (e.g., Stahl & Klauer, 2008; Stahl & Klauer, 2009). We found that, in combination, verbatim and gist repetition revealed a pattern of true and false recognition that necessitates multiple latent causes and, thus, provides a challenging testbed to test the GCM-account of false recognition. Based on a theoretical model analysis we implemented a simple differentiation mechanisms in GCM as an increase of item-specific dissimilarity sensitivity and an increase in memory strength with repeated study (Criss, 2006; Shiffrin & Steyvers, 1997). With this extension GCM captured the multicausal pattern of true and false recognition.

Together these results suggest that gist and verbatim retrieval may be better thought of as independent familiarity increments by partial and exact matches between probes and memory traces in a unitary memory system. Here, we used the same approach to test this single-process account of false recognition in short-term memory. We combined selective influence manipulations of gist and verbatim memory in a short-term memory task and fit both CRM and GCM to the data.

5.3 The current study

To summarize, in light of current debates about the influence of semantic and categorical relations between stimuli on visual short-term memory, we performed a short-term false recognition experiment with categorical verbal and photographic materials. To further test, whether verbal and visual short-term memory follow similar principles as long-term memory, we used selective influence manipulations of gist and verbatim memory and fit CRM and GCM to the responses. We previously fit GCM data from a long-term false recognition experiment with photographic material. Here, we extend this work by testing whether GCM can also account for false recognition in short-term memory with both photographic and verbal material. Moreover, given that we previously found that repeated study causes memory traces to become differentiated, we sought to test whether differentiation also occurs in short-term memory.

5.4 Methods

We performed an old/new-recognition experiment with selective influence manipulations targeting G and V parameters of the CRM. The general procedure was an adaptation of the procedure used by Flegal et al. (2010).

5.4.1 Participants

Eighty-one students of the University of Cologne, sampled from our lab database to be fluent in German, participated in the experiment in exchange for 7€ or course credit. Participants' mean age was 25.33 years (SD = 7.39), 25.33% were female, 87.65% were native speakers, and 33.33% studied psychology. All participants provided informed consent.

5.4.2 Material

We constructed each study list from one of three categorized materials: colored object photographs, greyscale object photographs, and words. For colored photographs we used the *Massive Memory* database (Brady et al., 2008). The database consists of 240 categories of 16 everyday objects. We used greyscale photographs from the database provided by Migo et al. (2013), which consists of 50 categories of 25 everyday objects. Empirical estimates of similarities for exemplars within each category are available for both databases (Hout et al., 2014; Migo et al., 2013). Finally, as verbal material we used 72 categorized lists of 5 words each used by Stahl and Klauer (2008) in their selective influence studies of the simplified CRM. Using the spatial arrangement method (Hout et al., 2013) we estimated the similarities for exemplars within each category via multidimensional individual difference scaling of 34-36 participants ($M = 35.18$, $SD = 0.83$). Based on visual inspections of plots of Kruskal's Stress, we assumed that all similarity spaces can be adequately approximated by three component dimensions.

We used distinct lists of items as primacy and recency buffers. For colored and greyscale object photographs we sampled from 128 colored and 88 greyscale images of road signs, respectively. For words we sampled from a list of 128 given names. No road signs or given names were included in either of the databases of study list items.

5.4.3 Procedure

Participants worked on computers in individual booths. Instructions, stimulus presentation, response collection were fully computerized. The experimental procedure consisted of a short-term and a subsequent surprise long-term recognition phase.

Each short-term recognition trial consisted of a short study list presentation, an intervening math equation verification task, and the recognition test. To construct the study lists we sampled six exemplars from two stimulus categories. On every trial, one category served as the small the other as the large category, Figure 5.1. We presented one exemplar from the small but four exemplars from the large category. Fuzzy-trace theory predicts that the increase in category size selectively increase gist retrieval, as assessed by the G parameter in CRM. Moreover, we always repeated one exemplar—either the exemplar from the small category or a randomly selected exemplar from the large category. Fuzzy-trace theory predicts that repeated target presentations selectively increase verbatim retrieval, as assessed by the V parameter in CRM. In the following we refer to trials in which the exemplar from the small category was repeated as *negative association trials* because we expected a negative association between true and false recognition for the small and large categories, respectively. Conversely, we expected a positive association between true and false for the small and large categories for trials in which an exemplar from the large category was repeated; we therefore refer to these trials as *positive association trials*.

At the beginning of each trial we presented a fixation cross for 1.5 s followed by a rapid succession of study list items. We presented words for 250 ms and images for 750 ms with no inter-stimulus interval (e.g., Nosofsky et al., 2011). The rapid visual presentation served to discourage subjects from rehearsing the material. Words were displayed in lower case. We randomly presented one or two buffer items prior to and two or three subsequent to the six critical memory items.

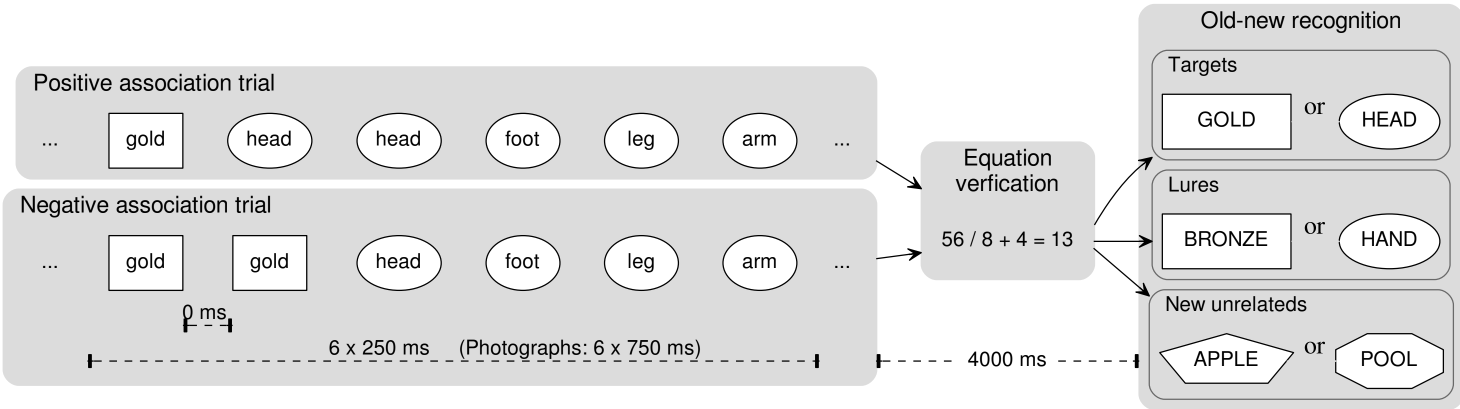

Figure 5.1: Diagram of the short-term recognition procedure for the word material. In positive association trials, we presented one exemplar from the small category (rectangles) but four exemplars from the large category (ovals), one of which was repeated. In negative association trials, we repeated the exemplar from the small category and presented each of the four exemplars from the large category only once. Following a 4 s-retention interval filled with an equation verification task, we presented one of three memory probes. We randomly probed the small or the large category with a previously presented item (target), an unpresented exemplar from the same category (lure), or a new item from an unpresented category (new). We presented study lists in random order; word stimuli were in German. Shapes represent different stimulus categories.

We instructed participants to closely follow and memorize the study list items as we would subsequently test their memory. The short-term recognition test probed participants memory for either the small or large category, Figure 5.1. We probed participants memory with either a previously presented exemplar (target), a related exemplar from the same category (lure), or an unrelated exemplar from a new category (new unrelated probe). Word probes were displayed in upper case. We asked participants to decided whether the displayed stimulus was old or new as fast and as accurately as possible. We then, displayed a four-level confidence scale and asked participants to indicate how confident they were in their response ranging from *very unsure* to *very sure*.[1]

In the intervening equation verification task, each equation consisted of multiplication or division of two integers a and b and subsequent addition or subtraction of a third integer c, where $a \in [2, 81]$, $b \in [2, 9]$, $c \in [2, 9]$. The correct result of the equation, too, was always an integer. The displayed result was correct on half of the trials and deviated from the correct result randomly by no more than 10. For example, the to-be-verified math equation may have been $56/8 + 4 = 13$. 500 ms after display onset, we asked participants to decide whether the display equation was correct as fast and accurately as possible. We gave immediate feedback for at least 500 ms by coloring the selected response green or red if it was correct or incorrect, respectively. Responses slower than 3.5 s were treated as incorrect; a countdown beneath the equation indicated the remaining time to respond. Overall the equation verification task lasted 4 s.

Participants completed the short-term recognition task separately for each stimulus material. Due to the varying number of categories available for the three stimulus materials the number trials differed between the blocks. For colored objects and words, we used 72 categories. We randomly selected 64 categories to construct the 32 study lists and reserved the remaining 8 categories for new unrelated memory probes. We unevenly assigned the 32 study lists to the three memory probe types: 12 targets, 12 lures, and 8 new unrelated probes. We distributed the 12 target and lure study lists evenly across the four combination of category size (1 vs. 4) and number of target presentations (once vs. twice), yielding three trials each. The 8 new unrelated probe trials were evenly assigned to positive and negative association trials, yielding four trials each. To increase the number of observa-

[1] For reasons of brevity, we do not report analyses of confidence judgments here.

tion, participants completed a second run through the three stimulus materials in the same order with everything else randomized anew—totaling six blocks.

After completion of the short-term recognition phase, participants unknowingly entered an approximately 20-minute retention interval for the subsequent surprise long-term recognition phase. During the retention interval participants completed a second unrelated experiment on evaluative conditioning (Stahl & Heycke, 2016). The long-term recognition task followed the same procedure as the short-term recognition task. We treated all short-term recognition study lists as targets for the long-term recognition task. However, we only used items as targets that had not been probed during the short-term recognition task.[2]

To summarize, the experiment realized a 2 (*Probe type*: Target vs. Lure) × 2 (*Target presentations*: Once vs. Twice) × 2 (*Category size*: 1 vs. 4) × 3 (*Material*: Color photos vs. Greyscale photos vs. Words) within-participant design. Unrelated new distractors constituted an additional probe type condition that was only crossed with material but not with the other within-participant factors. Overall, the duration of the experiment was approximately 95 minutes.

5.4.4 Data analysis

Our data analysis approach corresponds to the one used in Chapter 4 except where noted otherwise. We report Bayes factors as relative measures of evidence for our statistical and cognitive models (e.g, Wagenmakers et al., 2010). The Bayes factor may be equally interpreted as a model's prior predictive accuracy or as relative likelihood of the observed data given the model and the specified prior distributions on its parameters. Hence, Bayes factors are directly interpretable as graded measures of evidence. When Bayes factors are not available, we rely on posterior inference and compare marginal 95% highest density interval (HDI) estimates.

We used R (Version 3.6.3; R Core Team, 2017) and the R-packages *afex* (Version 0.23.0; Singmann et al., 2017), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *bridgesampling* (Version 0.6.0; Gronau & Singmann, 2018), *dplyr* (Version 0.8.4; Wickham & Francois, 2016), *drake* (Version

[2] We do not report analyses of the long-term recognition task here.

7.12.4; Landau, 2018), *emmeans* (Version 1.3.5; Lenth, 2018), *ggplot2* (Version 3.2.1; Wickham, 2016a), *papaja* (Version 0.1.0.9997; Aust & Barth, 2017), *purrr* (Version 0.3.3; Wickham, 2016b), and *tidyr* (Version 1.0.2; Wickham, 2017) for all our analyses.

5.4.4.1 Cognitive modelling

As in the previous chapter, we implemented CRM and GCM as Bayesian hierarchical models and estimated the joint posterior distribution of parameters by No-U-Turn sampling. To compare models using Bayes factors we estimated their marginal likelihoods using Warp-III bridge sampling (Gronau et al., 2017; Gronau & Singmann, 2018). We quantified the estimation error from 25 repeated sampling runs and report 2.5% and 97.5% qunatiles when the estimation error exceeds 5%.

5.4.4.1.1 Conjoint recognition model

We assumed that the probability of responding "Old" was the same for positive and negative association trials and, therefore, constrained b to be equal across trial types. Our ANOVA analysis of observed responses and by the fit of the model to data indicated that this assumption is tenable.

5.4.4.1.2 Generalized context model

In the absence of distance estimates for stimuli from different categories, we estimated three auxiliary parameters $d_{\text{inter}}^{(m)}$, common to all participants, as a stand-in for any $d(x_i, y_j)$ in material m, where x_i and y_j are exemplars from different categories. As in our previous application, we assumed that our primacy and recency buffers mitigate the differential effects of forgetting on memory traces and assumed constant memory strength for all items. Inspection of true recognition responses supported this assumption. Furthermore, we assumed that on each trial, only the current study list items entered the global matching process. That is, we did not model possible interference from previous trials across the course of the experiment.

While fitting the model, we observed non-trivial misfit for the word material. Based on research that suggests visually presented verbal material may be processed in an all-or-none fashion (Swagman et al., 2015), we hypothesized that participants may have experienced attentional lapses during the rapid study list presentation. We reasoned that attentional lapses would manifest as random guessing at test (cf., p. 381, Donkin & Nosofsky, 2012b; p. 140; van den Berg et al., 2014). Hence, we estimated the probability of attentional lapses p_{lapse} by modelling trial-wise response probabilities as a mixture of memory-based responses—as predicted by GCM—and randomly guessing old with probability $p = .5$.

5.5 Results

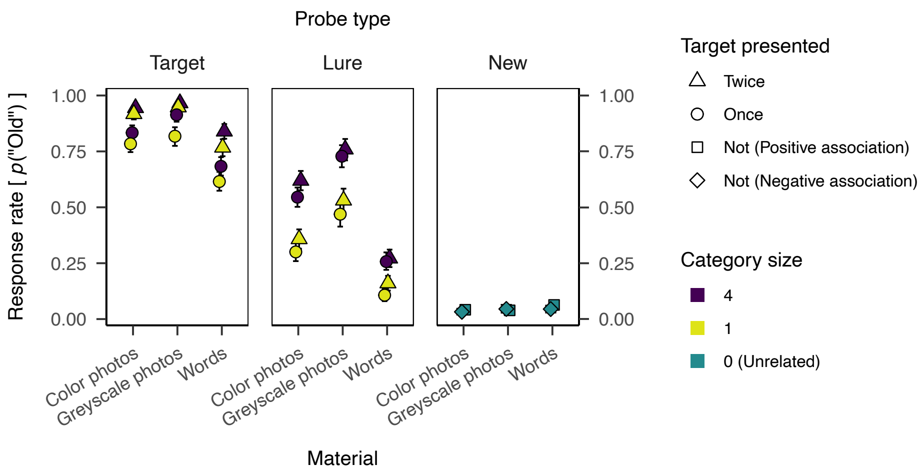

Figure 5.2: Rate of "Old"-responses as a function of category size and number of target presentations for each material. Based on previous research we predicted that the number of target presentations would increase "Old"-response rates to targets but not to lures. Additionally, we predicted that "Old"-response rates to targets and lures would increase with category size. Unrelated new probes do not fully conform to the factorial experimental design because they cannot be ascribed to one of the two presented categories. Hence, there is no corresponding target and the category size is zero. They can, however, be ascribed to positive and negative association trials, see Procedure for details. We expected no difference in "Old"-response rates between trial types. Point represent condition means, error bars represent bootstrap confidence intervals based on 10,000 samples.

5.5.1 False recognition rates

True and false recognition is typically quantified relative to the baseline of "Old"-response rates to unrelated new probes to account for guessing

Table 5.1: Adjusted false recognition rates for conditions when targets were presented once on the study list.

Category size	Material	M	95% CI	$t(80)$	p_{adj}	BF_{10}
1	Words	0.04	$[0.01, \infty]$	2.78	.010	8.69
	Greyscale photos	0.43	$[0.37, \infty]$	14.52	$< .001$	2.04×10^{21}
	Color photos	0.26	$[0.21, \infty]$	10.52	$< .001$	1.30×10^{14}
4	Words	0.19	$[0.14, \infty]$	8.31	$< .001$	8.19×10^{9}
	Greyscale photos	0.69	$[0.62, \infty]$	22.63	$< .001$	2.97×10^{33}
	Color photos	0.50	$[0.44, \infty]$	17.08	$< .001$	3.26×10^{25}

Note. The rate of false recognition was adjusted by subtracting "Old"-response rates to unrelated new probes from those of lures. p_{adj} = p-values Bonferroni-adjusted for 3 comparisons within each category size; for the Bayesian analysis we used a scale of $r = \sqrt{2}/2$ for the prior distribution. All tests are one-sided.

or memory background noise. Unrelated new probes do not fully conform to the factorial experimental design because they cannot be ascribed to one of the two presented categories. Hence, there is no corresponding target and the category size is zero. They can, however, be ascribed to positive and negative association trials, see Procedure for details. We expected no difference in "Old"-response rates between trial types. Hence, we first analyzed response rates to unrelated new probes to rule out any differences between positive and negative association trials, Figure 5.2. A 2 (*Trial type*: Positive vs. negative association) × 3 (*Material*: Color photos vs. Greyscale photos vs. Words)-ANOVA[3] provided evidence against any difference in "Old"-response rates between trial types. Overall, the "Old"-response rate to unrelated new probes was low, $M = 0.04$, 95% CI $[0.04, \infty]$, $t(80) = 8.12$, $p < .001$, $\text{BF}_{10} = 1.35 \times 10^3$. The Bayesian analysis indicated that the data supported absence of a main effect of trial type ($F(1, 80) = 0.85$, $MSE = 0.01$, $p = .359$, $\hat{\eta}^2_G = .002$, $\text{BF}_{01} = 6.39$) as well as an interaction with material, $F(1.69, 135.06) = 1.13$, $MSE = 0.01$, $p = .319$, $\hat{\eta}^2_G = .003$, $\text{BF}_{01} = 10.44$. Finally, "Old"-response rates appeared to be comparable across materials, $F(1.86, 148.41) = 1.35$, $MSE = 0.01$, $p = .262$, $\hat{\eta}^2_G = .006$, $\text{BF}_{01} = 8.54$.

For all subsequent analyses, we assessed the rate of adjusted false recog-

[3] For the Bayesian analysis we used a scale of $r = 0.5$ for the prior distribution, because we expected, if any, small effects.

Table 5.2: Pairwise comparisons of adjusted "Old"-response rates between category size 4 and 1 for target and lure probes and each material.

Probe type	Material	ΔM	95% CI	$t(80)$	p	p_{adj}	BF_{10}
Lure	Words	0.13	$[0.10, \infty]$	6.56	< .001	< .001	4.63×10^{6}
	Greyscale photos	0.24	$[0.20, \infty]$	9.56	< .001	< .001	1.97×10^{12}
	Color photos	0.25	$[0.22, \infty]$	12.40	< .001	< .001	3.81×10^{17}
Target	Words	0.07	$[0.04, \infty]$	3.65	< .001	.008	96.14
	Greyscale photos	0.06	$[0.03, \infty]$	3.79	< .001	< .001	148.59
	Color photos	0.04	$[0.02, \infty]$	2.85	.003	.001	10.45

Note. p_{adj} = p-values Bonferroni-adjusted for 3 comparisons within each probe type; for the Bayesian analysis we used a scale of $r = \sqrt{2}/2$ for the prior distribution. All tests are one-sided.

nition by subtracting "Old"-response rates to unrelated new probes from those to lures. To compare our findings to the results of previous studies we examined adjusted false recognition rates when targets were presented only once, Table 5.1. We detected false recognition in all conditions. Word lists with four exemplars yielded false recognition rates that were comparable to those previously reported for DRM lists (e.g, Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008; Flegal et al., 2010); for color and greyscale photos of everyday objects, however, we found considerably larger rates. Moreover, we detected false recognition even when the study lists included only a single related item, albeit for words the evidence for false recognition was weak.

