Cognitive Psychology 104 (2018) 106–142

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Modeling the dynamics of recognition memory testing with an integrated model of retrieval and decision making

Adam F. Osth[a,*], Anna Jansson[b], Simon Dennis[a], Andrew Heathcote[c]

[a] University of Melbourne, Australia
[b] University of Newcastle, Australia
[c] University of Tasmania, Australia

ARTICLE INFO

Keywords:
Recognition memory
Global matching models
Diffusion decision model
Testing effects

ABSTRACT

A robust finding in recognition memory is that performance declines monotonically across test trials. Despite the prevalence of this decline, there is a lack of consensus on the mechanism responsible. Three hypotheses have been put forward: (1) interference is caused by learning of test items (2) the test items cause a shift in the context representation used to cue memory and (3) participants change their speed-accuracy thresholds through the course of testing. We implemented all three possibilities in a combined model of recognition memory and decision making, which inherits the memory retrieval elements of the Osth and Dennis (2015) model and uses the diffusion decision model (DDM: Ratcliff, 1978) to generate choice and response times. We applied the model to four datasets that represent three challenges, the findings that: (1) the number of test items plays a larger role in determining performance than the number of studied items, (2) performance decreases less for strong items than weak items in pure lists but not in mixed lists, and (3) lexical decision trials interspersed between recognition test trials do not increase the rate at which performance declines. Analysis of the model's parameter estimates suggests that item interference plays a weak role in explaining the effects of recognition testing, while context drift plays a very large role. These results are consistent with prior work showing a weak role for item noise in recognition memory and that retrieval is a strong cause of context change in episodic memory.

A major constraint on models of memory concerns how the number of items present in memory affects memory performance. Such manipulations of memory set size have constrained models of recognition memory at both short (McElree & Dosher, 1989; Nosofsky, Little, Donkin, & Fific, 2011; Sternberg, 1966) and long (Clark & Gronlund, 1996; Dennis & Humphreys, 2001; Gillund & Shiffrin, 1984; McClelland & Chappell, 1998; Osth & Dennis, 2015; Shiffrin & Steyvers, 1997) time scales. Much theoretical interest concerns how the number of *studied* items in memory affects performance. However, of recent focus in recognition memory research is how the number of *tested* items affects memory.

Almost universally, recognition memory performance decreases throughout the course of testing. This finding was first reported by Peixotto (1947), but has frequently been replicated in the decades since (Annis, Malmberg, Criss, & Shiffrin, 2013; Averell, Prince, & Heathcote, 2016; Criss, Malmberg, & Shiffrin, 2011; Kiliç, Criss, Malmberg, & Shiffrin, 2017; Malmberg, Criss, Gangwani, & Shiffrin, 2012; Murdock & Anderson, 1975; Schulman, 1974). However, despite its status as an empirical regularity, theoretical interest in the nature of the testing effect has emerged only more recently (e.g.; Criss et al., 2011; Osth & Dennis, 2015). The decrease has been referred to as "output interference" in the literature, but we will reference it as the *test position effect*, or TPE, to avoid

* Corresponding author.
E-mail address: adamosth@gmail.com (A.F. Osth).

https://doi.org/10.1016/j.cogpsych.2018.04.002
Accepted 26 April 2018

Table 1
Key terms and definitions used throughout the article.

Term	Definition
Test position effect (TPE)	Finding that recognition memory performance declines across test trials
List length effect	Finding that recognition memory performance is worse for longer study lists
Mirror effect	Effect of a manipulation that has opposite effects on hit rates (HRs) and false alarm rates (FAR). Low frequency (LF) words, for instance, have higher HRs and lower FARs than high frequency (HF) words (Glanzer & Adams, 1985)
Global matching model	Process model framework for recognition memory. Probe item is matched against all memories simultaneously; each similarity value is summed together to produce an index of memory strength that is used to make a decision
Context representation	Representation that defines a learning episode
Item noise	Interference generated by the studied items that are not the probe item (match in context, mismatch on item)
Context noise	Interference generated by occurrences of the probe item that were acquired prior to the study episode (match in item, mismatch in context)
Background noise	Interference generated by memories of items that are not on the study list (mismatch on item and context)
Context drift	Change in the context representation in response to events, such as study or test trials

commitment to the idea that the effect is driven by interference from items learned at test. In addition, we will demonstrate later in this article that changes in decision dynamics through the course of testing play a role in the observed phenomenon. Modeling of the TPE has explored causal factors acting through the decision process or memory retrieval, but not both. Our work aims to bridge this gap by introducing a combined model of memory retrieval and decision making that addresses the changes in both choice probabilities and response time (RT) distributions through the course of testing. Due to the large amount of nomenclature in the paper, a list of key terms can be found in Table 1.

1. Causes of the test position effect

One of the earliest attempts to explore the nature of the TPE was through Ratcliff (1978)'s application of the Diffusion Decision Model (DDM), an evidence accumulation model of the decision process (see Fig. 1). In the DDM, evidence begins at a starting point z and accumulates in a noisy fashion toward one of two response boundaries; an upper response boundary denoted by the parameter a (corresponding to a "YES" decision in recognition memory) and a lower boundary at zero (corresponding to a "NO" decision). The boundary first reached determines the choice made by the participant, while the time taken to reach the boundary, when added to the time for non-decision processes, determines the response time (RT). The phenomenon of the speed-accuracy tradeoff, whereby faster decisions are made less accurately, is captured by changes in the a parameter; increases in the boundary make errors less likely but increase the RT due to the longer distance that the process has to travel in order to reach a boundary. Memory strength in the model is conceptualized as the rate of evidence accumulation, or the *drift rate*; increases in the drift rate increase the proportion of correct responses and decrease the RT. Drift rate is not fixed but varies from trial-to-trial, which is analogous to cross-trial variability in memory strength in signal detection theory (SDT). Finally, non-decision time components, such as perceptual processing and response output, are modeled by parameter t_{ER}. Variability in non-decision time is assumed to have a uniform distribution with width s_t.

One might naively assume that changes in performance are due to changes in memory strength alone. However, according to the DDM, changes in performance across conditions can also be due to participants setting different speed-accuracy thresholds. Drift rate and speed-accuracy thresholds can be separably estimated in the DDM due to their differential effects on the RT distribution – increases in drift rate primarily decrease the skew of the RT distribution, while increases in response boundaries increase both the

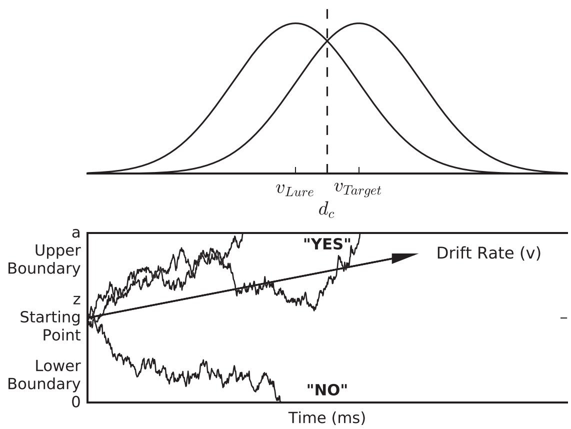

Fig. 1. The diffusion decision model (DDM). The drift rate is a sample from one of the normal distributions in the above panel. Evidence accumulation is noisy, such that diffusion processes with the same drift rate samples can reach different boundaries and produce different RTs. Depicted are three sample trajectories with the same drift rate. See the main text for more details.