Journal of Memory and Language 88 (2016) 51–69

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Fundamental causes of systematic and random variability in recognition memory

Lee Averell[a], Melissa Prince[b], Andrew Heathcote[c,*]

[a] The School of Psychology, The University of Newcastle, Australia
[b] School of Psychology, Flinders University, Australia
[c] School of Medicine, University of Tasmania, Australia

ARTICLE INFO

Article history:
Received 17 December 2012
revision received 23 December 2015
Available online 13 January 2016

Keywords:
Recognition memory
Item-context facilitation
Study-test lag
Test interference
Bayesian random item effects
Receiver Operating Characteristic

ABSTRACT

Progress in understanding recognition memory has been hampered by confounding among effects associated with the study position, test position and study-test lag factors that are intrinsic to the widely used study-test list paradigm. For example, the list-length effect – once considered a robust benchmark phenomenon – is now known to be either weak or absent when confounding effects associated with these factors are controlled. We investigate two effects of recent theoretical interest – item-context facilitation (occurring when items studied together are tested together) and test-position interference (with performance decreasing over a sequence of test trials) – and one effect of long-standing interest – decreasing performance as study-test lag increases. Traditional analyses of our experiment, whose design affords control over a range of confounds and allows us to disentangle the three effects, affirms all three as fundamental causes of systematic variability in recognition accuracy. These conclusions are strengthened and expanded by model-based analyses of recognition confidence and random item effects that also take into account non-systematic sources of variability.

Introduction

In an episodic recognition memory experiment, participants are asked to discriminate between items that were encountered (old) or not (new) in a specific episode (typically the previous study list). The simplicity of such study-test list paradigms has made them a cornerstone of the memory literature, and has revealed one of the most salient characteristics of human memory – variable and error-prone performance. Memory failures have been attributed to a number of systematic causes including problems at encoding (e.g., not attending to the name of someone when they introduced to you), during storage (e.g., failed consolidation; Wixted, 2004) or at retrieval (e.g., Mandler's, 1980, anecdote about failing to recognize his butcher in the unfamiliar context of a bus). Perhaps the most time-honoured cause of variability is associated with the delay between study and test (i.e., lag; Ebbinghaus, 1885). However, substantial unsystematic variations among people and among the items to be remembered are also ubiquitous and there has been a growing realization of the importance of formally treating them as random effects in memory analyses (e.g., Rouder & Lu, 2005). It is not surprising then, that much of the memory literature has been driven by the aim of understanding and identifying both the systematic and random factors that cause variability in memory performance.

Unfortunately, it is also a consequence of the appealing simplicity of study-test list paradigms that progress in identifying the systematic causes of variable performance

* Corresponding author at: Division of Psychology, Social Sciences Building, Sandy Bay, 7005 Tasmania, Australia.
E-mail address: andrew.heathcote@utas.edu.au (A. Heathcote).

http://dx.doi.org/10.1016/j.jml.2015.12.010

has been impeded because the factors of interest are intrinsically confounded with each other. Such confounding co-variation makes it difficult to isolate the relative contribution of each effect by means of experimental controls. Moreover, traditional methods of analysis can have problems associated with bias and efficacy (e.g., being overly optimistic about the significance of results due to ignoring random item effects) when this co-variation is ignored.

Here we aim to disentangle some of the primary systematic causes of variable performance in an item recognition memory task. In our experiment, participants and items as well as study and test conditions were all relatively homogenous and in line with procedures used in much past research. Because our overall experimental design follows the standard procedures implemented in the literature, the potential systematic causes of variable performance that we aim to disentangle are fundamental in the sense that they are present in the majority of study-test paradigms. We disentangle the systematic and random causes through a combination of two strategies.

First, we manipulated (between-subjects) the order in which items were tested. This enables a range of confounds to be controlled and allows us to disentangle the contribution of three ostensible systematic effects associated with:

1. *Study-test lag* – a decrease in accuracy with increasing lag between study and test;
2. *Test position* – a decrease in accuracy associated with increasing position in the test list;
3. *Item context* – an increase in accuracy when study and test item contexts match.