5.5.2 Target repetition and category size

Next, we assessed the effects of our experimental manipulations on adjusted rates "Old"-responses in a 2 (*Probe type*: Target vs. Lure) × 2 (*Target presentations*: Once vs. Twice) × 2 (*Category size*: 1 vs. 4) × 3 (*Material*: Color photos vs. Greyscale photos vs. Words)-ANOVA. The number of target presentations affected adjusted "Old"-response rates to target and lure probes differently, $F(1,80) = 24.00$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .007$, $\text{BF}_{10} = 179.93$. As predicted, adjusted "Old"-response rates for targets increased when we presented targets on the study list twice rather than once, $\Delta M = 0.13$, 95%

CI $[0.11, \infty]$, $t(80) = 10.60$, $p < .001$, $\mathrm{BF}_{10} = 1.82 \times 10^{14}$. Unlike in previous experiments on long-term memory, repeated presentations also increased the adjusted rates for lures, although not as much as to targets, $\Delta M = 0.06$, 95% CI $[0.03, \infty]$, $t(80) = 3.91$, $p < .001$, $\mathrm{BF}_{10} = 215.76$. As predicted, the adjusted "Old"-response rates also increased when we increased the category size from 1 to 4 exemplars, but the magnitude of the effect varied across combinations of probe types and materials, $F(1.78, 142.80) = 9.18$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .006$, $\mathrm{BF}_{10} = 9.89$, Table 5.2. The response rate increase for targets was modest and the difference between materials small. The effect on lures, in comparison to targets, was 2–5 times larger and differed substantially between materials: For words the effect was only half of that for color and greyscale photographs. The data provided moderate to overwhelming evidence for the absence of other interaction effects, Table C.1.

To summarize, we found false recognition with all materials even when only a single related category exemplar was presented as part of the study list. As predicted, increasing the category size from one to four exemplars caused a large increase in false recognition and a small increase in true recognition. Conversely, presenting a target on the study list a second time caused a large increase in true recognition but also a small increase in false recognition.

5.5.3 State-trace analysis

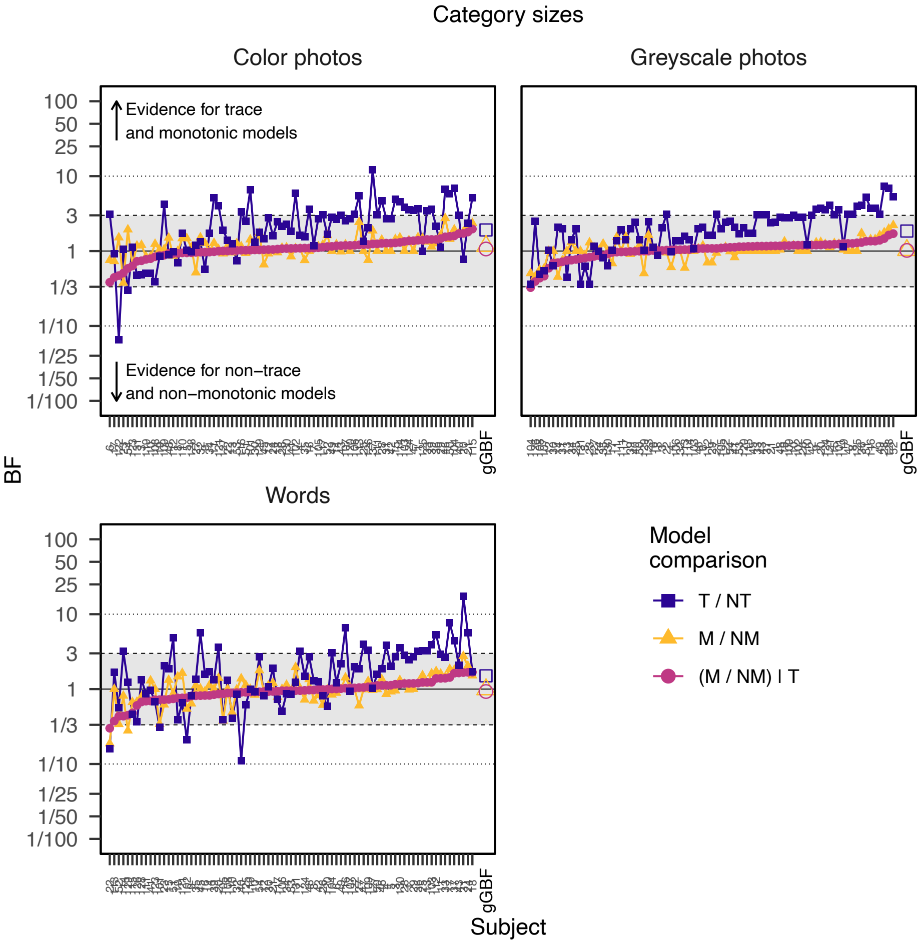

Figure 5.3: Bayes factors for individual participants in favor of the trace model (T) representing the assumption that true and false recognition always increase as studied category size increases in both dimensions relative to the non-trace model (NT), which posits a violation of the trace assumption in at least one dimension (T vs. NT) or in favor of a monotonic model of true and false recognition (M) relative to a non-monotonic model (NM) with (M vs. NM | T) and without incorporating the trace assumption (M vs. NM). Higher values indicate support for trace and monotonic models. Participants are ordered according to their evidential value for the monotonic relative to the non-monotonic model given the trace assumption. Big unfilled points represent the geometric mean of the group Bayes factor (gGBF).

Table 5.3: Results of the Bayesian state-trace analysis for each stimulus material.

	Trace model			Monotonic trace model		
Material	$\text{gGBF}_{\text{T/NT}}$	$\text{GBF}_{\text{T/NT}}$	ABF	$\text{gGBF}_{\text{(M/NM)} \mid \text{T}}$	$\text{GBF}_{\text{(M/NM)} \mid \text{T}}$	ABF
Color photos	1.91	6.40×10^{22}	13.50	1.07	169.60	1.25×10^{-3}
Greyscale photos	1.85	4.27×10^{21}	12.57	1.02	4.78	1.71
Words	1.50	2.16×10^{14}	14.22	0.93	2.36×10^{-3}	0.06

Note. The trace model (T) constrains true and false recognition to increase as studied category size increases in both dimensions and is complementary to the non-trace model (NT). The monotonic model (M | T) enforces a monotonic association between true and false recognition and is complementary to the non-monotonic model (NM | T); both models additionally enforced the trace constraint. gGBF = geometric mean of the group Bayes factor; GBF = group Bayes factor; ABF = non-parametric aggregated Bayes factor in favor of homogeneity relative to heterogeneity.

The results of the Bayesian state-trace analysis of the probit-transformed proportions of participants' "Old"-responses are shown in Figures 5.4 and 5.3; the statistical results are summarized in Table 5.3. We analyzed the data for each material separately.

Individual participants' responses were barely informative to the comparison of monotonic and non-monotonic models. To increase the informational value of the model comparison we incorporated the trace assumption into both models. As expected, this additional assumption was supported by most participants' data in all materials. In particular colored and greyscale objects revealed that most participants' responses conformed to predictions of the trace model; only four participant's responses to words and two participants' responses to greyscale objects provided modest evidence for a violation of the trace assumption. Accordingly, the aggregated data indicated that no participant's responses violated the trace assumption, $\mathrm{ABF}^{(\text{Words, T})} = 14.22$, $\mathrm{ABF}^{(\text{Greyscale photos, T})} = 12.57$, $\mathrm{ABF}^{(\text{Color photos, T})} = 13.50$. In aggregate the data overwhelmingly favored the trace model, $\mathrm{GBF}^{(\text{Words})}_{\mathrm{T/NT}} = 2.16 \times 10^{14}$, $\mathrm{GBF}^{(\text{Greyscale photos})}_{\mathrm{T/NT}} = 4.27 \times 10^{21}$, $\mathrm{GBF}^{(\text{Color photos})}_{\mathrm{T/NT}} = 6.40 \times 10^{22}$.

Unfortunately, incorporation of the trace constraint did not improve the evidential value of our analyses. The geometric group Bayes factors for all materials were inconclusive. In aggregate the data strongly favored the non-monotonic model $\mathrm{GBF}_{(\mathrm{NM/M})\,|\,\mathrm{T}} = 422.94$-to-1 for words but favored the monotonic model $\mathrm{GBF}_{(\mathrm{M/NM})\,|\,\mathrm{T}} = 4.78$-to-1 and $\mathrm{GBF}_{(\mathrm{M/NM})\,|\,\mathrm{T}} = 169.60$-to-1 for greyscale and color photos, respectively. Note that the group Bayes factor assumes homogeneity of participants' responses, that is, that the true association between true and false recognition of all participants is either monotonic or non-monotonic. However, the aggregated data provided substantial evidence against homogeneous monotonicity for the color photographs ($\mathrm{ABF}^{(\mathrm{M}\,|\,\mathrm{T})} = 1.25 \times 10^{-3}$) and was inconclusive for the greyscale photographs, $\mathrm{ABF}^{(\mathrm{M}\,|\,\mathrm{T})} = 1.71$.

To summarize, while the responses to the categorized word lists provide evidence for multiple latent causes. The results are less clear cut for color and greyscale photographs but favor a single latent dimension. Hence, accounting for the observed responses to categorized words constitutes a strong test of the single-process global matching perspective of false recognition. Responses to greyscale and colored objects on the other hand do themselves do not necessitate the assumption of a second latent variable and, hence, are a priori consistent with single-process models. As the next section will show, this does not, however, guarantee that a specific instantiation of a single-process model will be able to adequately describe the results.

5.5.4 Cognitive modelling

We fit a simplified version of the conjoint recognition model (see Chapter 4) to test the selective influence of our experimental manipulations on the memory-related parameter estimates. We then fit the Generalized Context Model to explore the models ability to describe the observed results.

5.5.4.1 Conjoint-recognition model

Based on our previous work, we used the model with constraints on V and G ($\mathscr{M}_{VG}^{\text{CRM}}$) as a baseline for our model comparisons. This model constrained V to increase from one to two target presentations within participants, while being unaffected by category size. It also constrained G to be unaffected by the number of target presentations, but to increase from category size 1 to 4 within participants. We compared this baseline model to models that relaxed either the constraint on V, but maintained the constraint on G ($\mathscr{M}_{G}^{\text{CRM}}$), or relaxed the constraint on G, but maintained the constraint on V ($\mathscr{M}_{V}^{\text{CRM}}$). We did not constrain parameters across materials.

The baseline CRM with guessing parameters b constrained to be equal across trial types within participants described the data well—in accordance with our previous analysis that found no difference in "Old"-response rates between trial types, Figure 5.4A.

The estimates of the population-level model parameters suggested that our selective influence manipulations, by and large, had the predicted effects, Table 5.4. First, repeating targets on the study list increased estimates of verbatim memory whereas increasing the category size did not. Conversely, increasing the category size increased estimates of gist memory whereas presenting targets repeatedly did not. These conclusions were also confirmed by our model comparisons. The constraint on V was overwhelmingly supported $\text{BF}_{VG/G} = 6.08 \times 10^{18}$ $[1.19 \times 10^{16},\ 6.39 \times 10^{20}]$-to-1 and

Table 5.4: Median posterior parameter estimates of the population probability medians $\widetilde{\mu}^{(\theta)}$ (95% HDI in parentheses) for the minimally constrained Conjoint Recognition model $\mathscr{M}_0^{\text{CRM}}$.

Material	Target presented	Category size	G	V
Color photos	Once	1	.25 [.20, .31]	.71 [.63, .78]
	Once	4	.53 [.46, .60]	.64 [.54, .72]
	Twice	1	.32 [.26, .37]	.90 [.85, .94]
	Twice	4	.62 [.56, .68]	.89 [.82, .95]
Greyscale photos	Once	1	.44 [.38, .51]	.66 [.56, .76]
	Once	4	.75 [.68, .82]	.69 [.52, .83]
	Twice	1	.51 [.44, .57]	.90 [.84, .95]
	Twice	4	.77 [.71, .83]	.87 [.77, .95]
Words	Once	1	.03 [.01, .06]	.57 [.51, .64]
	Once	4	.18 [.13, .24]	.58 [.50, .65]
	Twice	1	.09 [.06, .13]	.74 [.68, .80]
	Twice	4	.20 [.14, .25]	.81 [.75, .87]

Note. We constrained the populations means of the parameter b for guessing "Old" to be equal across trial types within participants, $\widetilde{\mu}^{(b)}_{\text{Color photos}} = .03\ [.02, .04]$, $\widetilde{\mu}^{(b)}_{\text{Greyscale photos}} = .04\ [.02, .05]$, $\widetilde{\mu}^{(b)}_{\text{Words}} = .05\ [.03, .07]$.

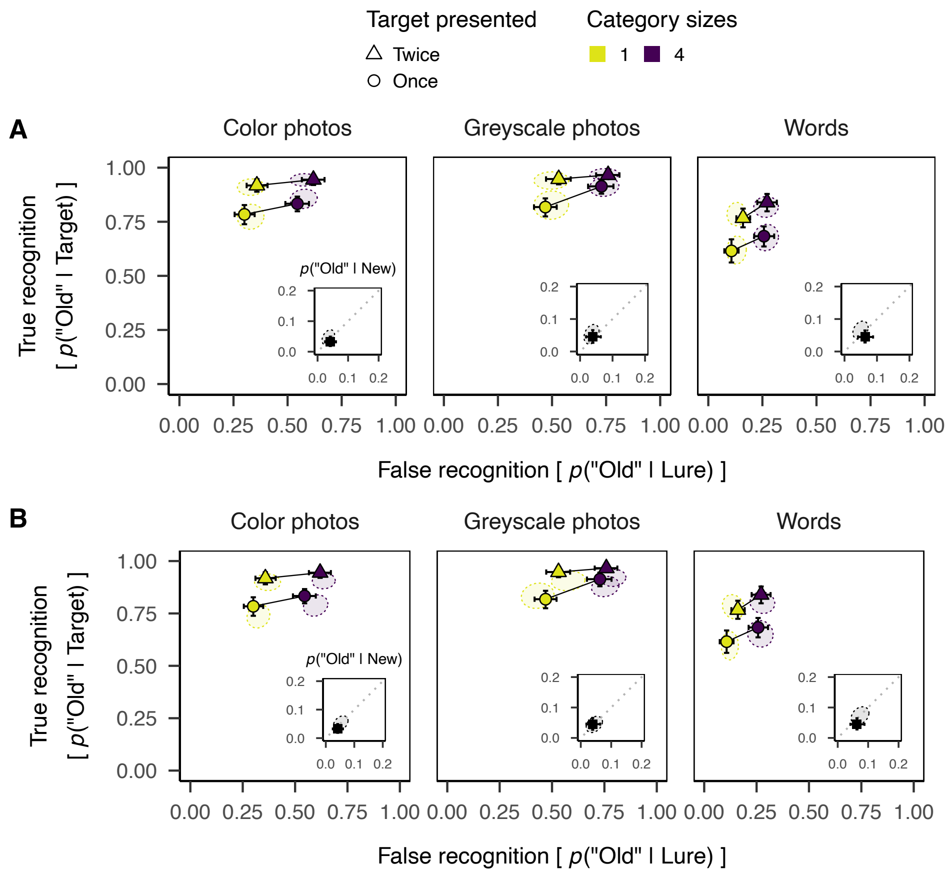

Figure 5.4: State-trace plot of averaged responses and posterior predictions of $\mathscr{M}_{VG}^{\text{CRM}}$ (**A**) and $\mathscr{M}_{0}^{\text{GCM}}$ (**B**). Points represent average observed rates of *old*-responses; error bars indicate 95% bootstrap confidence intervals based on 10,000 bootstrap samples. Ellipses represent multivariate normal-approximations to 95% credible regions posterior predictions. The inset shows the rate of "Old"-responses to unrelated new probes for positive (X-axis) and negative association trials (Y-axis).

the constraint on G was supported $\mathrm{BF}_{VG/V} = 1.20 \times 10^9$ $[1.27 \times 10^7,\ 1.43 \times 10^{11}]$-to-1. Visual inspection (Figure 5.4A) and quantitative assessment of posterior predictions indicated that all models describe the data adequately, Figure C.1.

Taken together these results support the validity of the gist and verbatim memory estimates in short-term false recognition. Thus, these results further support the applicability of models of long-term recognition to paradigms with much shorter study lists and retention intervals. Finally, the CRM fits indicate that the multiple latent dimensions identified by our state-trace analysis for the word material are consistent with independent verbatim and gist representations.

5.5.4.2 Generalized context model

Global matching models posit that the multiple latent dimensions are the result of separable contributions of exact and partially matched memory traces, see Chapter 4. To assess the adequacy of this alternative account, we next fit the GCM $\mathscr{M}_0^{\mathrm{GCM}}$ with parameters $\mu^{(c)}$, $\mu^{(c')}$, $\mu^{(m')}$, $\mu^{(k)}$, $\mu^{(\gamma)}$, and $\mu^{(b)}$ free to vary across materials. For comparison we fit a model that constrained memory strength and dissimilarity sensitivity to be equal, $\mathscr{M}_\lambda^{\mathrm{GCM}}$.

Estimates of the population-level model parameters indicate that processing of words differed from that of color and greyscale photographs, Table 5.5. For color and greyscale photographs, parameter estimates were similar.[4] In the unconstrained model $\mathscr{M}_0^{\mathrm{GCM}}$, the most notable descriptive difference was a slightly larger memory strength for color photographs—but the interval estimates indicated no difference. Nonetheless, only for color photographs the increase in memory strength differed from the baseline value of 1. In $\mathscr{M}_\lambda^{\mathrm{GCM}}$, the same difference was evident in λ, but again the interval estimates overlapped. The estimates suggest that a common set of parameters may suffice to account for the responses to both photographic materials. The estimates of the lapse rates p_{lapse} were low and provided only limited evidence for a pure guessing state as the posterior probability amassed at zero.

[4]Not that estimates of average inter-category distance d_{inter} and lapse rates p_{lapse} trade-off to explain "Old"-responses to unrelated new exemplars and, therefore, the estimates for d_{inter} are likely unreliable as also indicated by the large estimate uncertainty—except for color photographs where we used an informative prior based on or previous work.

Table 5.5: Median posterior parameter estimates of the population means in natural scale $\exp(\mu^{(\theta)})$ and $\exp(\mu^{(p_{lapse})})$ (95% HDI in parentheses) for the unconstrained Generalized Context Model ($\mathcal{M}_0^{\text{GCM}}$) and the constraint model assuming $\lambda = c = m$, $\mathcal{M}_\lambda^{\text{GCM}}$.

Parameter	Materials		
	Color photos	Greyscale photos	Words
$\mathcal{M}_0^{GCM}$			
c	1.42 [1.18, 1.68]	1.78 [1.40, 2.18]	3.65 [3.05, 4.42]
c'	1.82 [1.33, 2.58]	1.81 [1.40, 2.32]	4.23 [2.96, 8.98]
m'	2.71 [1.61, 4.30]	1.13 [0.77, 1.96]	1.09 [0.78, 1.64]
k	0.28 [0.18, 0.38]	0.20 [0.12, 0.28]	0.72 [0.58, 0.86]
γ	0.90 [0.74, 1.09]	1.14 [0.92, 1.36]	1.63 [1.21, 2.13]
p_{lapse}	0.04 [0.02, 0.07]	0.04 [0.02, 0.08]	0.12 [0.08, 0.16]
d_{inter}	6.26 [5.88, 6.61]	6.15 [4.05, 11.09]	4.59 [2.22, 10.49]
$\mathcal{M}_\lambda^{GCM}$			
λ	1.23 [1.01, 1.48]	2.07 [1.63, 2.59]	3.90 [3.24, 4.76]
λ'	1.41 [1.09, 1.88]	2.07 [1.63, 2.59]	4.60 [3.26, 7.28]
k	0.45 [0.36, 0.55]	0.56 [0.41, 0.72]	9.64 [7.43, 12.22]
γ	0.99 [0.82, 1.17]	1.01 [0.82, 1.22]	1.32 [1.00, 1.68]
p_{lapse}	0.04 [0.01, 0.06]	0.04 [0.01, 0.07]	0.12 [0.08, 0.15]
d_{inter}	6.19 [5.83, 6.53]	4.64 [3.49, 7.01]	4.44 [2.18, 9.79]

Parameter estimates for words, on the other hand, were markedly different. Estimates of dissimilarity sensitivity, the criterion k, and the probability of lapses p_{lapse} was higher than for the photographic materials. Given recent evidence that word identification may be an all-or-none process (e.g., Swagman et al., 2015; Aust & Stahl, 2016), it seems plausible that the rapid presentation rate in the current study may have caused a non-negligible lapse rate and, at the same time, high discriminability when the stimuli were processed.

