Journal of Memory and Language 93 (2017) 55–66

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Eye movements in forced-choice recognition: Absolute judgments can preclude relative judgments

Jeffrey J. Starns *, Tina Chen, Adrian Staub

University of Massachusetts Amherst, United States

ARTICLE INFO

Article history:
Received 8 December 2015
revision received 30 August 2016

Keywords:
Forced-choice testing
Recognition memory
Eye tracking

ABSTRACT

Forced choice recognition is usually assumed to involve a relative judgment process in which each test alternative is matched to memory and the one with the highest memory strength is selected. We monitored eye movements during a forced-choice recognition test to determine if absolute judgments also play a role; that is, do participants ever select an item because its memory strength exceeds an absolute criterion without comparing it to the other item? The results strongly supported a role for absolute judgments. Participants sometimes selected a response without looking at one of the test alternatives, and they were most likely to do this when they looked at the target first and correctly selected it as the studied item. Participants were also more accurate when they looked at the target first than when they looked at the lure first, which would be expected if they sometimes failed to consider the actual target word because they made an incorrect absolute judgment that the lure was studied. Finally, response times were faster when the word selected as the studied item was considered first than when it was considered second; that is, correct responses were faster when the target was viewed first and errors were faster when the lure was viewed first. This would be expected if participants sometimes make absolute judgments that the first item was studied, thus eliminating the additional time needed to consider the second item.

Introduction

Life often confronts us with tests of recognition memory; that is, situations in which we must decide whether or not a stimulus was previously encountered in a certain context. For example, someone who is not sure which baggage claim she should go to can look at the people waiting at each possible location and decide whether or not she saw them on the flight. Understanding recognition memory is an interesting goal in its own right, and recognition tasks have played an important role in testing general theories of event memory. One recurring theme in theoretical debates is whether the effects of specific variables are driven by changes in memory, changes in decision processes, or both (e.g., Criss, 2006; Glanzer & Bowles, 1976; Hicks & Marsh, 1998; Starns, White, & Ratcliff, 2010). Other common theoretical disputes concern whether recognition judgments are based on independent memory systems or a single evidence strength signal (e.g., Wixted, 2007; Yonelinas & Parks, 2007), and whether information retrieved from memory is discrete or can vary along a full continuum (e.g., Bröder & Schütz, 2009; Pazzaglia, Dube, & Rotello, 2013).

All of these debates have been informed by experiments using a forced-choice testing procedure. In the typical forced-choice procedure, participants study a list of items (usually words) and then complete a test in which two

* Corresponding author at: Department of Psychological and Brain Sciences, University of Massachusetts, 135 Hicks Way, Amherst, MA 01003-9271, United States.
E-mail address: jstarns@psych.umass.edu (J.J. Starns).

http://dx.doi.org/10.1016/j.jml.2016.09.001
0749-596X/

items are displayed on each test trial, one that was studied on the earlier list (the "target") and one that was not (the "lure"). Participants attempt to select the item that they recognize from the study list. This procedure can be contrasted with the old/new test format, in which the participant considers one item at a time and indicates whether or not it appeared on the study list.

Many theorists have assumed that forced-choice responding relies on relative judgments. According to this account, participants compare their memories for the two items and select the item with the strongest evidence of being studied on the list or guess if the items have identical evidence states (e.g., Green & Moses, 1966; Tanner & Swets, 1954). Our goal is to determine if forced choice responding indeed relies solely on relative judgments, or if absolute judgments also play a role. For example, participants might select one of the test alternatives because its absolute evidence exceeds an internal response criterion regardless of the evidence value for the other item. We used eye movement monitoring to track how participants consider each test alternative before making a selection. In the following sections, we first describe some important applications of forced-choice testing and show how interpretations in these contexts rely on the relative-judgment assumption, and then we explain how eye movement data can potentially help identify the influence of relative and absolute judgments.