Our initial analyses focus on the effects of these factors on the recognition rates (and monotonic transformations thereof), the measures that have been traditionally used to examine performance in a largely theory-free way.

Second, we provide a rigorous re-evaluation of these fundamental effects using Pratte and Rouder's (2011) model-based hierarchical Bayesian methods for the analysis of confidence ratings about new and old choices. This approach has the advantage of modelling the complete set of dependent variables collected on test trials (i.e., both old vs. new recognition decisions and associated confidence ratings), while also simultaneously accounting for both participant and item variability. It also allows us to examine the causes of observed performance in terms of the latent processes assumed by the unequal-variance signal-detection model (UVSD; Heathcote, 2003; Wixted, 2007). We also performed analyses based on dual-process signal-detection (DPSD; Yonelinas, 1994) models of recognition memory. Because the DPSD model did not provide as good an account of our data, and in any case supported conclusions very similar to the UVSD analysis, we provide this analysis in Supplementary materials.

In the following sections, we begin by reviewing previous evidence for these causes of variable performance. We then introduce our paradigm and explain how we are able to tease apart typically confounded effects. Finally, we report the modelling results that allow us to separate memory effects from decision effects, such as response bias. This de-confounding requires a theory of the underlying psychological processes, and although our primary focus is on memory effects it is important to also take account of how participants modulate bias in their decision processes in order to understand how effects on directly observed recognition rates are explained.

Variability associated with study

As well as three systematic causes (lag, test position and context) and two random causes (participants and items) we also study a third random cause only associated with studied items. Episodic recognition memory studies indicate that studied items not only have greater memory strength than non-studied items, but also are associated with greater variability in strength. This result is intuitively plausible given it is unlikely that the degree of strength added to an item during study is exactly the same for every item – participants may have fluctuations in attention or motivation over the course of a study list, they may have experienced an item outside of the experimental setting, and the time between study and test presentations varies, likely making some items easier to remember than others. This difference in new and old item variability is reflected by a ubiquitous Receiver Operating Characteristic (ROC) asymmetry, which was first demonstrated in the late 1950s and has since been replicated numerous times and with many procedural variations (e.g., Glanzer, Kim, Hilford, & Adams, 1999; Gronlund & Elam, 1994; Heathcote, 2003; Mickes, Wixted, & Wais, 2007; Ratcliff, Sheu, & Gronlund, 1992; Wixted & Stretch, 2004; Yonelinas, 1994).

ROCs plot the probability of correctly identifying a test item as having been studied (i.e., hit rate; HR) against the probability of falsely identifying an item as having been studied when it was in fact new (i.e., false alarm rate; FAR), across a range of decision criteria. These decision criteria can be varied by manipulating base rates or pay-offs in the experiment, but a more efficient and common approach is to ask participants to rate their confidence for binary recognition decisions.[1] Yonelinas and Parks (2007) note that it is a direct consequence of early ROC studies that we see a dominance of memory theories framed in terms of signal-detection theory.

The prototypical version of signal detection theory (Green & Swets, 1966) represents old and new items as equal-variance Gaussian distributions located on a single 'memory-strength' dimension; that is, the equal-variance signal-detection (EVSD) model. Items are classified as new or old with varying degree of confidence depending on where their memory-strength falls relative to a set of fixed criteria (e.g., increasing criteria demarcating high-confidence new, medium-confidence new and so on, up to high-confidence old ratings). However, EVSD predicts symmetric ROCs. Consequently, the equal variance

[1] Although there has been much recent debate concerning the use of ratings ROCs (e.g., Bröder & Schütz, 2009; Dubé & Rotello, 2012), Dube and Rotello found ROCs were not distorted by the ratings procedure and recommend the use of confidence ratings in analysing recognition memory data.