In sum, the estimates corroborate the results of our state-trace analysis and suggest that the multidimensionality of responses to the word material may be due to attentional lapses and consequent random guessing. Compared to our results in long-term memory, the only notable difference was the absence of a differentiation effect, as indicated by the negligible differences between c and c' or λ and λ'.

The fit of the unconstrained model $\mathscr{M}_0^{\text{GCM}}$ is shown in Figure 5.4B—the predictions of the constrained model $\mathscr{M}_\lambda^{\text{GCM}}$ were highly similar. The model exhibited some quantitative misfit. Mirroring our findings from long-term memory, for color photographs the model again predicted too little true and too much false recognition when more than one category exemplar was presented. For color photographs and words there was a tendency to over-predict "Old" responses to unrelated new distractors. However, keeping in mind the constraint exerted by the inter-item similarities as well as our simplifying assumptions, we conclude that the fit of the GCM to the group-level data is satisfactory.

5.6 Discussion

The aim of this study was to assess the influence of categorical relations between stimuli on verbal and visual short-term memory. To do so, we performed a short-term false recognition experiment with categorical verbal and photographic materials. In an extension of prior work on long-term memory, we used selective influence manipulations of gist and verbatim memory and fit CRM and GCM to the responses. First, we observed false recognition effects following a filled 4 s-retention interval for all materials, even when the study list included only one exemplar from the probed category. Particularly the latter finding is noteworthy because to our knowledge all previous studies on false recognition in short-term memory presented at least 4 related items. As we will discuss shortly, it remains an open question to what extent encoding noise rather than semantic or categorical interference contributes to these effects. Second, as expected we found that presenting 4 instead of 1 category exemplar increased true as well as false recognition; CRM parameter estimates confirmed that these increases reflected a selective influence on gist memory. Moreover, we found that that presenting study list items twice instead of once also increased true as well as false recognition. Although this result seems to contradict the hypothesized selective influence on verbatim memory, CRM parameter estimates indicated that the data supported a selective influence on verbatim memory.

Despite the success of our selective influence manipulations, state-trace analyses indicated that the observed true and false recognition rates for the photographic materials were not diagnostic with respect to the number of latent causes. Only responses to the verbal material favored multiple latent causes. These findings indicate that, in particular, responses to the verbal material provide a challenging testbed for the GCM-account of false recognition. Indeed, as in our previous work in long-term memory, GCM was able to describe the observed rates of true and false recognition to the photographic material adequately but exhibited non-trivial misfit for the verbal material. Additionally exploration suggests that this misfit may be attributed to attentional lapses during the rapid study list presentation that result in random guessing at test. We further discuss the implications of this finding below. After making an allowance for attentional lapses, GCM described the our results adequately albeit the fits could have been closer. It is worth noting, however, that the fits reported are likely a lower bound because we relied on group-level estimates of inter-item similarities form an independent sample rather than individual estimates for each participant (p. 289, Nosofsky et al., 2011).

5.6.1 Theoretical implications

As we have shown in Chapter 4, in long-term memory GCM requires a differentiation extension to account for the selective increase in true recognition caused by presenting targets five times compared to once. Here, we tested whether differentiation also occurs over the short term. Our results indicate that presenting targets twice compared to once caused no notable differentiation of memory representations as also indicated by observed the increase in false recognition. At this point the implications of this finding remain somewhat ambiguous. On the one hand it is possible that a single target repetition is insufficient to produce detectable differentiation effects. In line with this interpretation in our experiment on long-term false recognition we presented targets five times—a typical procedure in the study of differentiation (e.g., Criss, 2006; Criss, 2010; Criss et al., 2012). Also, the rapid presentation of the study list may have been less conducive to memory consolidation and differentiation than the slower presentation rate and longer inter-stimulus intervals typically used in long-term memory experiments. On the other hand, a lack of differentiation could be indicative of a qualitative difference between the operating principles and processes of short- and long-term memory. Exploring these issues further is an interesting avenue for future research.

The contribution of random guessing due to loss of information from short-term memory performance has been a corner stone of recent theorizing on short-term memory, and in particular visual short-term memory (e.g., Adam et al., 2017; Bays et al., 2009; Miner et al., 2019; van den Berg et al., 2012; Zhang & Luck, 2008). As noted above, in our modelling we made an allowance for random guessing at test. Resembling the debate in visual short-term memory, the CRM naturally predicts random guessing, whereas GCM assumes that all responses are based on a memory signal and ordinarily denies any role of random guessing. It could be argued that by incorporating a guessing process in GCM we have fundamentally altered the nature of the model. We would counter this assertion by pointing out that we attribute the cause for random guessing to the encoding stage, rather than to a maintenance failure as assumed by slots-models of short-term memory. Specifically, we assume that during the rapid visual presentation of study list items, participants on some trials fail to attend the stimulus display and hence do not encode the items into memory (cf., p. 381, Donkin & Nosofsky, 2012b; p. 140; van den Berg et al., 2014). Our analysis indicate that such attentional lapses affect primarily processing of verbal material but less so of photographs. This finding is in line with research suggesting that, in contrast to images, words may be processed in an all-or-none fashion (Swagman et al., 2015). Moreover, an as-of-yet unpublished experiment from our lab shows that under the conditions of the current experiment photographs may be partially or noisily encoded, whereas words are not

(Aust & Stahl, 2016). These findings, too, suggest that words are either represented fully and accurately or not at all and substantiate the assumption that random guessing may have contributed to our results. Hence, our assumption of attentional lapses and random guessing at test are not at odds with the nature of GCM or models of short-term memory that posit the absence of guessing due to loss of information from short-term memory.

The results from our unpublished experiment further indicate that encoding noise likely contributed to the high rates of false recognition for photographic material observed here. It remains to be seen to what degree the observed false recognition rates are caused by semantic and categorical similarity or encoding noise—especially those for study lists, which included only one exemplar from the probed category. Based on the robust false recognition effects for verbal material, we are however confident that our results are not solely the result of encoding noise. Also, our unpublished results indicate the need for a refinement to our modelling approach: GCM assumes error-free encoding, but this assumption appears to be violated, at least under the current rapid encoding conditions. The Noisy Exemplar Model (Kahana & Sekuler, 2002) is closely related to GCM and relaxes this assumption. We leave it for future research to explore whether model fits can be improved by accounting for encoding noise.

5.6.2 Conclusions

We have presented the results from an experiment on visual and verbal false recognition over the short-term. Our results are consistent with previous research which suggests that short-term false recognition follows very similar principles as in long-term memory. Moreover, we demonstrated that GCM, a global matching model that successfully accounts for a variety of findings from short- and long-term memory, can also account for short-term false recognition. Thus, we conclude that the fits reported here indicate that GCM is a serious alternative to dual-process accounts of false recognition in short-term memory and, we hope, will inspire further research.

References

Abadie, M., & Camos, V. (2019). False memory at short and long term. *Journal of Experimental Psychology: General*, *148*(8), 1312–1334. https://doi.org/10.1037/xge0000526

Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97. https://doi.org/10.1016/j.cogpsych.2017.07.001

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature News*, *508*(7496), 312. https://doi.org/10.1038/508312a

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, *36*(1), 74–81. https://doi.org/10.3758/MC.36.1.74

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown* [R package version 0.1.0.9485]. https://github.com/crsh/papaja

Aust, F., & Stahl, C. (2016). *False recognition in short-term memory may be due to encoding errors* [Poster presented at the Annual Meeting of the Psychonomic Society 2016, Boston, MA].

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 1–11. https://doi.org/10.1167/9.10.7

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working Memory Has Better Fidelity Than Long-Term Memory: The Fidelity Constraint Is Not a General Property of Memory After All. *Psychological Science*, *30*(2), 223–237. https://doi.org/10.1177/0956797618813538

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4–4. https://doi.org/10.1167/11.5.4

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*(1), 160–179. https://doi.org/10.1037/0033-295X.106.1.160

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 307–327. https://doi.org/10.1037/0278-7393.27.2.307

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*(1), 37–60. https://doi.org/10.3758/BF03210740

Coane, J. H., McBride, D. M., Raulerson, B. A., & Scott, J. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*(1), 62–70. https://doi.org/10.1027/1618-3169.54.1.62

Cowan, N. (1999). An Embedded-Processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (1st ed., pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*(4), 461–478. https://doi.org/10.1016/j.jml.2006.08.003

Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 484–499. https://doi.org/10.1037/a0018435

Criss, A. H., & Koop, G. J. (2015). Differentiation in episodic memory. In J. G. W. Raaijmakers, A. H. Criss, R. Goldstone, R. M. Nosofsky, & M. Steyvers (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin* (pp. 112–125). Psychology Press.

Criss, A. H., Wheeler, M. E., & McClelland, J. L. (2012). A Differentiation Account of Recognition Memory: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *25*(3), 421–435. https://doi.org/10.1162/jocn_a_00292

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, *21*(2), 142–145. https://doi.org/10.3758/BF03202725

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dimsdale-Zucker, H. R., Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2018). Serial position-dependent false memory effects. *Memory*, *0*(0), 1–13. https://doi.org/10.1080/09658211.2018.1513039

Donkin, C., & Nosofsky, R. M. (2012b). The structure of short-term memory scanning: An investigation using response time distribution models. *Psychonomic Bulletin & Review*, *19*(3), 363–394. https://doi.org/10.3758/s13423-012-0236-8

Festini, S. B., & Reuter-Lorenz, P. A. (2013). The short- and long-term consequences of directed forgetting in a working memory task. *Memory*, *21*(7), 763–777. https://doi.org/10.1080/09658211.2012.754900

Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1331–1338. https://doi.org/10.1037/a0019903

Flegal, K. E., & Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Memory & Cognition*, (42), 1–11. https://doi.org/10.3758/s13421-013-0391-9

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*(Supplement C), 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Gronau, Q. F., & Singmann, H. (2018). *Bridgesampling: Bridge sampling for marginal likelihoods and bayes factors* [R package version 0.6-0]. https://CRAN.R-project.org/package=bridgesampling

Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Tenzing: Documenting contributorship using CRediT. *MetaArXiv*. https://doi.org/10.31222/osf.io/b6ywe

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A Multidimensional Scaling Database with Similarity Ratings for 240 Object Categories from the Massive Memory Picture Database. *PLoS ONE*, *9*(11), e112644. https://doi.org/10.1371/journal.pone.0112644

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256–281. https://doi.org/10.1037/a0028860

Iidaka, T., Harada, T., & Sadato, N. (2014). False memory for face in short-term memory and neural activity in human amygdala. *Brain Research*, *1591*, 74–85. https://doi.org/10.1016/j.brainres.2014.10.003

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The Mind and Brain of Short-Term Memory. *Annual Review of Psychology*, *59*(1), 193–224. https://doi.org/10.1146/annurev.psych.59.103006.093615

Jou, J., Arredondo, M. L., Li, C., Escamilla, E. E., & Zuniga, R. (2016). The effects of increasing semantic-associate list length on the Deese–Roediger–McDermott false recognition memory: Dual false-memory process in retrieval from sub- and supraspan lists. *The Quarterly Journal of Experimental Psychology*, *0*(0), 1–18. https://doi.org/10.1080/17470218.2016.1222446

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*(18), 2177–2192. https://doi.org/10.1016/S0042-6989(02)00118-9

Kelly, M. A., Mewhort, D. J. K., & West, R. L. (2017). The memory tesseract: Mathematical equivalence between composite and separate storage memory models. *Journal of Mathematical Psychology*, *77*, 142–155. https://doi.org/10.1016/j.jmp.2016.10.006

Landau, W. M. (2018). The drake r package: A pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *3*(21). https://doi.org/10.21105/joss.00550

Lenth, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.1.2]. https://CRAN.R-project.org/package=emmeans

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), nn.3655. https://doi.org/10.1038/nn.3655

Macé, A.-L., & Caza, N. (2011). The role of articulatory suppression in immediate false recognition. *Memory*, *19*(8), 891–900. https://doi.org/10.1080/09658211.2011.613844

McDermott, K. B., & Watson, J. M. (2001). The Rise and Fall of False Recall: The Impact of Presentation Duration. *Journal of Memory and Language*, *45*(1), 160–176. https://doi.org/10.1006/jmla.2000.2771

Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods*, *45*(2), 344–354. https://doi.org/10.3758/s13428-012-0255-4

Miner, A., Schurgin, M., & Brady, T. F. (2019). Is working memory inherently more 'precise' than long-term memory? Extremely high fidelity visual long-term memories for frequently encountered objects. *PsyArXiv*. https://doi.org/10.31234/osf.io/qr6b9

Morey, R. D., & Rouder, J. N. (2015). *Bayesfactor: Computation of bayes factors for common designs* [R package version 0.9.12-2]. https://CRAN.R-project.org/package=BayesFactor

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269. https://doi.org/10.3758/BF03213879

Nairne, J. S. (2002). Remembering Over the Short-Term: The Case Against the Standard Model. *Annual Review of Psychology*, *53*(1), 53–81. https://doi.org/10.1146/annurev.psych.53.100901.135131

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (2011a). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Nosofsky, R. M. (2016, January 1). An Exemplar-Retrieval Model of Short-term Memory Search: Linking Categorization and Probe Recognition. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 47–84). Academic Press. https://doi.org/10.1016/bs.plm.2016.03.002

Nosofsky, R. M., Cao, R., Harding, S. M., & Shiffrin, R. M. (2020). Modeling short- and long-term memory contributions to recent event recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000812

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1524–1539. https://doi.org/10.1037/xlm0000015

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315. https://doi.org/10.1037/a0022494

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Palmeri, T. J. (2014). An Exemplar-Based Random-Walk Model of Categorization and Recognition. In J. Busemeyer, J. Townsend, Z. Wang, & A. Eidels (Eds.), *Mathematical and Computational Models of Cognition*. Oxford University Press.

O'Donnell, R. E., Clement, A., & Brockmole, J. R. (2018). Semantic and functional relationships among objects increase the capacity of visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1151–1158. https://doi.org/10.1037/xlm0000508

Olszewska, J. M., Reuter-Lorenz, P. A., Munier, E., & Bendler, S. A. (2015). Misremembering what you see or hear: Dissociable effects of modality on short- and long-term false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1316–1325. https://doi.org/10.1037/xlm0000115

Osth, A. F., & Dennis, S. (2020). Global matching models of recognition memory. *PsyArXiv*. https://doi.org/10.31234/osf.io/mja6c

Quinlan, P. T., & Cohen, D. J. (2016). The Precategorical Nature of Visual Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000274

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Reyna, V. F., & Brainerd, C. J. (1995a). Fuzzy-trace theory: Some foundational issues. *Learning and Individual differences*, *7*(2), 145–162. https://doi.org/10.1016/1041-6080(95)90028-4

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., & McDermott, K. B. (2000). Tricks of Memory. *Current Directions in Psychological Science*, *9*(4), 123–127. https://doi.org/10.1111/1467-8721.00075

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Schurgin, M. W., & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition*, *23*(7), 859–862. https://doi.org/10.1080/13506285.2015.1093243

Schurgin, M. W., & Flombaum, J. I. (2018). Visual working memory is more tolerant than visual long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(8), 1216–1227. https://doi.org/10.1037/xhp0000528

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2019). Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. *Nature Human Behaviour*, 325472. https://doi.org/10.1101/325472

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2017). *Afex: Analysis of factorial experiments* [R package version 0.17-4]. https://github.com/singmann/afex

Stahl, C., & Heycke, T. (2016). Evaluative Conditioning with Simultaneous and Sequential Pairings Under Incidental and Intentional Learning Conditions. *Social Cognition*, *34*(5), 382–412. https://doi.org/10.1521/soco.2016.34.5.382

Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 570–586. https://doi.org/10.1037/0278-7393.34.3.570

Stahl, C., & Klauer, K. C. (2009). Measuring phantom recollection in the simplified conjoint recognition paradigm. *Journal of Memory and Language*, *60*(1), 180–193. https://doi.org/10.1016/j.jml.2008.08.001

Swagman, A. R., Province, J. M., & Rouder, J. N. (2015). Performance on Perceptual Word Identification is Mediated by Discrete States. *Psychonomic Bulletin & Review*, *22*(1), 265–273. https://doi.org/10.3758/s13423-014-0670-x

Tehan, G. (2010). Associative relatedness enhances recall and produces false memories in immediate serial recall. *Canadian Journal of Experimental Psychology*, *64*(4), 266–272. https://doi.org/10.1037/a0021375

Tse, C.-S., Li, Y., & Altarriba, J. (2011). The effect of semantic relatedness on immediate serial recall and serial recognition. *The Quarterly Journal of Experimental Psychology*, *64*(12), 2425–2437. https://doi.org/10.1080/17470218.2011.604787

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

van Vugt, M. K., Sekuler, R., Wilson, H. R., & Kahana, M. J. (2013). An electrophysiological signature of summed similarity in visual working memory. *Journal of Experimental Psychology: General*, *142*(2), 412–425. https://doi.org/10.1037/a0029759

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wickham, H. (2016a). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. http://ggplot2.org

Wickham, H. (2016b). *Purrr: Functional programming tools* [R package version 0.2.2]. https://CRAN.R-project.org/package=purrr

Wickham, H. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions* [R package version 0.6.1]. https://CRAN.R-project.org/package=tidyr

Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation* [R package version 0.5.0]. https://CRAN.R-project.org/package=dplyr

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11. https://doi.org/10.1167/4.12.11

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Conclusions

Chapter 6

Conclusions

FREDERIK AUST

In Chapter 1, I laid out the motivation for my work: Despite some notable counterexamples, the theoretical and empirical exchange between the fields of learning and memory is limited. I discussed the historical origins and epistemological difference that arguably contributed to this disconnect. Learning and memory research traditionally sought explanations at different levels of analysis—the computational level and the representational and algorithmic level, respectively (Marr, 1982). I rejected the argument that different levels of analysis prevent or oppose theoretical exchange and argued that both fields have much to gain from joint efforts to explain their respective phenomena. Contemporary learning researchers, like memory researchers, are interested in representations and algorithms and I identified explanations at this level as a promising starting point for integrative theoretical work.

In the work presented here, I explored whether learning and memory may be conceptualized as distinct algorithms that operate on the same representation of past experiences (cf. p. 707, Nosofsky, 1988). The identification of a representational format that is applicable to a broad range of phenomena would be theoretically interesting because it may serve as a basis for broad-scoped, integrative, well-constrained, explanations. In Chapter 2, I reviewed representational and process assumptions in learning and memory, by the example of evaluative conditioning and false recognition, and identified important similarities in the theoretical debates. Both fields distinguish between single- and dual-process models. Among those dual-process models a finer distinction can be made between dual-retrieval models and dual-representation models. Dual-retrieval models assume that dissociations arise at the response stage of the processing cascade and are due to different retrieval mechanisms. Dual-representation models, in contrast, attribute dissociations to different encoding processes that yield qualitatively distinct representations. I argued that models which attribute dissociations to later stages are more parsimonious and should be preferred until refuted and based on my review of the literature I concluded that the available evidence currently does not necessitate dual-representations.

Finally, I identified a common element in several successful theories of learning and memory: They postulate an informationally rich unitary representation of past events combined with parallel similarity-based retrieval. This representational format satisfies Marr's principle of least commitment, which states that a domain-general representation likely stores stimulus information after minimal preprocessing because unprocessed representations are conducive to the flexible deployment of different algorithms to meet the demands of different tasks (pp. 485-486, Marr, 1976). Thus, a rich unitary representation is conducive to an integrative theoretical approach to learning and memory. Specifically, I identified global matching memory models and their exemplar representation as a promising candidate for a common representational substrate that satisfies the principle of least commitment.

I then presented two cases in which exemplar-based global matching models, which take characteristics of the stimulus material and context into account, suggest parsimonious explanations for empirical dissociations in evaluative conditioning (EC; Chapter 3) and long-term false recognition (Chapter 4). These explanations suggest reinterpretations of findings that are commonly taken as evidence for dual-process and in particular dual-representation models.