Applications of forced-choice testing

One common use for forced-choice testing has been to control for potential differences in response bias between conditions (e.g., Glanzer & Bowles, 1976; Hicks & Marsh, 1998). The idea is that participants do not need to set a criterion for the amount of memory evidence that they require to recognize an item, because they can select an item simply by comparing the relative evidence strength of the two options. Thus, any difference between conditions on a forced-choice test must demonstrate a difference in memory evidence and not a difference in the criterion used to evaluate the evidence. For example, words that were not on the study list are less likely to be falsely recognized if they are low-frequency words than if they are high-frequency words (Glanzer & Adams, 1985). This finding could mean that high-frequency words produce higher levels of illusory memory evidence than low-frequency words, or it could mean that people apply a more stringent retrieval criterion for low frequency words because they expect to have better memory for these items. Glanzer and Bowles attempted to discriminate these possibilities with a forced-choice test in which they included some trials with two lures – one high frequency and one low frequency – instead of a lure and a target (without informing participants that some trials would not include a target item). Participants consistently chose the high-frequency lure in these pairs, and the authors interpreted this as demonstrating that high-frequency lures produced higher levels of (illusory) memory evidence than low-frequency lures.

Researchers have also used forced-choice testing to explore the nature of memory retrieval in recognition (Jang, Wixted, & Huber, 2009; Kellen & Klauer, 2011; Kellen, Singmann, Vogt, & Klauer, 2015; Kroll, Yonelinas, Dobbins, & Frederick, 2002; Parks & Yonelinas, 2009; Province & Rouder, 2012; Smith & Duncan, 2004). One goal in this literature is to distinguish models that have continuous retrieval states, discrete retrieval states, or a mixture of continuous and discrete retrieval states from independent memory systems. Several studies have compared these models in terms of their ability to explain the relationship between accuracy on old/new and forced-choice recognition tests (Jang et al., 2009; Kroll et al., 2002; Smith & Duncan, 2004). Other studies used tests with more than two options to evaluate each model's predictions for how often participants can select the target item on their second try after they initially choose a lure (Kellen & Klauer, 2011; Parks & Yonelinas, 2009). Forced-choice data have also been useful for testing models that specify how information is stored in and retrieved from memory, including global matching models (e.g., Clark, Hori, & Callan, 1993) and neurally-inspired models (e.g., Norman & O'Reilly, 2003). Forced-choice testing has even helped to define the nature of hippocampal amnesia (e.g., Bayley, Wixted, Hopkins, & Squire, 2008; Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000).

In all of the applications listed above, conclusions about memory rely on the relative-evidence assumption. If absolute judgments are common in forced-choice testing, then these conclusions should be re-evaluated. For example, absolute judgments require a response criterion, so simply demonstrating an effect on a forced-choice test does not necessarily mean that the effect is based on memory differences as opposed to changes in decision processes. Imagine that participants use a simple strategy that combines absolute and relative judgments: they select the first item that passes an absolute criterion for being a studied item, or they choose the stronger item if neither passes the criterion. For lure-lure trials in the Glanzer and Bowles (1976) experiment, this strategy would lead to more selections for the high-frequency lure even if memory strength is equivalent for the two frequency classes as long as people use a higher absolute criterion for low-frequency words, because high-frequency lures would have a better chance of triggering a response based on an absolute judgment. Thus, a forced-choice test does not necessarily help to discriminate memory and decision processes if absolute judgments play a role. Similarly, all of the models relating old-new and forced-choice accuracy assume that relative judgments are made for every trial (Jang et al., 2009; Kellen & Klauer, 2011; Kellen et al., 2015; Kroll et al., 2002; Parks & Yonelinas, 2009; Province & Rouder, 2012; Smith & Duncan, 2004), and the predicted relationship between the two tasks changes when absolute judgments are allowed (we explore this in more detail in the General Discussion).

Although the relative-evidence assumption underlies a range of theoretical conclusions, we know of no strong evidence that rules out the possibility that absolute judgments also play a role. In fact, Hockley (1984) reported results that might indicate a role for absolute judgments. Hockley evaluated response time (RT) data in a forced choice recognition task, and he found that RTs for correct