In EC, dissociations between US expectancy and CS evaluation are typically interpreted as demonstration that EC is resistant to extinction, and consequently, that EC is driven by a simple association-based learning process. We tested whether these results are instead, information dissociations that are caused by different affordances of the dependent measures. Based on this hypothesis, we conducted simulations using the exemplar-based global matching model MINERVA 2, which were subsequently corroborated by three experiments. The results suggest that CS evaluations are by default integrative judgments—summaries of large portions of the learning history—whereas US expectancy reflects momentary judgments that focus on recent events. As the simulations showed, the different summaries can be retrieved from memory by using and reinstating appropriate contextual cues. Hence, dissociations between US expectancy and CS evaluation are consistent with single-process memory models and do not necessitate dual-representation assumptions.

Dissociations between true and false recognition are often attributed to independent verbatim and gist memory traces. Gist traces are conceptual summaries; their retrieval promotes true as well as false recognition. In contrast, verbatim traces are detailed reflections of an episode and support only true recognition. In a theoretical model analysis and an experiment we illustrate that a simpler exemplar-based global-matching explanation can also account for these findings. We argue that dissociations between true and false recognition result from distinct patterns of probe-trace similarities: False recognition results from deceptive familiarity caused by partial matches with similar (but non-identical) traces. Exact matches, on the other hand, are unique to true recognition and make an independent contribution to familiarity. This explanation is corroborated by an adequate fit of the Generalized Context Model (GCM) to our experimental data. Hence, dissociations between true and false recognition are consistent with a single-process memory models and do not necessitate dual-representation assumptions.

Finally, I report an experiment that shows that GCM also accounts for false recognition in short-term memory (Chapter 5). False recognition in short-term memory poses a problem to models positing that short-term memory is insulated from long-term memory and thereby rule out effects of semantic and categorical relations on memory performance. In contrast, unitary memory models naturally account for false recognition effects in short- and long-term memory. Illustrating the broad explanatory scope of exemplar-based global matching models, GCM was able to account for the rates of true and false recognition observed in our experiment. Hence, GCM can account for false recognition over the short and long term and is thus a candidate for a unitary memory model.

Taken together, this work provides further evidence that it is not necessary to assume dual-representations; in all cases, we were able to reattribute dissociations, which had previously been located at encoding, to the response stage. Moreover, our modeling results illustrates the broad explanatory scope and the integrative potential of exemplar-based global matching models.

6.1 Future directions

These findings are encouraging but need to be developed further. In Chapter 3, we presented a first attempt to formalize a memory-based model of EC. As discussed, this model must be considered a proof-of-principle. More theoretical as well as experimental work is needed to develop this approach into a full theory of EC (Stahl & Aust, 2018). An important next step is to identify a set of experimental results that define the scope of the phenomenon and could serve as benchmarks for theories of EC (cf. Oberauer et al., 2018; Jamieson et al., 2012). Such a set of benchmarks would enable further simulation work that could inform the refinement of our model. More broadly, I expect that the elaboration of an exemplar-based account of EC, and the identification of information dissociations, will encourage more careful considerations about the measures used to asses attitudes or liking and how participants use them. The so-gained knowledge about the dependent measures will in turn help to further develop formal process models of EC that are needed to investigate how the information that drives evaluations is represented:

> *The process models that would help make sense out of data on trace features need not be complex or highly sophisticated or "correct." But they must enter the picture in some form. Even a bad process model is better than none at all. It would help make clear the logic of the method of specifying trace properties, aid com- munication, and facilitate cross-comparisons of data obtained with different methods. Moreover, it can be improved, revised, or re- placed with a better one. A nonexistent model cannot. (p. 297, Tulving & Bower, 1974)*

Similarly, in Chapter 4 we derived theoretically, and corroborated empirically, the need for a differentiation mechanism in GCM. Inspired by an observation in previous research, we proposed a psychologically plausible and mathematically simple implementation by scaling the exponential similarity gradient to a proper probability density function. Although this proposal was in part inspired by previous research, strongly supported by our data, and fit the data of our subsequent experiment in short-term memory, it needs to be tested further. First, it should be confirmed that fixing the scaling of the similarity gradient does not impair the models ability to account for previous results, for example from categorization or perception. It is encouraging that our differentiation mechanism does not change the predictions of the model unless a subsets of memory traces are selectively strengthened, for example by extended training, or weakened, for example by study-test lag. Hence, an obvious next step is to explore whether the model can still explain the results from previous experiments that examine forgetting or selectively strengthen subsets of the study material. Second, as previously discussed in differentiation models (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997), differentiation emerges mechanistically as noisily encoded memory traces are updated with additional study. It would be interesting to explore the functional relationship between the differentiation effects in these models and our implementation in GCM. Relatedly, it could be interesting to explore more principled extensions of GCM based on related models that allow for noisy representations (e.g., Kahana & Sekuler, 2002).

At a broader level, our findings on long-term memory highlight the importance of computational level analysis. The detailed analysis of the stimulus space that is necessary to apply GCM was a critical to its success. Using essentially the same approach Schurgin et al. (2019) were recently able to develop a compelling unitary storage account for a broad range of short-term memory data called Target Confusability Competition (TCC). The success of these models shows that a disregard of the precise characteristics of the environment can result in overly complex cognitive process assumptions. Or, put differently,

> *By concentrating on purely cognitive-processing stages, researchers have largely ignored the fact that allegedly cognitive phenomena (biases, preparedness for learning) can be already built into the environmental stimulus input that impinges on the human mind. (p. 69, Fiedler, 2016)*

In Chapter 5, we demonstrated that GCM can account for short-term false recognition and suggest that GCM may be a serious contender for a unitary model of short- and long-term memory. Although this contention is supported by independent research (Nosofsky, Cox, et al., 2014; Nosofsky et al., 2020), it too requires additional testing. Following the approach taken in previous research, it would be interesting to test whether the model can simultaneously account for short- and long-term false recognition in the same subjects. Such modeling may also further elucidate the role of differentiation in short-term memory, which we were unable to explore conclusively. Finally, it will be interesting to see how GCM relates to the recently proposed TCC (Schurgin et al., 2019) given that both models make very similar assumptions about stimulus representations.

6.2 On the parsimony of least commitment

I want to end with some brief thoughts on the parsimony of least-committed representations. Exemplar models of memory have an impressive record accounting for findings from domains as diverse as attention, learning, categorization, and memory (Cox & Shiffrin, 2017; Logan, 2002; Osth & Dennis, 2020; Schmidt et al., 2016). These achievements are, of course, the best testament to the potential of exemplar representations. I have additionally argued from a system design perspective, that exemplars satisfy Marr's principle of least commitment and, thus, qualify as domain-general representational format. However, it is precisely this assumption of retaining rich memory traces of each episode that may seem prohibitive with respect to the models theoretical parsimony as well as psychological and biological plausibility.

In the context of single- and dual-process models one may be skeptical that a single-process model with least-committed representations and complex processing capabilities would be more parsimonious than a dual-representation model that assumes an inflexible process with simple lossy representations. Although dual-representation models are not necessarily more complex, as I noted in Chapter 2, there is typically a close relationship between competing theories: dual-process models postulate one process that is identical or closely related to the process posited by a competing single-process model. That is, the second inflexible process with simple lossy representations is an add-on that necessarily increases the complexity of the model. Thus, although parsimonious dual-representation models are conceivable, in practice least-committed single-process models are more parsimonious.

But are exemplar representations psychologically and biologically plausible? Does it make sense to assume that every experience leaves a distinct trace in memory and is retrievable with an appropriate memory cue? Could such a memory system be implemented in the brain? A recent theoretical analysis shows that MINERVA 2 and other exemplar models can be implemented as distributed neural models (Kelly et al., 2017). Moreover, through a combination of neuroanatomical and -computational methods, the dentate gyrus of the hippocampus has been identified as a likely locus of the formation of exemplar-like representations (Kumaran et al., 2016). Although more often than not we complain about lapses, distortions, and other failures, by introspection most people can attest to the impressive storage capacity of human memory. The ability to retain detailed representations of large numbers of objects and scene has also been demonstrated in experimental studies (e.g., Brady et al., 2008; Konkle et al., 2010). But probably the most convincing case for the massive capacity of human memory is made by individuals with highly superior autobiographical memories (HSAM; Parker et al., 2006). These individuals exhibit an impressive memory for events of their own life. Given a date they can name the day of the week, the clothes they wore, and details of what happened without practice or the use of mnemonics. For those details that are independently verifiable LePort et al. (2012) found the reports to be correct 97% of the time. Interestingly, the memory of HSAM individuals may not be fundamentally different from that of normal healthy adults. A series of false memory studies revealed that HSAM individuals are no less susceptible to false memories caused by post-event misinformation or associatively related word lists (Patihis et al., 2013). Hence, exemplar representations cannot be dismissed by appealing to the implausibility of the implementation in the brain.

Despite their merits and evidence in favor of rich representations, efforts to identify simpler representational formats should continue. A well-adapted cognitive architecture matches the complexity of its representations to the informational value of the environment:

> *There is a point where too much information and too much information processing can hurt. Cognition is the art of focusing on the relevant and deliberately ignoring the rest. (p. 20-21, Gigerenzer & Todd, 1999)*

The key to optimizing psychological fitness is to develop *appropriately* rich representations.

References

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature News*, *508*(7496), 312. https://doi.org/10.1038/508312a

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. https://doi.org/10.1037/rev0000076

Fiedler, K. (2016). Functional research and cognitive-process research in behavioural science: An unequal but firmly connected pair. *International Journal of Psychology*, *51*(1), 64–71. https://doi.org/10.1002/ijop.12163

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Tenzing: Documenting contributorship using CRediT. *MetaArXiv*. https://doi.org/10.31222/osf.io/b6ywe

Jamieson, R. K., Crump, M. J. C., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning & Behavior*, *40*(1), 61–82. https://doi.org/10.3758/s13420-011-0046-2

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*(18), 2177–2192. https://doi.org/10.1016/S0042-6989(02)00118-9

Kelly, M. A., Mewhort, D. J. K., & West, R. L. (2017). The memory tesseract: Mathematical equivalence between composite and separate storage memory models. *Journal of Mathematical Psychology*, *77*, 142–155. https://doi.org/10.1016/j.jmp.2016.10.006

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene Memory Is More Detailed Than You Think: The Role of Categories in Visual Long-Term Memory. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, *20*(7), 512–534. https://doi.org/10.1016/j.tics.2016.05.004

LePort, A. K. R., Mattfeld, A. T., Dickinson-Anson, H., Fallon, J. H., Stark, C. E. L., Kruggel, F., Cahill, L., & McGaugh, J. L. (2012). Behavioral and neuroanatomical investigation of Highly Superior Autobiographical Memory (HSAM). *Neurobiology of Learning and Memory*, *98*(1), 78–92. https://doi.org/10.1016/j.nlm.2012.05.002

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*(2), 376–400. https://doi.org/10.1037/0033-295X.109.2.376

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *275*(942), 483–519. https://doi.org/10.1098/rstb.1976.0090

Marr, D. (1982). *Vision*. W. H. Freeman and Company.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*(4), 724–760. https://doi.org/10.1037/0033-295X.105.4.734-760

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M., Cao, R., Harding, S. M., & Shiffrin, R. M. (2020). Modeling short- and long-term memory contributions to recent event recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000812

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1524–1539. https://doi.org/10.1037/xlm0000015

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Osth, A. F., & Dennis, S. (2020). Global matching models of recognition memory. *PsyArXiv*. https://doi.org/10.31234/osf.io/mja6c

Parker, E. S., Cahill, L., & McGaugh, J. L. (2006). A Case of Unusual Autobiographical Remembering. *Neurocase*, *12*(1), 35–49. https://doi.org/10.1080/13554790500473680

Patihis, L., Frenda, S. J., LePort, A. K. R., Petersen, N., Nichols, R. M., Stark, C. E. L., McGaugh, J. L., & Loftus, E. F. (2013). False memories in highly superior autobiographical memory individuals. *Proceedings of the National Academy of Sciences*, *110*(52), 20947–20952. https://doi.org/10.1073/pnas.1314373110

Schmidt, J. R., De Houwer, J., & Rothermund, K. (2016). The Parallel Episodic Processing (PEP) model 2.0: A single computational model of stimulus-response binding, contingency learning, power curves, and mixing costs. *Cognitive Psychology*, *91*, 82–108. https://doi.org/10.1016/j.cogpsych.2016.10.004

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2019). Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. *Nature Human Behaviour*, 325472. https://doi.org/10.1101/325472

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Stahl, C., & Aust, F. (2018). Evaluative Conditioning as Memory-Based Judgment. *Social Psychological Bulletin*, *13(3)*, e28589. https://doi.org/10.5964/spb.v13i3.28589

Tulving, E., & Bower, G. H. (1974). The Logic of Memory Representations. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 265–301). Academic Press. https://doi.org/10.1016/S0079-7421(08)60457-0

Bibliography

Abadie, M., & Camos, V. (2019). False memory at short and long term. *Journal of Experimental Psychology: General*, *148*(8), 1312–1334. https://doi.org/10.1037/xge0000526

Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97. https://doi.org/10.1016/j.cogpsych.2017.07.001

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature News*, *508*(7496), 312. https://doi.org/10.1038/508312a

Amsel, A. (1967). Partial Reinforcement Effects On Vigor and Persistence: Advances in Frustration Theory Derived from a Variety of Within-Subjects Experiments. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (pp. 1–65). Academic Press. https://doi.org/10.1016/S0079-7421(08)60512-5

Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of a week. *Psychonomic Bulletin & Review*, 1–6. https://doi.org/10.3758/s13423-015-0800-0

Anderson, J. R. (1990). *The adaptive character of thought*. L. Erlbaum Associates.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123. https://doi.org/10.1037/h0033773

Arndt, J., & Hirshman, E. (1998). True and False Recognition in MINERVA2: Explanations from a Global Matching Perspective. *Journal of Memory and Language*, *39*(3), 371–391. https://doi.org/10.1006/jmla.1998.2581

Ashby, F. G., & Maddox, W. T. (1993). Relations between Prototype, Exemplar, and Decision Bound Models of Categorization. *Journal of Mathematical Psychology*, *37*(3), 372–400. https://doi.org/10.1006/jmps.1993.1023

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, *36*(1), 74–81. https://doi.org/10.3758/MC.36.1.74

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (pp. 89–195). Academic Press.

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown* [R package version 0.1.0.9485]. https://github.com/crsh/papaja

Aust, F., Haaf, J. M., & Stahl, C. (2018). A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000600

Aust, F., & Stahl, C. (2016). *False recognition in short-term memory may be due to encoding errors* [Poster presented at the Annual Meeting of the Psychonomic Society 2016, Boston, MA].

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion*, *34*(1), 128–143. https://doi.org/10.1080/02699931.2019.1604322

Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996). No Evidence for Modulation of Evaluative Flavor–Flavor Associations in Humans. *Learning and Motivation*, *27*(2), 200–241. https://doi.org/10.1006/lmot.1996.0012

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy*, *10*(4), 179–199. https://doi.org/10.1016/0146-6402(88)90014-8

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). *Behaviour Research and Therapy*, *33*(7), 825–831. https://doi.org/10.1016/0005-7967(95)00021-O

Baeyens, F., Díaz, E., & Ruiz, G. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition and Emotion*, *19*(2), 245–268. https://doi.org/10.1080/02699930441000300

Baeyens, F., Hendrickx, H., Crombez, G., & Hermans, D. (1998). Neither Extended Sequential nor Simultaneous Feature Positive Training Result in Modulation of Evaluative Flavor–Flavor Conditioning in Humans. *Appetite*, *31*(2), 185–204. https://doi.org/10.1006/appe.1998.0167

Baeyens, F., Vansteenwegen, D., & Hermans, D. (2009a). Associative learning requires associations, not propositions. *Behavioral and Brain Sciences*, *32*(02), 198. https://doi.org/10.1017/S0140525X09000867

Baeyens, F., Vansteenwegen, D., & Hermans, D. (2009b). Associative learning requires associations, not propositions. *Behavioral and Brain Sciences*, *32*(02), 198. https://doi.org/10.1017/S0140525X09000867

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, *19*(2), 137–181. https://doi.org/10.1016/0022-2496(79)90016-6

Bar-Anan, Y., & Moran, T. (2018). Simple First: A Skeleton for an Evaluative Learning Model. *Social Psychological Bulletin*, *13*(3), 1–23. https://doi.org/10.5964/spb.v13i3.28761

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, *4*(2), 207–228. https://doi.org/10.1016/0010-0285(73)90012-1

Bartlema, A., Lee, M., Wetzels, R., & Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: Case studies in selective attention and representation in category learning. *Journal of Mathematical Psychology*, *59*, 132–150. https://doi.org/10.1016/j.jmp.2013.12.002

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 1–11. https://doi.org/10.1167/9.10.7

Bechtel, W., & Shagrir, O. (2015). The Non-Redundant Contributions of Marr's Three Levels of Analysis for Explaining Information-Processing Mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322. https://doi.org/10.1111/tops.12141

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working Memory Has Better Fidelity Than Long-Term Memory: The Fidelity Constraint Is Not a General Property of Memory After All. *Psychological Science*, *30*(2), 223–237. https://doi.org/10.1177/0956797618813538

Biegler, P., & Vargas, P. (2013). Ban the Sunset? Nonpropositional Content and Regulation of Pharmaceutical Advertising. *The American Journal of Bioethics*, *13*(5), 3–13. https://doi.org/10.1080/15265161.2013.776127

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation*, *39*(1), 58–70. https://doi.org/10.1016/j.lmot.2007.03.003

Bodner, G. E., & Dypvik, A. T. (2005). Masked priming of number judgments depends on prime validity and task. *Memory & Cognition*, *33*(1), 29–47. https://doi.org/10.3758/bf03195295

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*(1), 80–99. https://doi.org/10.1037/0033-2909.114.1.80

Bouton, M. E. (2007). *Learning and behavior: A contemporary synthesis*. Sinauer Associates.

Bower, G. H. (2000). A brief history of memory research. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 3–32). Oxford University Press.

Bower, J. M., & Beeman, D. (1998). *The book of GENESIS: Exploring realistic neural models with the GEneral NEural SImulation System* (2nd ed.). TELOS.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4–4. https://doi.org/10.1167/11.5.4

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brainerd, C. J., Gomes, C. F. A., & Nakamura, K. (2015). Dual Recollection in Episodic Memory. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000084

Brainerd, C. J., Nakamura, K., Chang, M., & Bialer, D. M. (2019). Verbatim editing: A general model of recollection rejection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1776–1790. https://doi.org/10.1037/xlm0000684

Brainerd, C. J., Nakamura, K., & Murtaza, Y. A. (2020). Explaining complementarity in false memory. *Journal of Memory and Language*, *112*, 104105. https://doi.org/10.1016/j.jml.2020.104105

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-Trace Theory and Children's False Memories. *Journal of Experimental Child Psychology*, *71*(2), 81–129. https://doi.org/10.1006/jecp.1998.2464

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory* (Vol. 14). Oxford University Press.

Brainerd, C. J., & Reyna, V. F. (2018). Complementarity in false memory illusions. *Journal of Experimental Psychology: General*, *147*(3), 305–327. https://doi.org/10.1037/xge0000381

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*(3), 343–382. https://doi.org/10.1037/0033-2909.134.3.343

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-Recognition Reversal: When Similarity is Distinctive. *Journal of Memory and Language*, *34*(2), 157–185.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*(1), 160–179. https://doi.org/10.1037/0033-295X.106.1.160

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*(4), 762–784. https://doi.org/10.1037/0033-295X.110.4.762

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 307–327. https://doi.org/10.1037/0278-7393.27.2.307

Brainerd, C., & Reyna, V. (2002). Fuzzy-Trace Theory and False Memory. *Current Directions in Psychological Science*, *11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old–new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 77–102. https://doi.org/10.1037/0278-7393.26.1.77

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Buchanan, L., Brown, N. R., Cabeza, R., & Maitson, C. (1999). False Memories and Semantic Lexicon Arrangement. *Brain and Language*, *68*, 172–177. https://doi.org/10.1006/brln.1999.2072

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition*, *51*(3), 251–261. https://doi.org/10.1016/S0278-2626(03)00030-7

Capaldi, E. J. (1967). A Sequential Hypothesis of Instrumental Learning. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (pp. 67–156). Academic Press. https://doi.org/10.1016/S0079-7421(08)60513-7

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*(1), 37–60. https://doi.org/10.3758/BF03210740

Coane, J. H., McBride, D. M., Raulerson, B. A., & Scott, J. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*(1), 62–70. https://doi.org/10.1027/1618-3169.54.1.62

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, *30*(7), 1138–1147. https://doi.org/10.3758/BF03194331

Corneille, O., & Stahl, C. (2019a). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

Corneille, O., & Stahl, C. (2019b). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

Cowan, N. (1998). Children's Memories According to Fuzzy-Trace Theory: An Endorsement of the Theory's Purpose and Some Suggestions to Improve Its Application. *Journal of Experimental Child Psychology*, *71*(2), 144–154. https://doi.org/10.1006/jecp.1998.2466

Cowan, N. (1999). An Embedded-Processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (1st ed., pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In J.-C. Lacaille, W. S. Sossin, V. F. Castellucci, & S. Belleville (Eds.), *Progress in Brain Research* (pp. 323–338). Elsevier.

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, *119*(3), 480–499. https://doi.org/10.1037/a0027791

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. https://doi.org/10.1037/rev0000076

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*(4), 461–478. https://doi.org/10.1016/j.jml.2006.08.003

Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 484–499. https://doi.org/10.1037/a0018435

Criss, A. H., & Howard, M. W. (2015). Models of episodic memory. In J. R. Busemeyer, Z. J. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (pp. 165–183). Oxford University Press.

Criss, A. H., & Koop, G. J. (2015). Differentiation in episodic memory. In J. G. W. Raaijmakers, A. H. Criss, R. Goldstone, R. M. Nosofsky, & M. Steyvers (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin* (pp. 112–125). Psychology Press.

Criss, A. H., Wheeler, M. E., & McClelland, J. L. (2012). A Differentiation Account of Recognition Memory: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *25*(3), 421–435. https://doi.org/10.1162/jocn_a_00292

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, *21*(2), 142–145. https://doi.org/10.3758/BF03202725

Davis-Stober, C. P., Morey, R. D., Gretton, M., & Heathcote, A. (2016). Bayes factors for state-trace analysis. *Journal of Mathematical Psychology*, *72*, 116–129. https://doi.org/10.1016/j.jmp.2015.08.004

De Houwer, J. (1998). Leren is eenvoudig, doen is moeilijk: Een nieuwe visie op het onderscheid tussen evaluatieve en Pavloviaanse conditionering [Learning is simple, performance is difficult: A new view on the relation between evaluative and Pavlovian conditioning]. *Gedragstherapie*, *31*, 49–66.

De Houwer, J. (2007). A Conceptual and Theoretical Analysis of Evaluative Conditioning. *The Spanish Journal of Psychology*, *10*(02), 230–241. https://doi.org/10.1017/S1138741600006491

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*(1), 1–20. https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2011). Evaluative conditioning: A review of procedure knowledge and mental process theories. In T. R. Schachtman & S. Reilly (Eds.), *Applications of learning and conditioning*. Oxford University Press.

De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin*, *13(3)*, e28046. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture–picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(2), 237–242. https://doi.org/10.1037/0097-7403.26.2.237

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*, *20*(4), 631–642. https://doi.org/10.3758/s13423-013-0386-3

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853–869. https://doi.org/10.1037/0033-2909.127.6.853

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dennett, D. C. (1981). *Brainstorms: Philosophical essays on mind and psychology*. Bradford Books.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, *13*(1), 1–21. https://doi.org/10.3758/BF03193807

Dimsdale-Zucker, H. R., Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2018). Serial position-dependent false memory effects. *Memory*, *0*(0), 1–13. https://doi.org/10.1080/09658211.2018.1513039

Dodson, C. S., & Schacter, D. L. (2001). If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*(1), 155–161. https://doi.org/10.3758/BF03196152

Donkin, C., & Nosofsky, R. M. (2012a). A Power-Law Model of Psychological Memory Strength in Short- and Long-Term Recognition. *Psychological Science*, *23*(6), 625–634. https://doi.org/10.1177/0956797611430961

Donkin, C., & Nosofsky, R. M. (2012b). The structure of short-term memory scanning: An investigation using response time distribution models. *Psychonomic Bulletin & Review*, *19*(3), 363–394. https://doi.org/10.3758/s13423-012-0236-8

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209. https://doi.org/10.1037/0033-295X.106.1.180

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*(3), 286–303. https://doi.org/10.1037/0096-3445.127.3.286

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*, *115*(2), 426–446. https://doi.org/10.1037/0033-295X.115.2.426

Dunn, J. C., & Kalish, M. L. (2018). *State-trace analysis*. Springer.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*(1), 91–101. https://doi.org/10.1037/0033-295X.95.1.91

Dwyer, D. M., Jarratt, F., & Dick, K. (2007). Evaluative conditioning with foods as CSs and body shapes as USs: No evidence for sex differences, extinction, or overshadowing. *Cognition and Emotion*, *21*(2), 281–299. https://doi.org/10.1080/02699930600551592

Ebbinghaus, H. (1885). *Über das Gedächtnis [On memory]*. Duncker & Humblot.

Eliasmith, C., & Kolbeck, C. (2015). Marr's Attacks: On Reductionism and Vagueness. *Topics in Cognitive Science*, *7*(2), 323–335. https://doi.org/10.1111/tops.12133

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Festini, S. B., & Reuter-Lorenz, P. A. (2013). The short- and long-term consequences of directed forgetting in a working memory task. *Memory*, *21*(7), 763–777. https://doi.org/10.1080/09658211.2012.754900

Fiedler, K. (2016). Functional research and cognitive-process research in behavioural science: An unequal but firmly connected pair. *International Journal of Psychology*, *51*(1), 64–71. https://doi.org/10.1002/ijop.12163

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, *25*(4), 639–656. https://doi.org/10.1080/02699931.2010.513497

Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1331–1338. https://doi.org/10.1037/a0019903

Flegal, K. E., & Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Memory & Cognition*, (42), 1–11. https://doi.org/10.3758/s13421-013-0391-9

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pitbull: EC effects depend on CS–US relations. *Cognition and Emotion*, *26*(3), 534–540. https://doi.org/10.1080/02699931.2011.588687

French, R. M., & Thomas, E. (2015). Interactive Effects of Explicit Emergent Structure: A Major Challenge for Cognitive Computational Modeling. *Topics in Cognitive Science*, *7*(2), 206–216. https://doi.org/10.1111/tops.12135

Gallo, D. (2006). *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*. Psychology Press.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. Wiley.

Gast, A. (2018). A Declarative Memory Model of Evaluative Conditioning. *Social Psychological Bulletin*, *13(3)*, e28590. https://doi.org/10.5964/spb.v13i3.28590

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative Conditioning From the Perspective of the Associative-Propositional Evaluation Model. *Social Psychological Bulletin*, *13*(3), 1–33. https://doi.org/10.5964/spb.v13i3.28024

Gawronski, B., Gast, A., & De Houwer, J. (2014). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition and Emotion*, 1–15. https://doi.org/10.1080/02699931.2014.947919

Gawronski, B., Rydell, R. J., De Houwer, J., Brannon, S. M., Ye, Y., Vervliet, B., & Hu, X. (2018). Contextualized Attitude Change. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology*. Academic Press. https://doi.org/10.1016/bs.aesp.2017.06.001

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General*, *139*(4), 683–701. https://doi.org/10.1037/a0020315

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015). *Bayesian Data Analysis*. CRC press.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Glass, A. L., Holyoak, K. J., & Santa, J. L. (1979). *Cognition*. Addison-Wesley.

Greve, A., Donaldson, D. I., & van Rossum, M. C. W. (2010). A single-trace dual-process model of episodic memory: A novel computational account of familiarity and recollection. *Hippocampus*, *20*(2), 235–251. https://doi.org/10.1002/hipo.20606

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*(Supplement C), 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Gronau, Q. F., & Singmann, H. (2018). *Bridgesampling: Bridge sampling for marginal likelihoods and bayes factors* [R package version 0.6-0]. https://CRAN.R-project.org/package=bridgesampling

Groninger, L. K., & Groninger, L. D. (1980). A comparison of recognition and savings as retrieval measures: A reexamination. *Bulletin of the Psychonomic Society*, *15*(4), 263–266. https://doi.org/10.3758/BF03334527

Hanus, D. (2016). Causal reasoning versus associative learning: A useful dichotomy or a strawman battle in comparative psychology? *Journal of Comparative Psychology*, *130*(3), 241–248. https://doi.org/10.1037/a0040235

Hebscher, M., Wing, E., Ryan, J., & Gilboa, A. (2019). Rapid Cortical Plasticity Supports Long-Term Memory Formation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.09.009

Heck, D. W., & Wagenmakers, E.-J. (2016). Adjusted priors for Bayes factors involving reparameterized order constraints. *Journal of Mathematical Psychology*, *73*, 110–116. https://doi.org/10.1016/j.jmp.2016.05.004

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In S. P. Shohov (Ed.), *Advances in psychology research* (pp. 17–40). Nova Science Publishers.

Heycke, T., Aust, F., Banaji, M. R., Cone, J., Van Dessel, P., Ferguson, M., Hu, X., Jiang, C., Kurdi, B., Rydell, R. J., Spitzer, L., Stahl, C., Vitiello, C., & De Houwer, J. (in prep). *Of Two Minds: A registered replication*.

Heycke, T., Aust, F., & Stahl, C. (2017). Subliminal influence on preferences? A test of evaluative conditioning for brief visual conditioned stimuli using auditory unconditioned stimuli. *Royal Society Open Science*, *4*, 160935. https://doi.org/10.1098/rsos.160935

Heycke, T., & Gawronski, B. (2020). Co-occurrence and relational information in evaluative learning: A multinomial modeling approach. *Journal of Experimental Psychology: General*, *149*(1), 104–124. https://doi.org/10.1037/xge0000620

Heycke, T., Gehrmann, S., Haaf, J. M., & Stahl, C. (2018). Of two minds or one? A registered replication of Rydell et al. (2006). *Cognition and Emotion*, *32*(8), 1708–1727. https://doi.org/10.1080/02699931.2018.1429389

Heycke, T., & Stahl, C. (2020). No evaluative conditioning effects with briefly presented stimuli. *Psychological Research*, *84*(4), 1020–1027. https://doi.org/10.1007/s00426-018-1109-1

Hintzman, D. L. (1990). Human Learning and Memory: Connections and Dissociations. *Annual Review of Psychology*, *41*(1), 109–139. https://doi.org/10.1146/annurev.ps.41.020190.000545

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1987). Recognition and recall in MINERVA 2: Analysis of the 'recognition-failure' paradigm. *Modelling cognition* (pp. 215–229). John Wiley & Sons.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Hintzman, D. L. (1993). Twenty-Five Years of Learning and Memory: Was the Cognitive Revolution a Mistake? In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 359–391). MIT Press.

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 667–680. https://doi.org/10.1037/0278-7393.18.4.667

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Högden, F., Hütter, M., & Unkelbach, C. (2018). Does evaluative conditioning depend on awareness? Evidence from a continuous flash suppression paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(10), 1641–1657. https://doi.org/10.1037/xlm0000533

Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Tenzing: Documenting contributorship using CRediT. *MetaArXiv*. https://doi.org/10.31222/osf.io/b6ywe

Holland, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, *37*(1-2), 105–131. https://doi.org/10.1016/0010-0277(90)90020-K

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A Multidimensional Scaling Database with Similarity Ratings for 240 Object Categories from the Massive Memory Picture Database. *PLoS ONE*, *9*(11), e112644. https://doi.org/10.1371/journal.pone.0112644

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256–281. https://doi.org/10.1037/a0028860

Houwer, J. D. (2011). Why the Cognitive Approach in Psychology Would Profit From a Functional Approach and Vice Versa. *Perspectives on Psychological Science*, *6*(2), 202–209. https://doi.org/10.1177/1745691611400238

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional Versus Dual-Process Accounts of Evaluative Conditioning: I. The Effects of Co-Occurrence and Relational Information on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *43*(1), 17–32. https://doi.org/10.1177/0146167216673351

Hughes, S., Houwer, J. D., & Perugini, M. (2016). The functional-cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology*, *51*(1), 4–14. https://doi.org/10.1002/ijop.12239

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208–233. https://doi.org/10.1037/0033-295X.96.2.208

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, *33*(1), 36–67. https://doi.org/10.1016/0022-2496(89)90003-5

Hütter, M., & Rothermund, K. (2020). Automatic processes in evaluative learning. *Cognition and Emotion*, *34*(1), 1–20. https://doi.org/10.1080/02699931.2019.1709315

Hütter, M., & Sweldens, S. (2013). Implicit misattribution of evaluative responses: Contingency-unaware evaluative conditioning requires simultaneous stimulus presentations. *Journal of Experimental Psychology: General*, *142*(3), 638–643. https://doi.org/10.1037/a0029989

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*(3), 539–557. https://doi.org/10.1037/a0026477

Iidaka, T., Harada, T., & Sadato, N. (2014). False memory for face in short-term memory and neural activity in human amygdala. *Brain Research*, *1591*, 74–85. https://doi.org/10.1016/j.brainres.2014.10.003

Jamieson, R. K., Crump, M. J. C., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning & Behavior*, *40*(1), 61–82. https://doi.org/10.3758/s13420-011-0046-2

Jamieson, R. K., & Hauri, B. R. (2012). An exemplar model of performance in the artificial grammar task: Holographic representation. *Canadian Journal of Experimental Psychology*, *66*(2), 98–105. https://doi.org/10.1037/a0027023

Jamieson, R. K., & Mewhort, D. J. K. (2009a). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *The Quarterly Journal of Experimental Psychology*, *62*(3), 550–575. https://doi.org/10.1080/17470210802055749

Jamieson, R. K., & Mewhort, D. J. K. (2009b). Applying an exemplar model to the serial reaction-time task: Anticipating from experience. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1757–1783. https://doi.org/10.1080/17470210802557637

Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial-grammar task: String completion and performance on individual items. *The Quarterly Journal of Experimental Psychology*, *63*(5), 1014–1039. https://doi.org/10.1080/17470210903267417

Jenkins, H. M. (1979). Animal learning and behavior theory. In E. Hearst (Ed.), *The first century of experimental psychology* (pp. 177–228). Erlbaum Hillsdale, NJ.

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General*, *149*(2), 230–248. https://doi.org/10.1037/xge0000648

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*(5), 662–672. https://doi.org/10.3758/PBR.17.5.662

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, *96*(5), 933–948. https://doi.org/10.1037/a0014747

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*(1), 1–32. https://doi.org/10.1037/a0034190

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The Mind and Brain of Short-Term Memory. *Annual Review of Psychology*, *59*(1), 193–224. https://doi.org/10.1146/annurev.psych.59.103006.093615

Jou, J., Arredondo, M. L., Li, C., Escamilla, E. E., & Zuniga, R. (2016). The effects of increasing semantic-associate list length on the Deese–Roediger–McDermott false recognition memory: Dual false-memory process in retrieval from sub- and supraspan lists. *The Quarterly Journal of Experimental Psychology*, *0*(0), 1–18. https://doi.org/10.1080/17470218.2016.1222446

Jou, J., & Flores, S. (2013). How are false memories distinguishable from true memories in the Deese–Roediger–McDermott paradigm? A review of the findings. *Psychological Research*, *77*(6), 671–686. https://doi.org/10.1007/s00426-012-0472-6

Jozefowiez, J. (2018). Associative versus predictive processes in Pavlovian conditioning. *Behavioural Processes*, *154*, 21–26. https://doi.org/10.1016/j.beproc.2017.12.016

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*(18), 2177–2192. https://doi.org/10.1016/S0042-6989(02)00118-9

Kalish, M. L., Dunn, J. C., Burdakov, O. P., & Sysoev, O. (2016). A statistical test of the equality of latent orders. *Journal of Mathematical Psychology*, *70*, 1–11. https://doi.org/10.1016/j.jmp.2015.10.004

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.2307/2291091

Kattner, F. (2014). Reconsidering the (in)sensitivity of evaluative conditioning to reinforcement density and CS–US contingency. *Learning and Motivation*, *45*, 15–29. https://doi.org/10.1016/j.lmot.2013.09.002

Kattner, F., & Green, C. S. (2015). Cue competition in evaluative conditioning as a function of the learning process. *Acta Psychologica*, *162*, 40–50. https://doi.org/10.1016/j.actpsy.2015.09.013

Kellen, D., & Klauer, K. C. (2019). Selecting Amongst Multinomial Models: An Apologia for Normalized Maximum Likelihood. *PsyArXiv*. https://doi.org/10.31234/osf.io/9v2hd

Kelly, M. A., Mewhort, D. J. K., & West, R. L. (2017). The memory tesseract: Mathematical equivalence between composite and separate storage memory models. *Journal of Mathematical Psychology*, *77*, 142–155. https://doi.org/10.1016/j.jmp.2016.10.006

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*(2), 133–159. https://doi.org/10.1016/0749-596X(90)90069-C

Klaassen, F., Zedelius, C. M., Veling, H., Aarts, H., & Hoijtink, H. (2018). All for one or some for all? Evaluating informative hypotheses using multiple N = 1 studies. *Behavior Research Methods*, *50*(6), 2276–2291. https://doi.org/10.3758/s13428-017-0992-5

Klauer, K. C. (2009). Spontaneous evaluations. In F. Strack & J. Förster (Eds.), *Social cognition: The basis of human interaction* (pp. 199–217). Psychology Press.

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology*, *43*(4), 648–655. https://doi.org/10.1016/j.jesp.2006.06.003

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1882–1905. https://doi.org/10.1037/0096-1523.25.6.1882

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene Memory Is More Detailed Than You Think: The Role of Categories in Visual Long-Term Memory. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What Do We Know About What We Cannot Remember? Accessing the Semantic Attributes of Words That Cannot Be Recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1095–1105. https://doi.org/10.1037/0278-7393.29.6.1095

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C., & Jackson, E. M. (1999). Perceptually Based False Recognition of Novel Objects in Amnesia: Effects of Category Size and Similarity to Category Prototypes. *Cognitive Neuropsychology*, *16*(3-5), 317–341. https://doi.org/10.1080/026432999380816

Kukken, N., Hütter, M., & Holland, R. W. (2020). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating the effects of CS-US pairings and their meaning. *Cognition and Emotion*, *34*(1), 170–187. https://doi.org/10.1080/02699931.2019.1617112

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, *20*(7), 512–534. https://doi.org/10.1016/j.tics.2016.05.004

Lachman, S. J. (1997). Learning is a Process: Toward an Improved Definition of Learning. *The Journal of Psychology*, *131*(5), 477–480. https://doi.org/10.1080/00223989709603535

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550617697177

Landau, W. M. (2018). The drake r package: A pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *3*(21). https://doi.org/10.21105/joss.00550

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-8). University of Florida, Gainesville, FL.

López, F. J., & Shanks, D. R. (2008). Models of Animal Learning and Their Relations to Human Learning. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychol-*

ogy (1st ed., pp. 589–611). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.026

Lenth, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.1.2]. https://CRAN.R-project.org/package=emmeans

LePort, A. K. R., Mattfeld, A. T., Dickinson-Anson, H., Fallon, J. H., Stark, C. E. L., Kruggel, F., Cahill, L., & McGaugh, J. L. (2012). Behavioral and neuroanatomical investigation of Highly Superior Autobiographical Memory (HSAM). *Neurobiology of Learning and Memory*, *98*(1), 78–92. https://doi.org/10.1016/j.nlm.2012.05.002

Lipp, O. V., Mallan, K. M., Libera, M., & Tan, M. (2010). The effects of verbal instruction on affective and expectancy learning. *Behaviour Research and Therapy*, *48*(3), 203–209. https://doi.org/10.1016/j.brat.2009.11.002

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation*, *34*(3), 219–239. https://doi.org/10.1016/S0023-9690(03)00011-0

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 431–440. https://doi.org/10.1037/0097-7403.32.4.431

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*(2), 376–400. https://doi.org/10.1037/0033-295X.109.2.376

Love, B. C. (2015). The Algorithmic Level Is the Bridge Between Computation and Brain. *Topics in Cognitive Science*, *7*(2), 230–242. https://doi.org/10.1111/tops.12131

Lovibond, P. F. (2004). Cognitive Processes in Extinction. *Learning & Memory*, *11*(5), 495–500. https://doi.org/10.1101/lm.79604

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3–26. https://doi.org/10.1037/0097-7403.28.1.3

Ly, M., Murray, E., & Yassa, M. A. (2013). Perceptual versus conceptual interference and pattern separation of verbal stimuli in young and older adults. *Hippocampus*, *23*(6), 425–430. https://doi.org/10.1002/hipo.22110

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), nn.3655. https://doi.org/10.1038/nn.3655

Macé, A.-L., & Caza, N. (2011). The role of articulatory suppression in immediate false recognition. *Memory*, *19*(8), 891–900. https://doi.org/10.1080/09658211.2011.613844

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the Brain's Algorithm for Categorization from Its Neural Implementation. *Current Biology*, *23*(20), 2023–2027. https://doi.org/10.1016/j.cub.2013.08.035

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*(4), 335–384. https://doi.org/10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., Holden, J. E., & Shiffren, R. M. (2004). Modeling the Effects of Repetitions, Similarity, and Normative Word Frequency on Old-New Recognition and Judgments of Frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 319–331. https://doi.org/10.1037/0278-7393.30.2.319

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *275*(942), 483–519. https://doi.org/10.1098/rstb.1976.0090

Marr, D. (1982). *Vision*. W. H. Freeman and Company.

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy*, *1*(2), 57–101. https://doi.org/10.1016/0146-6402(78)90013-9

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matute, H., Lipp, O. V., Vadillo, M. A., & Humphreys, M. S. (2011). Temporal contexts: Filling the gap between episodic memory and associative learning. *Journal of Experimental Psychology: General*, *140*(4), 660–673. https://doi.org/10.1037/a0023862

Matute, H., Vegas, S., & De Marez, P.-J. (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 714–725. https://doi.org/10.1037/0278-7393.28.4.714

Mazur, J. E. (2012). *Learning and Behavior* (7th ed.). Pearson.

McClamrock, R. (1991). Marr's three levels: A re-evaluation. *Minds and Machines*, *1*(2), 185–196. https://doi.org/10.1007/BF00361036

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356. https://doi.org/10.1016/j.tics.2010.06.002

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*(4), 724–760. https://doi.org/10.1037/0033-295X.105.4.734-760

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1987). *Psychological and Biological Models*. MIT Press.

McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model: A Dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 204–218). The Guilford Press.

McDermott, K. B., & Watson, J. M. (2001). The Rise and Fall of False Recall: The Impact of Presentation Duration. *Journal of Memory and Language*, *45*(1), 160–176. https://doi.org/10.1006/jmla.2000.2771

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Hu-*

man Perception and Performance, *21*(1), 128–148. https://doi.org/10.1037/0096-1523.21.1.128

McLaren, I. P. L., Forrest, C. L. D., McLaren, R. P., Jones, F. W., Aitken, M. R. F., & Mackintosh, N. J. (2014). Associations and propositions: The case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory*, *108*, 185–195. https://doi.org/10.1016/j.nlm.2013.09.014

Merikle, P. M., & Reingold, E. M. (1998). On demonstrating unconscious perception: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General*, *127*(3), 304–310. https://doi.org/10.1037/0096-3445.127.3.304

Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods*, *45*(2), 344–354. https://doi.org/10.3758/s13428-012-0255-4

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* Henry Holt and Co. https://doi.org/10.1037/10039-000

Miller, N. E. (1959). Liberalization of basic S-R Concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science* (pp. 196–202). McGraw-Hill.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386. https://doi.org/10.1037/0033-2909.117.3.363

Miller, R. R., & Escobar, M. (2001). Contrasting Acquisition-Focused and Performance-Focused Models of Acquired Behavior. *Current Directions in Psychological Science*, *10*(4), 141–145. https://doi.org/10.1111/1467-8721.00135

Miner, A., Schurgin, M., & Brady, T. F. (2019). Is working memory inherently more ‘precise’ than long-term memory? Extremely high fidelity visual long-term memories for frequently encountered objects. *PsyArXiv*. https://doi.org/10.31234/osf.io/qr6b9

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183–198. https://doi.org/10.1017/S0140525X09000855

Moran, T., & Bar-Anan, Y. (2013a). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion*, *27*(4), 743–752. https://doi.org/10.1080/02699931.2012.732040

Moran, T., & Bar-Anan, Y. (2013b). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion*, *27*(4), 743–752. https://doi.org/10.1080/02699931.2012.732040

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2016). The Assimilative Effect of Co-Occurrence on Evaluation Above and Beyond the Effect of Relational Qualifiers. *Social Cognition*, *34*(5), 435–461. https://doi.org/10.1521/soco.2016.34.5.435

Moran, T., Hughes, S., Hussey, I., Vadillo, M. A., Olson, M. A., Aust, F., Bading, K. C., Balas, R., Benedict, T., Corneille, O., Douglas, S. B., Ferguson, M. J., Fritzlen, K. A., Gast, A., Gawronski, B., Giménez-Fernández, T., Hanusz, K., Heycke, T., Högden, F., . . . De Houwer, J. (in press). Incidental Attitude Formation via the Surveillance Task: A Preregistered Replication of Olson and Fazio (2001). *Psychological Science*. https://doi.org/10.31234/osf.io/kavx4

Morey, R. D., & Rouder, J. N. (2015). *Bayesfactor: Computation of bayes factors for common designs* [R package version 0.9.12-2]. https://CRAN.R-project.org/package=BayesFactor

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269. https://doi.org/10.3758/BF03213879

Nairne, J. S. (2002). Remembering Over the Short-Term: The Case Against the Standard Model. *Annual Review of Psychology*, *53*(1), 53–81. https://doi.org/10.1146/annurev.psych.53.100901.135131

Navarro, D. J. (2007). On the interaction between exemplar-based concepts and a response scaling process. *Journal of Mathematical Psychology*, *51*(2), 85–98. https://doi.org/10.1016/j.jmp.2006.11.003

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts.

Newell, A. (1982). The knowledge level. *Artificial Intelligence*, *18*(1), 87–127. https://doi.org/10.1016/0004-3702(82)90012-1

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*(8), 285–290. https://doi.org/10.1016/j.tics.2008.04.009

Norman, K. A., Detre, G., & Polyn, S. M. (2008). Computational models of episodic memory. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 189–225). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.011

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27. https://doi.org/10.1037/0096-1523.17.1.3

Nosofsky, R. M. (2011a). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Nosofsky, R. M. (2011b). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Nosofsky, R. M. (2016, January 1). An Exemplar-Retrieval Model of Short-term Memory Search: Linking Categorization and Probe Recognition. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 47–84). Academic Press. https://doi.org/10.1016/bs.plm.2016.03.002

Nosofsky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, *75*, 97–129. https://doi.org/10.1016/j.cogpsych.2014.08.003

Nosofsky, R. M., Cao, R., Harding, S. M., & Shiffrin, R. M. (2020). Modeling short- and long-term memory contributions to recent event recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000812

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1524–1539. https://doi.org/10.1037/xlm0000015

Nosofsky, R. M., & Gold, J. (2016). Memory strength versus memory variability in visual change detection. *Attention, Perception, & Psychophysics*, *78*(1), 78–93. https://doi.org/10.3758/s13414-015-0992-4

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315. https://doi.org/10.1037/a0022494

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Palmeri, T. J. (2014). An Exemplar-Based Random-Walk Model of Categorization and Recognition. In J. Busemeyer, J. Townsend, Z. Wang, & A. Eidels (Eds.), *Mathematical and Computational Models of Cognition*. Oxford University Press.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*(2), 530–556. https://doi.org/10.3758/s13428-017-0884-8

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations Between Categorization and Recognition in Amnesic and Normal Individuals: An Exemplar-Based Interpretation. *Psychological Science*, *9*(4), 247–255. https://doi.org/10.1111/1467-9280.00051

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 924–940. https://doi.org/10.1037/0278-7393.28.5.924

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

O'Donnell, R. E., Clement, A., & Brockmole, J. R. (2018). Semantic and functional relationships among objects increase the capacity of visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1151–1158. https://doi.org/10.1037/xlm0000508

O'Donohue, W., Ferguson, K. E., & Naugle, A. E. (2003). The structure of the cognitive revolution: An examination from the philosophy of science. *The Behavior Analyst*, *26*(1), 85–110.

Olson, M. A., & Fazio, R. H. (2001). Implicit Attitude Formation Through Classical Conditioning. *Psychological Science*, *12*(5), 413–417. https://doi.org/10.1111/1467-9280.00376

Olszewska, J. M., Reuter-Lorenz, P. A., Munier, E., & Bendler, S. A. (2015). Misremembering what you see or hear: Dissociable effects of modality on short- and long-term false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1316–1325. https://doi.org/10.1037/xlm0000115

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary Learning Systems. *Cognitive Science*, *38*(6), 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311–345. https://doi.org/10.1037/0033-295X.108.2.311

Osth, A. F., & Dennis, S. (2020). Global matching models of recognition memory. *PsyArXiv*. https://doi.org/10.31234/osf.io/mja6c

Parker, E. S., Cahill, L., & McGaugh, J. L. (2006). A Case of Unusual Autobiographical Remembering. *Neurocase*, *12*(1), 35–49. https://doi.org/10.1080/13554790500473680

Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, *114*(1), 188–201. https://doi.org/10.1037/0033-295X.114.1.188

Patihis, L., Frenda, S. J., LePort, A. K. R., Petersen, N., Nichols, R. M., Stark, C. E. L., McGaugh, J. L., & Loftus, E. F. (2013). False memories in highly superior autobiographical memory individuals. *Proceedings of the National Academy of Sciences*, *110*(52), 20947–20952. https://doi.org/10.1073/pnas.1314373110

Pavlov, I. P. (1927/1960). *Conditioned reflex: An investigation of the physiological activity of the cerebral cortex*. Dover Publications.

Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, *30*(2), 73–95. https://doi.org/10.3758/BF03192911

Peters, K. R., & Gawronski, B. (2011). Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569. https://doi.org/10.1177/0146167211400423

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Pidgeon, L. M., & Morcom, A. M. (2014). Age-related increases in false recognition: The role of perceptual and conceptual similarity. *Frontiers in Aging Neuroscience*, *6*, 283. https://doi.org/10.3389/fnagi.2014.00283

Pierce, B. H., Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2005). The modality effect in false recognition: Evidence for test-based monitoring. *Memory & Cognition*, *33*(8), 1407–1413. https://doi.org/10.3758/BF03193373

Pierce, W. D., & Cheney, C. D. (2013). *Behavior Analysis and Learning* (5th ed.). Psychology Press.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of experimental psychology. Learning, memory, and cognition*, *33*(1), 130–144. https://doi.org/10.1037/0278-7393.33.1.130

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, *2*(5), 509. https://doi.org/10.1037/0278-7393.2.5.509

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1163–1175. https://doi.org/10.1037//0096-1523.28.5.1163

Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, *17*(1), 78–99. https://doi.org/10.1037/a0025809

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.

Quinlan, P. T., & Cohen, D. J. (2016). The Precategorical Nature of Visual Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000274

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Raes, A. K., De Houwer, J., Verschuere, B., & De Raedt, R. (2011). Return of fear after retrospective inferences about the absence of an unconditioned stimulus during extinction. *Behaviour Research and Therapy*, *49*(3), 212–218. https://doi.org/10.1016/j.brat.2010.12.004

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275–280. https://doi.org/10.1037/h0027768

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rescorla, R. A. (2003). Contemporary Study of Pavlovian Conditioning. *The Spanish Journal of Psychology*, *6*(2), 185–195. https://doi.org/10.1017/S1138741600005333

Reyna, V. F., & Brainerd, C. J. (1995a). Fuzzy-trace theory: Some foundational issues. *Learning and Individual differences*, *7*(2), 145–162. https://doi.org/10.1016/1041-6080(95)90028-4

Reyna, V. F., & Brainerd, C. J. (1998). Fuzzy-Trace Theory and False Memory: New Frontiers. *Journal of Experimental Child Psychology*, *71*(2), 194–209. https://doi.org/10.1006/jecp.1998.2472

Reyna, V. F., & Brainerd, C. J. (1995b). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1–75. Retrieved July 5, 2016, from http://www.sciencedirect.com/science/article/pii/1041608095900314

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*(2), 178–191. https://doi.org/10.1037/0012-1649.30.2.178

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*(2), 95–123. https://doi.org/10.1016/S1041-6080(97)90002-9

Reyna, V. F., & Titcomb, A. L. (1997). Constraints on the suggestibility of eyewitness testimony: A fuzzy-trace theory analysis. *Intersections in basic and applied memory research*, 157–174.

Richter, J., & Gast, A. (2017). Distributed practice can boost evaluative conditioning by increasing memory for the stimulus pairs. *Acta Psychologica*, *179*, 1–13. https://doi.org/10.1016/j.actpsy.2017.06.007

Roediger, H. L. (2004). What Happened to Behaviorism. *APS Observer*, *17*(3).

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., & McDermott, K. B. (2000). Tricks of Memory. *Current Directions in Psychological Science*, *9*(4), 123–127. https://doi.org/10.1111/1467-8721.00075

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, *113*(3), 553–565. https://doi.org/10.1037/0033-2909.113.3.553

Rogers, T. T., & McClelland, J. L. (2014). Parallel Distributed Processing at 25: Further Explorations in the Microstructure of Cognition. *Cognitive Science*, *38*(6), 1024–1077. https://doi.org/10.1111/cogs.12148

Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, *8*(3), 183–189. https://doi.org/10.1002/bs.3830080302

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rumelhart, D. E., James L. McClelland, & PDP Research Group. (1987). *Foundations* (12. Aufl). MIT Press.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008. https://doi.org/10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of Two Minds: Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes. *Psychological Science*, *17*(11), 954–958. https://doi.org/10.1111/j.1467-9280.2006.01811.x

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing False Recognition in Younger and Older Adults: The Distinctiveness Heuristic. *Journal of Memory and Language*, *40*(1), 1–24. https://doi.org/10.1006/jmla.1998.2611

Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. *Neuropsychology*, *11*(3), 331. https://doi.org/10.1037/0894-4105.11.3.331

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schelling, T. C. (1981). *The strategy of conflict* (8. print). Harvard Univ. Pr.

Schmidt, J. R., De Houwer, J., & Rothermund, K. (2016). The Parallel Episodic Processing (PEP) model 2.0: A single computational model of stimulus-response binding, contingency learning, power curves, and mixing costs. *Cognitive Psychology*, *91*, 82–108. https://doi.org/10.1016/j.cogpsych.2016.10.004

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential Hypothesis Testing with Bayes Factors: Efficiently Testing Mean Differences. *Psychological Methods*. https://doi.org/10.1037/met0000061

Schurgin, M. W., & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition*, *23*(7), 859–862. https://doi.org/10.1080/13506285.2015.1093243

Schurgin, M. W., & Flombaum, J. I. (2018). Visual working memory is more tolerant than visual long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(8), 1216–1227. https://doi.org/10.1037/xhp0000528

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2019). Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. *Nature Human Behaviour*, 325472. https://doi.org/10.1101/325472

Shagrir, O., & Bechtel, W. (2018). Marr's Computational Level and Delineating Phenomena. In D. M. Kaplan (Ed.), *Explanation and Integration in Mind and Brain Science* (pp. 190–214). Oxford University Press.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, *60*(3), 291–309. https://doi.org/10.1080/17470210601000581

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367–395. https://doi.org/10.1017/S0140525X00035032

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Shevrin, H., & Dickman, S. (1980). The psychological unconscious: A necessary assumption for all psychological theory? *American Psychologist*, *35*(5), 421–434. https://doi.org/10.1037/0003-066X.35.5.421

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of experimental psychology. Learning, memory, and cognition*, *21*(2), 267–287. https://doi.org/10.1037/0278-7393.21.2.267

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review*, *3*(3), 314–321. https://doi.org/10.3758/BF03210755

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2017). *Afex: Analysis of factorial experiments* [R package version 0.17-4]. https://github.com/singmann/afex

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, *57*(4), 193–216.

Skinner, B. F. (1985). Cognitive science and behaviourism. *British Journal of Psychology*, *76*(3), 291–301. https://doi.org/10.1111/j.2044-8295.1985.tb01953.x

Skinner, B. (1938). *The Behavior of Organisms: An Experimental Analysis*. Appleton-Century.

Smith, E. R., & DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review*, *4*(2), 108–131. https://doi.org/10.1207/S15327957PSPR0402_01

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*(01), 1. https://doi.org/10.1017/S0140525X00052432

Spear, N. E., & Riccio, D. C. (1994). *Memory: Phenomena and principles*. Allyn & Bacon.

Stahl, C., & Aust, F. (2018). Evaluative Conditioning as Memory-Based Judgment. *Social Psychological Bulletin*, *13(3)*, e28589. https://doi.org/10.5964/spb.v13i3.28589

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal Evaluative Conditioning? Above-Chance CS Identification May Be Necessary and Insufficient for Attitude Learning. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000191

Stahl, C., Henze, L., & Aust, F. (2016). False memory for perceptually similar but conceptually distinct line drawings. *PsyArXiv*. https://doi.org/10.31234/osf.io/zr7m8

Stahl, C., & Heycke, T. (2016). Evaluative Conditioning with Simultaneous and Sequential Pairings Under Incidental and Intentional Learning Conditions. *Social Cognition*, *34*(5), 382–412. https://doi.org/10.1521/soco.2016.34.5.382

Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 570–586. https://doi.org/10.1037/0278-7393.34.3.570

Stahl, C., & Klauer, K. C. (2009). Measuring phantom recollection in the simplified conjoint recognition paradigm. *Journal of Memory and Language*, *60*(1), 180–193. https://doi.org/10.1016/j.jml.2008.08.001

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(2), 286–291. https://doi.org/10.1037/a0013255

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*(3), 404–420. https://doi.org/10.1037/a0016196

Stanovich, K. E. (2005). *The robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press.

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Stewart, I. (2016). The fruits of a functional approach for psychological science. *International Journal of Psychology*, *51*(1), 15–27. https://doi.org/10.1002/ijop.12184

Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, *23*(3), 165–180. https://doi.org/10.2307/2295722

Swagman, A. R., Province, J. M., & Rouder, J. N. (2015). Performance on Perceptual Word Identification is Mediated by Discrete States. *Psychonomic Bulletin & Review*, *22*(1), 265–273. https://doi.org/10.3758/s13423-014-0670-x

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review*, *18*(2), 187–209. https://doi.org/10.1177/1088868314527832

Tehan, G. (2010). Associative relatedness enhances recall and produces false memories in immediate serial recall. *Canadian Journal of Experimental Psychology*, *64*(4), 266–272. https://doi.org/10.1037/a0021375

Teige-Mocigemba, S., & Klauer, K. C. (2008). 'Automatic' evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology*, *44*(5), 1414–1417. https://doi.org/10.1016/j.jesp.2008.04.004

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall Accuracy and Illusory Memories: When More is Less. *Memory*, *7*(2), 233–256. https://doi.org/10.1080/741944069

Tolman, E. C., & Honzik, C. H. (1930). "Insight" in rats. *University of California Publications in Psychology*, *4*, 215–232.

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. I. Orientation and the short-cut. *Journal of Experimental Psychology*, *36*(1), 13–24. https://doi.org/10.1037/h0053944

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Tse, C.-S., Li, Y., & Altarriba, J. (2011). The effect of semantic relatedness on immediate serial recall and serial recognition. *The Quarterly Journal of Experimental Psychology*, *64*(12), 2425–2437. https://doi.org/10.1080/17470218.2011.604787

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*(4), 344–354. https://doi.org/10.1037/h0043150

Tulving, E., & Bower, G. H. (1974). The Logic of Memory Representations. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 265–301). Academic Press. https://doi.org/10.1016/S0079-7421(08)60457-0

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, *21*, 437–484. https://doi.org/10.1146/annurev.ps.21.020170.002253

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*(1), 122–129. https://doi.org/10.1037/h0022014

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*(6), 559–573. https://doi.org/10.1037/h0028143

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

van der Helm, P. A. (2012). Cognitive architecture of perceptual organization: From neurons to gnosons. *Cognitive Processing*, *13*(1), 13–40. https://doi.org/10.1007/s10339-011-0425-9

van Vugt, M. K., Sekuler, R., Wilson, H. R., & Kahana, M. J. (2013). An electrophysiological signature of summed similarity in visual working memory. *Journal of Experimental Psychology: General*, *142*(2), 412–425. https://doi.org/10.1037/a0029759

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 71–79. https://doi.org/10.1037/0097-7403.32.1.71

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquistion context. *Behaviour Research and Therapy*, *43*(3), 323–336. https://doi.org/10.1016/j.brat.2004.01.001

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy*, *43*(3), 357–371. https://doi.org/10.1016/j.brat.2004.02.005

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Watson, J. B. (1924). *Psychology from the Standpoint of a Behaviorist* (2nd ed.). J. B. Lippincott.

Wellek, S. (2002). *Testing Statistical Hypotheses of Equivalence* (1 edition). Chapman and Hall/CRC.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8

Wickham, H. (2016a). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. http://ggplot2.org

Wickham, H. (2016b). *Purrr: Functional programming tools* [R package version 0.2.2]. https://CRAN.R-project.org/package=purrr

Wickham, H. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions* [R package version 0.6.1]. https://CRAN.R-project.org/package=tidyr

Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation* [R package version 0.5.0]. https://CRAN.R-project.org/package=dplyr

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11. https://doi.org/10.1167/4.12.11

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A Model of Dual Attitudes. *Psychological Review*, *107*(1), 101. https://doi.org/10.1037/0033-295X.107.1.101

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*(4), 1025–1054. https://doi.org/10.1037/a0020874

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, *133*(5), 800–832. https://doi.org/10.1037/0033-2909.133.5.800

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 1022–1041. https://doi.org/10.1037/0278-7393.27.4.1022

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, *43*(3), 155–165. https://doi.org/10.1016/j.lmot.2012.02.003

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Appendix A

Appendix to Chapter 3

A.1 MINERVA 2 simulation method

We assumed that trials were encoded as combinations of stimulus and context features. CSs and USs consisted of 10 unique features coded as 1. Hence, for simplicity we assumed that all stimuli were unrelated to each other. For USs 10 additional features coded stimulus valence with 1 for positive, -1 for negative, and 0 for neutral or no valence. The contexts were represented by 10 common and 20 distinguishing features. For the first context, 10 distinguishing features were coded as 1, indicating the presence of some contextual features, and the remaining 10 features were coded as -1, indicating the absence of other contextual features. The coding was reversed for the second context. The context for end-of-study pleasantness ratings was represented by the 10 common and 40 unique features, all coded as 1—the distinguishing context features of the learning procedure were coded as 0. Thus, we assumed participants would experience the end-of-study rating procedure as markedly different from the learning procedure.

We further assumed that participants' memory initially contained information unrelated to the experiment. This assumption was implemented by starting with a memory containing 100 episodes where each feature was randomly coded as -1, 0, or 1. Each CS-US pairing was appended to the memory as a new trace and features were correctly encoded into memory with a probability of $p = 0.60$ or as 0 otherwise. We simulated 10 trials for each context (i.e. acquisition and counterconditioning or extinction). Each simulation was repeated 30 times.

To predict US expectancy and CS pleasantness ratings, we reasoned that the CS in question and the current context act as cues to recall previous pairings with USs. Hence, we used the CS and context to probe memory and computed the normalized memory echo, in which features range from -1 to 1. The normalized memory echo represents the recalled information—a mixture of all learning episodes involving the CS. We then determined the valence of the recalled content by averaging across the recalled valence-coding features. If the recalled content was positive we predicted an expectation of a positive US and a positive CS evaluation. Thus, we predicted US expectancy and CS pleasantness ratings based on the same information. This approach is essentially equivalent to predicting US expectancy from orthogonal category-specific features (e.g., human, animal, or object features).

A.2 CS-US pairing memory

Here we report the analysis of participants' US category and US identity recognition responses.

A.2.1 Experiment 1

We analyzed US category and identity recognition responses using 2 (*US valence order*: US+ US− vs. US− US+) × 2 (*Context*: First vs. Second) repeated-measures ANOVAs.

Overall, US category recognition was quite accurate. We observed a small recency effect, that is, US category recognition was somewhat better for the second ($M = .87$, SD = .19) than for the first context ($M = .78$, SD = .22), $F(1, 36) = 10.44$, $MSE = 0.03$, $p = .003$, $\hat{\eta}^2_G = .051$, $\mathrm{BF}_{10} = 76.76$. We found no noteworthy evidence for any other effects of our experimental manipulations, all $p \geq .245$, all $\mathrm{BF}_{01} \geq 2.62$.

A one-way repeated-measures ANOVA of end-of-study pleasantness ratings of US categories indicated that participants remembered the valence of the US categories, $F(1.85, 66.57) = 115.47$, $MSE = 4.48$, $p < .001$, $\hat{\eta}^2_G = .710$, $\mathrm{BF}_{10} = 3.71 \times 10^{27}$. Without any exemplars available, participants rated the animal category as more pleasant than the object category, $\Delta M = 2.81$, 95%

CI [1.94, 3.68], $t(36) = 6.54$, $p < .001$, $\mathrm{BF}_{10} = 5.01 \times 10^6$, and the human category as less pleasant than the object category, $\Delta M = -4.32$, 95% CI [−5.23, −3.42], $t(36) = -9.67$, $p < .001$, $\mathrm{BF}_{10} = 2.26 \times 10^{14}$. Thus, recognition memory for US categories may be indicative of participants' US valence memory. Note, however, that participants rated US categories after the US identity recognition assessment during which we presented arrays containing all exemplars from each US category.

Recognition accuracy for the specific USs that had been paired with CSs followed a similar pattern. Overall, US identity recognition was quite accurate in both the first ($M = .73$, SD = .25) and the second context, $M = .82$, SD = .23. However, the observed recency effect in US identity recognition appeared to be largely due to CSs that had first been paired with positive and then with negative USs, $F(1,36) = 9.48$, $MSE = 0.03$, $p = .004$, $\hat{\eta}^2_G = .029$, $\mathrm{BF}_{10} = 26.39$. Participants were less accurate to recognize the positive USs that had been paired with CSs in the first learning phase than the corresponding negative USs from the first phase, $\Delta M = 0.17$, 95% CI [0.08, 0.26], $t(36) = 4.33$, $p < .001$ (adjusted for two comparisons), $\mathrm{BF}_{10} = 219.60$. There was some evidence, however, that there was no recency effect for CSs that had first been paired with negative USs and later with positive USs, $\Delta M = 0.01$, 90% CI [−0.06, 0.07], $t(36) = -1.93$, $p = .061$ (equivalence test adjusted for three comparisons), $\mathrm{BF}_{01} = 5.45$.

The memory-based judgment perspective assumes that EC requires memory of CS and US valence. We tested whether the observed changes in CS pleasantness across contexts was contingent on memory for CS-US pairs. Due to the overall high memory accuracy only small subsamples were available to test our hypotheses. Nonetheless, we found some evidence that the observed EC effects were contingent on memory for US categories, $F(1,6) = 7.67$, $MSE = 6.52$, $p = .032$, $\hat{\eta}^2_G = .113$, $\mathrm{BF}_{10} = 5.68$. This finding also corroborates that US category recognition is indicative of US valence memory. Our analyses regarding the role of memory for US identity were inconclusive, $F(1,11) = 2.31$, $MSE = 5.29$, $p = .157$, $\hat{\eta}^2_G = .008$, $\mathrm{BF}_{01} = 1.60$.

A.2.2 Experiment 2

A.2.2.1 Confirmatory results

We analyzed US category recognition accuracy using 2 (*Valence*: Positive vs. Negative) × 2 (*Learning procedure*: Acquisition vs. Extinction) × 2 (*Context*: First vs. Second) repeated-measures ANOVA. As in Experiment 1, US category recognition was quite accurate. However, participants better remembered that no US had been presented (M = .92, SD = .18 and M = .93, SD = .20 for acquisition and extinction, respectively) than the correct US category when a CS had been paired with a US, M = .80, SD = .31 and M = .80, SD = .29 for acquisition and extinction, respectively, $\mathrm{BF}_{10} = 1.66 \times 10^{12}$. Beyond the recognition advantage for US absence, we found evidence indicating that recognition performance was comparable between the learning procedures, $\mathrm{BF}_{01} = 7.69$. We found no noteworthy evidence for any other effects of our experimental manipulations, all $\mathrm{BF}_{10} \leq 2.21$.

We analyzed US identity recognition accuracy using 2 (*Valence*: Positive vs. Negative) × 2 (*Learning procedure*: Acquisition vs. Extinction) repeated-measures ANOVA. US identity recognition, too, was quite accurate in both acquisition (M = .87, SD = .26) and extinction procedures, M = .85, SD = .27. We found no noteworthy evidence for any effects of our experimental manipulations, all $\mathrm{BF}_{10} \leq 1.39$.

A.2.2.2 Exploratory results

As in Experiment 1, a one-way repeated-measures ANOVA of end-of-study pleasantness ratings of US categories indicated that participants remembered the valence of US categories, $\mathrm{BF}_{10} = 5.96 \times 10^{51}$. Without any exemplars available, participants rated the animal category as more pleasant than the object category, $\mathrm{BF}_{10} = 5.88 \times 10^{14}$, and the human category as less pleasant than the object category, $\mathrm{BF}_{10} = 1.63 \times 10^{29}$. Thus, recognition memory for US categories may be indicative of participants' US valence memory.

Memory for CS-US pairings was too accurate to test whether the observed differences in EC effects across referenced contexts was contingent on memory for CS-US pairs.

A.2.3 Experiment 3

A.2.3.1 Confirmatory analyses

We analyzed US category recognition accuracy using 2 (*Valence*: Positive vs. Negative) × 2 (*Learning procedure*: Acquisition vs. Extinction) × 2 (*Context*: First vs. Second) × 2 (*DV order*: Pleasantness first vs. Expectancy first) ANOVA with repeated-measures on the first three factors. Again, US category recognition was quite accurate. We found that the effect of context on US category memory differed between learning procedure, $BF_{10} = 1.29 \times 10^{14}$. Unlike in Experiment 2, we found evidence indicating that the recognition advantage for US absence was dependent on the learning procedure, $BF_{10} = 250.44$. Participants best remembered that a US was absent in the acquisition procedure ($M = .89$, SD = .26); however, memory for US absence in the extinction procedure ($M = .78$, SD = .34) was comparable to the memory for the correct category when a CS had been paired with a US ($M = .76$, SD = .31, and $M = .74$, SD = .32 for acquisition and extinction, respectively). These results were not affected by DV order, $BF_{01} = 8.13$; we found evidence that there were no other effects of our experimental manipulations, all $BF_{01} \geq 6.19$.

We analyzed US identity recognition accuracy using 2 (*Valence*: Positive vs. Negative) × 2 (*Learning procedure*: Acquisition vs. Extinction) × 2 (*DV order*: Pleasantness first vs. Expectancy first) ANOVA with repeated-measures on the first two factors. US identity recognition, too, was quite accurate in both acquisition ($M = .85$, SD = .28) and extinction procedure ($M = .85$, SD = .28). We found weak evidence suggesting that memory for negative USs ($M = .87$, SD = .26) was better than for positive USs ($M = .83$, SD = .29, $BF_{10} = 3.46$) but there was no noteworthy evidence indicating that any other experimental manipulation affected US identity recognition, all $BF_{10} \leq 1.80$.

A.2.3.2 Exploratory analyses

As in the previous experiments, a one-way repeated-measures ANOVA of end-of-study pleasantness ratings of US categories indicated that participants remembered the valence of the US categories, $BF_{10} = 2.11 \times 10^{150}$. Participants remembered the animal category as more pleasant than the object category, $BF_{10} = 4.04 \times 10^{34}$, and human category as less pleasant than object category, $BF_{10} = 1.15 \times 10^{79}$. Thus, recognition memory for US categories may be indicative of participants' US valence memory.

Memory for CS-US pairings again was too accurate to test whether the observed differences in EC effects across referenced contexts was contingent on memory for CS-US pairs.

A.3 Normative IAPS ratings for USs

Table A.1: Identifiers of IAPS pictures used in CS-US and as filler USs with mean normative pleasure and arousal ratings (standard deviations in parentheses).

	CS-US pairs			
	Positive	Neutral	Negative	US-US pairs
	1610	7000	2750	9280
	1604	7035	2312	5970
	1620	7002	3300	5611
	1600	7009	2900.1	5250
	1750	7004	2276	5660
	1500	7233	2753	5870
	1460	7090	2110	5720
	1721	7080	9041	5780
	1540	7006	9331	9000
	1440	7175	2399	
	1463	7705	2100	
	1590	7025	2455	
Pleasure	7.56 (0.44)	4.97 (0.17)	3.12 (0.49)	5.61 (1.94)
Arousal	4.15 (0.57)	2.53 (0.40)	4.36 (0.28)	3.95 (0.74)

Note. IAPS = International Affective Picture System (Lang, Bradley, & Cuthbert, 2008), CS = Conditioned stimulus, US = Unconditioned stimulus

A.3.1 References

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-8). University of Florida, Gainesville, FL.

Table A.2: Weighted means (and standard deviations) of normative US pleasantness in Experiment 1.

Valence	Pleasure	Arousal
Acquisition		
Positive	7.55 (0.42)	4.23 (0.44)
Negative	2.81 (0.48)	4.54 (0.36)
Counterconditioning		
Positive	7.54 (0.42)	4.20 (0.45)
Negative	2.82 (0.44)	4.52 (0.37)

Note. Normative ratings from Lang et al. (2008).
US = Unconditioned stimulus

Appendix B

Appendix to Chapter 4

B.1 Gist and verbatim activation from inter-item similarities

In the following, we examine the global matching interpretation of gist and verbatim activation at the level of inter-item similarities. We assume that the study list is composed of sublists of related items and that any item can serve as memory probe, as is the case in study lists composed of exemplars from various categories (e.g., Buchanan, Brown, Cabeza, & Maitson, 1999; Andermane & Bowers, 2015; Pierce, Gallo, Weiss, & Schacter, 2005; Stark, Yassa, Lacy, & Stark, 2013). Further, we assume that $\gamma = 1$ to enable expansion of the summed probe-trace similarities f: The familiarity for a lure $x_i = L$ drawn randomly from the non-studied exemplars of a category expands to

$$\begin{aligned} f^{(L)} &= \sum_{j=1}^{J} \eta(L, y_j) \\ &= \underbrace{\sum_{p=1}^{P} \eta(L, r_p)}_{\text{partial matches}} + \sum_{q=1}^{Q} \eta(L, u_q) \end{aligned}$$

where r_p are related traces of items from the same sublist and u_q are unrelated traces of items from different sublists. The sum of similarities between the lure L and the related traces r_p represents the contribution of partial matches.

Similarly, the familiarity for a new unrelated distractor $x_i = N$ drawn randomly from the non-studied sublists expands to

$$f^{(N)} = \sum_{p=1}^{P} \eta(N, r_p) + \sum_{q=1}^{Q} \eta(N, u_q)$$

Assuming that items from different sublists exhibit no systematic similarity to items of the lure's sublist, it follows that

$$\sum_{q=1}^{Q} \eta(N, u_q) \approx \sum_{q=1}^{Q} \eta(L, u_q) \approx \sum_{q=1}^{Q} \eta(T, u_q)$$

where T is a target probe[1], the numerator for G in Equation (4.5) simplifies to

$$f^{(L)} - f^{(N)} \approx \sum_{p=1}^{P} \eta(L, r_p) - \eta(N, r_p)$$

We expand the summed similarity for a target T to an additional addend for exact matches,

$$\begin{aligned} f^{(T)} &= \sum_{p=1}^{P} \eta(T, r_p) + \sum_{q=1}^{Q} \eta(T, u_q) \\ &= \underbrace{\sum_{o=1}^{O} \eta(T, T^*)}_{\text{exact matches}} + \underbrace{\sum_{p'=1}^{P'} \eta(T, r'_{p'})}_{\text{partial matches}} + \sum_{q=1}^{Q} \eta(T, u_q) \end{aligned}$$

where T^* is the trace left by T during study list presentation and $r'_{p'}$ are other related traces of study list items from the same sublist. Given that sublist items can be randomly selected to serve as target or lure,

$$\sum_{p'=1}^{P'} \eta(T, r'_{p'}) \approx \sum_{p'=1}^{P'} \eta(L, r'_{p'})$$

the numerator for V in Equation (4.4) simplifies to

[1] The approximate equality of summed similarities between any sublist item and all studied items from other sublists implies a mnemonic interpretation of the b parameter in the CRM (pp. 169-170; Brainerd, Reyna, & Mojardin, 1999)

$$
\begin{aligned}
f^{(T)} - f^{(L)} &\approx \sum_{o=1}^{O} \eta(T,T^*) - \eta(L,T^*) + \sum_{p'=1}^{P'} \eta(T,r'_{p'}) - \eta(L,r'_{p'}) \\
&\approx O[\eta(T,T^*) - \eta(L,T^*)] \\
&\approx O[1 - \eta(L,T^*)]
\end{aligned}
$$

B.1.1 Effects of differentiation

Besides emphasizing the separable contributions of exact and partial matches to true and false recognition, the above expressions clarify the effect of differentiation. First, consider the effect of differentiation on the incremental contribution of exact matches (V). As implemented here, differentiation yields a steeper similarity gradient for repeatedly studied items, $\eta(x_i, y_i) > \eta(x_i, y_i)^{\delta} = \exp(-c\delta \cdot d(x_i, y_j))$, where $\delta > 1$. Knowing that 0 < $\eta(L,T^*)$ < 1, the following clearly shows that estimates of V increase as the differentiation factor δ increases,

$$
\begin{aligned}
f^{(T)} - f^{(L)} &\approx O[\eta(T,T^*)^{\delta} - \eta(L,T^*)^{\delta}] \\
&\approx O[1 - \eta(L,T^*)^{\delta}]
\end{aligned}
$$

Expansion of the simplified numerator of Equation (4.5), which we presented above, shows that differentiation also reduces estimates of G (to a much smaller extent b).

$$
f^{(L)} - f^{(N)} \approx O[\eta(L,T^*)^{\delta} - \eta(N,T^*)^{\delta}] + \sum_{p'=1}^{P'} \eta(L,r'_{p'}) - \eta(N,r'_{p'})
$$

Because by definition $\eta(L,T^*) > \eta(N,T^*)$, differentiation attenuates $\eta(L,T^*)$ more than $\eta(N,T^*)$.

While GCM requires differentiation to account for selective influence of repeated study on true recognition, it is important to note that GCM can produce two-dimensional response patterns without differentiation, Figure B.1.

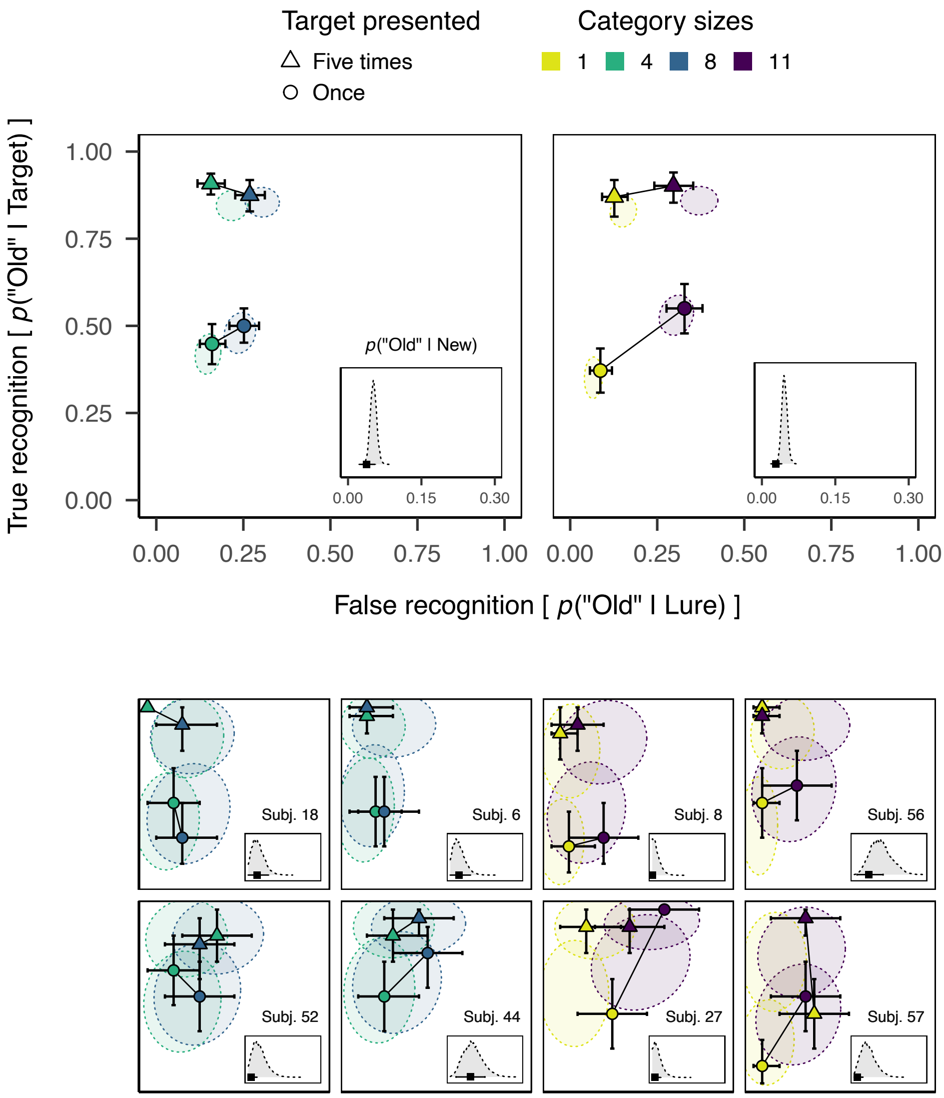

Figure B.1: State-trace plot of averaged (top) and individual responses (bottom) and posterior predictions of $\mathcal{M}^{\text{GCM}}_{\times\delta}$. From each group we show the two participants with strongest support for the non-monotonic (top row) and monotonic models (bottom row, see Figure 4.3). Points represent average observed rates of *old*-responses; error bars indicate 95% bootstrap confidence intervals based on 10,000 bootstrap samples. Ellipses represent multivariate normal-approximations to 95% credible regions posterior predictions. The inset shows the proportion of *old*-responses to unrelated new probes; kernel density estimates represent the posterior predictions.

B.1.2 References

Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of a week. *Psychonomic Bulletin & Review*, 1–6. https://doi.org/10.3758/s13423-015-0800-0

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*(1), 160–179. https://doi.org/10.1037/0033-295X.106.1.160

Buchanan, L., Brown, N. R., Cabeza, R., & Maitson, C. (1999). False Memories and Semantic Lexicon Arrangement. *Brain and Language*, *68*(1–2), 172–177. https://doi.org/10.1006/brln.1999.2072

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015). *Bayesian Data Analysis*. Boca Raton, FL: CRC press.

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A Multidimensional Scaling Database with Similarity Ratings for 240 Object Categories from the Massive Memory Picture Database. *PLoS ONE*, *9*(11), e112644. https://doi.org/10.1371/journal.pone.0112644

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*(5), 662–672. https://doi.org/10.3758/PBR.17.5.662

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Pierce, B. H., Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2005). The modality effect in false recognition: Evidence for test-based monitoring. *Memory & Cognition*, *33*(8), 1407–1413. https://doi.org/10.3758/BF03193373

Rosenthal, J. S. (2017). Simple confidence intervals for MCMC without CLTs. *Electronic Journal of Statistics*, *11*(1), 211–214. https://doi.org/10.1214/17-EJS1224

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

B.2 Cognitive modelling details

In the following we provide additional information on prior specification and model fits for CRM and GCM.

B.2.1 Conjoint-Recognition Model

Based on previous results and inspection of prior predictions we specified an informed prior on $\mu^{(b)}$ but used standard uninformative priors for the latent memory parameters for the gth group and hth condition,

$$\begin{aligned}\mu^{(b)} &\sim \mathcal{N}(-1.3, 0.5)\\ \mu^{(V)}_{gh} &\sim \mathcal{N}(0,1)\\ \mu^{(G)}_{gh} &\sim \mathcal{N}(0,1)\end{aligned}$$

For standard deviations and parameter correlations we used vaguely informative priors,

$$\begin{aligned}\sigma^{(\theta)}_{gh} &\sim \mathcal{G}(2,3)\\ \Omega_g &\sim \mathcal{LKJ}(2)\end{aligned}$$

B.2.1.1 Model fit

In the application of multinomial processing-tree models it is customary to test whether model predictions deviate from the observed category frequencies by means of G^2 tests. In the Bayesian hierarchical modeling framework model fit is often assessed visually and quantitatively (Chapter 6, Gelman et al., 2015). We visually compared posterior predictive distributions of each model to the observed data and computed summary T statistics to quantify model fit (Equations 17 and 18, Klauer, 2010). Posterior predictive checks, both visual and quantitative, rely on comparisons of observed data $\mathbf{n}^{obs}$ to new data generated from the model after it has been fit to the data. Thus, for each posterior sample of person-level parameters θ a new data set $\mathbf{n}^{pred}$ is generated randomly from the model. To quantify deviations of the data from model predictions a summary statistic $T(\mathbf{n},\theta)$ is computed for observed and predicted data. Model fit is assessed based on the estimated probability p that $T(\mathbf{n}^{pred},\theta) > T(\mathbf{n}^{obs},\theta)$. A small value of p indicates that the observed data deviate strongly from data that are plausible given the model.

We assessed the fit of our models with respect to three aspects of the data: participant-level frequency of old responses (T_P), mean frequency of old responses across participants (T_M), and covariance of frequency of old responses (T_{Cov}). For T_P and T_M we calculated expected frequencies of old responses for participant i in condition k as $\hat{n}_{ik} = N_k P(C_{ik}|\theta_i)$, where N_k is the number of observed responses. T_P was defined as

$$T_P(\mathbf{n},\theta) = \sum_{ik} \frac{(n_{ik} - \hat{n}_{ik})^2}{\hat{n}_{ik}},$$

where n_{ik} is the observed frequency of old responses by participant i in condition k. T_M was defined accordingly by averaging response frequencies across participants, i.e. $n_{.k} = \frac{1}{n}\sum_{i=1}^{n} n_{ik}$ and $\hat{n}_{.k} = \frac{1}{n}\sum_{i=1}^{n} \hat{n}_{ik}$. Similarly, deviations from the expected covariance of frequencies are quantified as

$$T_{Cov}(\mathbf{n},\theta) = \sum_{k}\sum_{l} \frac{(s_{k,l} - \hat{\sigma}_{k,l})^2}{\sqrt{\hat{\sigma}_{k,k}\hat{\sigma}_{l,l}}},$$

where s is the observed and $\hat{\sigma}$ the expected variance-covariance matrix. $\hat{\sigma}$ is calculated from expected frequencies $\hat{n}_{ik}$ and $\hat{n}_{il}$; variances on the diagonal are corrected[2] by addition of

$$\frac{\sum_i n p_k(1-p_k)}{n} \times \frac{(n-1)}{n}.$$

B.2.2 Generalized Context model

Based on previous results and inspection of prior predictions we specified vaguely informative priors on all unconstrained parameter means of the gth group,

[2]The term used by Klauer (2010) additionally corrects the expected covariances between response categories within each independent multinomial processing tree. Because we modeled the proportion of old responses as binomial outcome this correction was omitted.

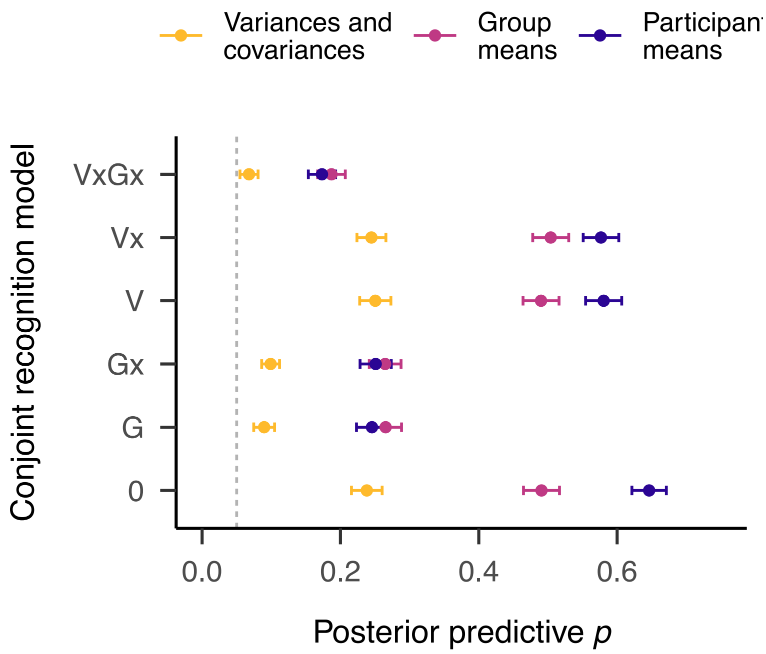

Figure B.2: Posterior predictive p-values for means at participant- and group-level as well as the variance-covariance matrix of participant means as predicted by our Conjoint Recognition Model variants. Error bars represent upper bound 95% Monte Carlo confidence intervals (Rosenthal, 2017). The vertical line indicates a posterior predictive p-value of 0.05. All models describe the data adequately.

$$\log(\mu_g^{(c)} + 0.5) \sim \mathcal{N}(0.1, 0.5)$$
$$\Phi^{-1}(1/\mu_g^{(\delta)}) \sim \mathcal{N}(-0.26, 0.23)$$
$$\log(\mu_g^{(m')}) \sim \mathcal{N}(1, 1)$$
$$\log(\mu_g^{(k)} + 0.33) \sim \mathcal{N}(1, 0.67)$$
$$\log(\mu_g^{(\gamma)} + 0.5) \sim \mathcal{N}(0.1, 0.5)$$
$$\log(d_{\text{inter}} + 3) \sim \mathcal{N}(1, 0.67)$$

For standard deviations and parameter correlations we again used vaguely informative priors,

$$\sigma_g^{(c)} \sim \mathcal{G}(2, 5)$$
$$\sigma_g^{(\delta)} \sim \mathcal{G}(2, 3)$$
$$\sigma_{m'}^{(\delta)} \sim \mathcal{G}(2, 3)$$
$$\sigma_g^{(k)} \sim \mathcal{G}(2, 3)$$
$$\sigma_g^{(\gamma)} \sim \mathcal{G}(2, 5)$$
$$\Omega_g \sim \mathcal{LKJ}(2)$$

B.2.2.1 Inter-category distance estimate

Without distance estimates for stimuli from different categories the GCM cannot predict old-responses to new distractors, i.e., exemplars from new categories. We, therefore, estimated an auxiliary parameter d_{inter} as a stand-in for any $d(x_i, y_j)$ where i and j were from different categories. The obtained estimate of $d_{\text{inter}} = 8.08$ 95% HDI [7.44, 8.81] is in a psychologically plausible range. Figure B.3 shows the posterior distribution in comparison to the empirical distribution of within-category distances obtained from MDS of similarity ratings from Hout, Goldinger, & Brady (2014). The inter-category distance is estimated to be considerably larger than all within-category distances. The inset of Figure B.3 shows a normal quantile-quantile plot comparing the distribution of within-category distances to a normal distribution. In line with the results reported by Johns & Jones (2010) for lexical semantics, the distribution is substantially skewed towards fewer high-similarity word pairs.

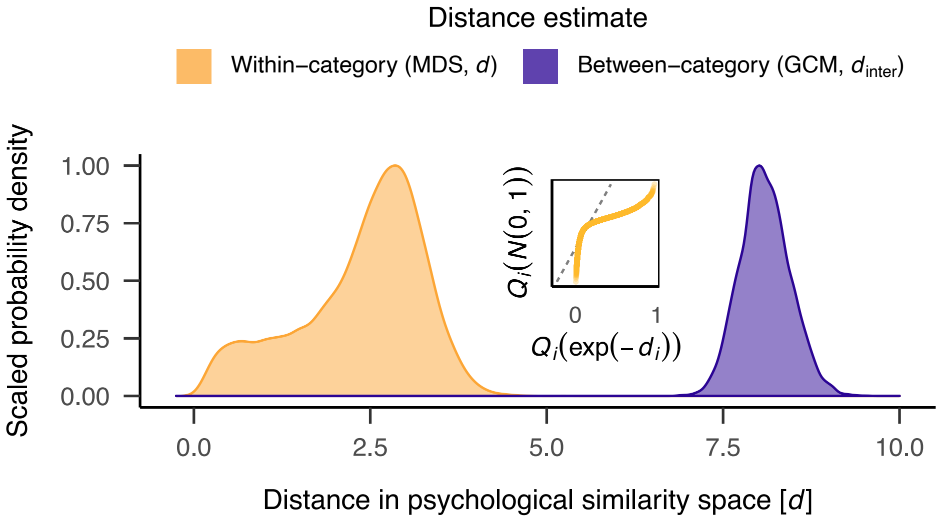

Figure B.3: Estimates of distances between study list items in psychological similarity space. The yellow distribution represents the the kernel density estimate of all pairwise within-category distances obtained from the multi-dimensional scaling (MDS) of similarity ratings from Hout et al. (2014). The solid purple distribution represents the posterior distribution for the auxiliary parameter d_{inter}, a stand-in for any distance $d(x_i, y_j)$ where x_i and y_j are items from different categories, in the fitted Generalized Context Models (GCM). The inset shows a normal quantile-quantile plot comparing the distribution of within-category distances to a normal distribution.

B.2.3 References

Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of a week. *Psychonomic Bulletin & Review*, 1–6. https://doi.org/10.3758/s13423-015-0800-0

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*(1), 160–179. https://doi.org/10.1037/0033-295X.106.1.160

Buchanan, L., Brown, N. R., Cabeza, R., & Maitson, C. (1999). False Memories and Semantic Lexicon Arrangement. *Brain and Language*, *68*(1–2), 172–177. https://doi.org/10.1006/brln.1999.2072

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015). *Bayesian Data Analysis*. Boca Raton, FL: CRC press.

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A Multidimensional Scaling Database with Similarity Ratings for 240 Object Categories from the Massive Memory Picture Database. *PLoS ONE*, *9*(11), e112644. https://doi.org/10.1371/journal.pone.0112644

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*(5), 662–672. https://doi.org/10.3758/PBR.17.5.662

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Pierce, B. H., Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2005). The modality effect in false recognition: Evidence for test-based monitoring. *Memory & Cognition*, *33*(8), 1407–1413. https://doi.org/10.3758/BF03193373

Rosenthal, J. S. (2017). Simple confidence intervals for MCMC without CLTs. *Electronic Journal of Statistics*, *11*(1), 211–214. https://doi.org/10.1214/17-EJS1224

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Appendix C

Appendix to Chapter 5

C.1 Additional results

Table C.1: Analysis of variance results for adjusted "Old"-response rates.

Effect	F	df_1^{GG}	df_2^{GG}	MSE	p	$\hat{\eta}^2_G$	BF_{10}
Material	189.33	1.99	159.22	0.09	$< .001$	.269	4.52×10^{143}
Target presentations	71.63	1	80	0.06	$< .001$	.044	7.74×10^{19}
Category size	266.64	1	80	0.03	$< .001$	.085	3.69×10^{40}
Probe type	768.24	1	80	0.11	$< .001$	.471	4.93×10^{290}
Material × Target presentations	2.17	1.96	156.94	0.05	.118	.003	0.16
Material × Category size	4.53	1.97	157.82	0.03	.013	.003	0.21
Target presentations × Category size	2.29	1	80	0.03	.135	.001	0.17
Material × Probe type	49.17	1.82	145.45	0.05	$< .001$	.048	7.03×10^{20}
Target presentations × Probe type	24.00	1	80	0.03	$< .001$	.007	179.93
Category size × Probe type	82.47	1	80	0.04	$< .001$	.031	3.14×10^{13}
Material × Target presentations × Category size	1.19	1.99	159.16	0.02	.308	.001	0.05
Material × Target presentations × Probe type	2.21	1.96	156.86	0.03	.114	.001	0.10
Material × Category size × Probe type	9.18	1.78	142.80	0.03	$< .001$	.006	9.89
Target presentations × Category size × Probe type	0.22	1	80	0.03	.641	.000	0.11
Material × Target presentations × Category size × Probe type	1.10	1.98	158.26	0.02	.336	.001	0.06

Note. The rate of false recognition was adjusted by subtracting "Old"-response rates to unrelated new probes from those to lures. For the Bayesian analysis we used a scale of $r = \sqrt{2}/2$ for the prior distribution.

Table C.1 summarizes the ANOVA results of the adjusted "Old"-response rates in our experiment.

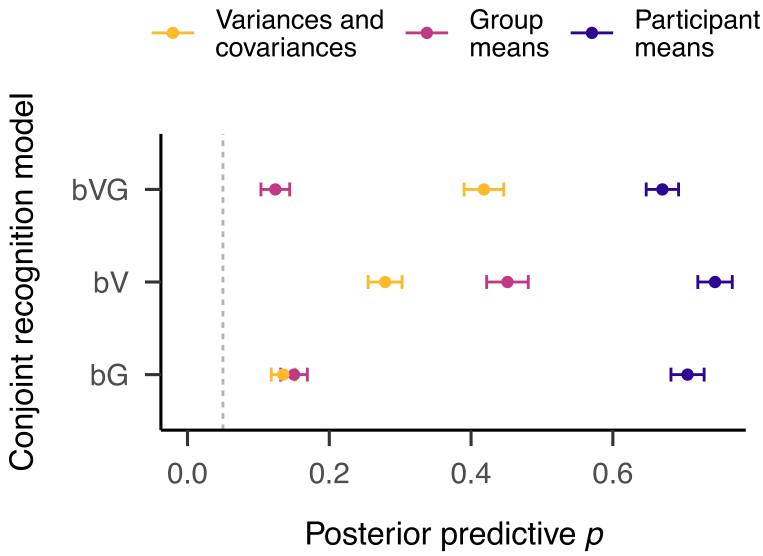

Figure C.1: Posterior predictive *p*-values for means at participant- and group-level as well as the variance-covariance matrix of participant means as predicted by our Conjoint Recognition Model variants. Error bars represent upper bound 95% Monte Carlo confidence intervals (Rosenthal, 2017). The vertical line indicates a posterior predictive *p*-value of 0.05. All models describe the data adequately. The increase in fit for the variance-covariance structure is likely due to an overly informative prior setting that needs to be explored further.

When fitting the CRM to our data, we assessed the fit of each model with respect to three aspects of the data: participant-level frequency of old responses (T_P), mean frequency of old responses across participants (T_M), and covariance of frequency of old responses (T_{Cov}; for details see Chapter 4). As shown in Figure C.1, all models describe the data adequately.

C.1.1 References

Rosenthal, J. S. (2017). Simple confidence intervals for MCMC without CLTs. *Electronic Journal of Statistics*, *11*(1), 211–214. https://doi.org/10.1214/17-EJS1224

Appendix D

Individual contributions

The following list describes my and each of my coauthors contributions to the empirical research chapters. I list roles according to the standardized Contributor Roles Taxonomy (CRediT; Allen et al., 2014; Holcombe et al., 2020).

A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning

Frederik Aust: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing

Julia M. Haaf: Conceptualization, Investigation, Methodology, Resources, Software, Validation, Writing - Original Draft Preparation, Writing - Review & Editing

Stahl Christoph: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing - Original Draft Preparation, Writing - Review & Editing

An exemplar-familiarity interpretation of verbatim and gist memory in false recognition

Frederik Aust: Conceptualization, Data Curation, Formal Analysis, Methodology, Validation, Visualization, Writing - Original Draft Preparation

Roscoe F. J. W. Araujo: Conceptualization, Investigation, Methodology, Project Administration, Resources, Software

Christoph Stahl: Conceptualization, Funding Acquisition, Investigation, Methodology, Resources, Software, Supervision

An exemplar-familiarity model of false recognition over the short term

Frederik Aust: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing - Original Draft Preparation

Christoph Stahl: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision