



The effects of weak versus strong relational judgments on response bias in Two-Alternative-Forced-Choice recognition: Is the test criterion-free?☆



Jerwen Jou^{a,*}, Shaney Flores^b, Hector M. Cortes^a, Bryce G. Leka^a

^a University of Texas – Rio Grande Valley, USA

^b Washington University in St. Louis, USA

ARTICLE INFO

Article history:

Received 22 November 2015

Received in revised form 26 March 2016

Accepted 30 March 2016

Available online 19 April 2016

Keywords:

Two-alternative-forced-choice

Criterion in recognition memory

2AFC recognition

Null test items in 2AFC

Old/new 2AFC versus comparative judgment

ABSTRACT

It is widely believed that a Two-Alternative-Forced-Choice (2AFC) in an old/new recognition memory test is made by comparing the two items and choosing the item with the higher strength. For this reason, it is considered to be criterion-free by some researchers. We found evidence that subjects probabilistically compromised the comparison by choosing the left item when they recognized it as old. Using both normal test pairs (comprised of one new and one old item) and two types of null pairs (comprised of both-new or both-old items), we found that a left-biased choice was coupled with higher hit and false alarm rates and a shorter left than right-choice RT for the normal pairs, consistent with the hypothesis of a bias for making a choice on the basis of a left individual-item recognition. For the null pairs, RT was much longer for the both-new than for the both-old pairs, providing additional evidence for basing decision on an individual-item's absolute, rather than a relative, familiarity. Additionally, subjects gave higher confidence ratings to choices for the both-old than the normal and both-new pairs, again suggesting that their decision was based on absolute familiarity of the items. The results were found to be not due to a fast-response instruction. A comparative judgment experiment in which subjects chose the item higher or lower in an attribute magnitude did not show the response side bias and RT asymmetry. The presence of bias in the former, and the absence of it in the latter can be explained by a weak versus strong relational judgment in the former and the latter type of 2AFC, respectively. We discuss the implications these findings have for the use of the 2AFC as a method for testing recognition memory.

© 2016 Elsevier B.V. All rights reserved.

Two most commonly used methods of testing recognition memory are the Yes/No (YN henceforth) recognition and the Two-Alternative-Forced-Choice (2AFC) test. In a YN test, a single test item is presented and subjects make either an “Old” or a “New” decision for the single test item. According to Signal Detection Theory (Egan, 1975; Green & Swets, 1966; Macmillan & Creelman, 1991; Swets, Tanner, & Birdsall, 1961; Tanner & Swets, 1954), there are two distributions along a continuum of memory strength (or familiarity) underlying this decision process, an old-item and a new-item distribution. On average, the old-item distribution has a higher memory strength than the new-item distribution. There are two factors in the signal detection process, the d' (or sensitivity measured as the z -score difference between the mean strength of the old-item and that of the new-item distribution), which

determines discriminability, and the *criterion*, which determines a bias towards responding Yes or No. The bias is affected by motivation and is supposed to be independent of the true discriminability measure d' . The criterion is a threshold placed somewhere between the two distributions, dividing the space into two categories. If a test item's familiarity falls above the criterion, a “Yes” response will be made. If it falls below it, a “No” response will be made. One can make more Yes's or No's responses depending on one's cost and benefit analysis of the consequence of the decision. But the change in bias, i.e., the shifting in the placement of criterion will not change one's true discriminating ability (d')¹ (Egan, 1975; Green & Swets, 1966; Macmillan & Creelman, 1991).

In a 2AFC test, one must choose only one of the test items as “Old” regardless of whether any or both of the two items are judged to have met the criterion. Thus, the criterion is thought to be irrelevant, meaning that one does not compare the test items against a familiarity

☆ Experiment 1A was presented at the 51st Annual Meeting of the Psychonomic Society in St. Louis, MO. We thank Marco Gonzalez, Mario Arredondo, Eric Escamilla, and Liann Pena for help with data collection. Jim Neely did a very thorough review and gave us detailed feedback, for which we are very grateful. We also thank an anonymous reviewer for his/her comments.

* Corresponding author at: Department of Psychological Science, University of Texas – Rio Grande Valley, 1201 W. University Drive, Edinburg, TX 78539-2999, USA.

E-mail address: jerwen.jou@utrgv.edu (J. Jou).

¹ Although it is commonly assumed that d' is a discrimination measure independent of bias or criterion, it is in fact not the case when the variance of the old-item distribution is not equal to that of the new items (with the variance of the old-item distribution typically larger than that of the new-item distribution). Data showed that in YN recognition, stricter criteria yield larger d' than lenient criteria (Cho & Neely, 2012; Jou, 2011; Kroll et al., 2002).

criterion one sets up as in a YN test. Instead, the two items are compared against each other and the one with the higher signal value is chosen (Egan, 1975; Glanzer & Bowles, 1976; Green & Swets, 1966; Hicks & Marsh, 1998; Macmillan & Creelman, 1991; Smith & Duncan, 2004; Zechmeister & Nyberg, 1982). According to this idea, we rely on the familiarity or signal-strength *difference* (relative familiarity or strength) between the two items to make a perceptual or memory recognition decision (Glanzer & Adams, 1990; Glanzer & Bowles, 1976; Green & Swets, 1966).

Because the decision is supposed to be strictly based on the strength *difference* between the two items rather than on an absolute criterion one sets up, some researchers refer to the 2AFC test as criterion-free (Egan, 1975; Green & Swets, 1966; Macmillan & Creelman, 1991; Zechmeister & Nyberg, 1982; Hicks & Marsh, 1998). In the context of signal detection, for example, Green and Swets (1966) 2AFC model posits that each of the two items generates a signal-to-noise likelihood ratio, and the observer chooses the one with the higher signal-to-noise ratio, which is to say that the choice is based on the *difference* between the two likelihood ratios. Thus, based on our understanding of the term “criterion-free” in the literature, it means regardless of where an absolute criterion is set, the signal strength *difference* between the two items does not change (analogous to two people standing on a floor; the taller of two persons is always taller whether they stand on a low or high floor, hence the height of the floor is irrelevant). Note that the difference cannot be computed without *comparing* the two items. The main purpose of the present study is to show some evidence that an old/new 2AFC in a recognition test may not be strictly based on the relative familiarity of the two items. A bias can occur toward making an individual-item recognition rather than strictly taking the relative familiarity as the basis of decision in a 2AFC for which subjects identify the old item. We call the old/new 2AFC a *weak-relational judgment*, meaning that the target sometimes can be recognized without comparing it with the distractor. On the other hand, in a comparative judgment where subjects choose the item with greater or lesser quantity of some attribute on a dimension (Birbaum & Jou, 1990; Jou, 2010; Moyer & Bayer, 1976; Moyer & Dumais, 1978), one cannot identify the target item without making a comparative judgment on the two choices. If this is indeed the case, the concept that an old/new 2AFC in a recognition test is strictly based on the difference or relative familiarity between the two items may be an oversimplification.

We hypothesize that the nature of the relation between the two items determines the degree to which a comparison between the two items is necessary in order to determine which of the two items is the target. If the choice is based on a purely relational judgment, then the decision cannot be made without comparing the two items. We suggest that in an old/new 2AFC, the choice decision does *not always* have to be based on a relational judgment. We emphasize the important distinction early in the paper that *not always* is different from *always not*. We do not assume that in an old/new 2AFC test subjects *always do not* make a relative judgment. Instead, we are proposing that people can *probabilistically* base their decision in an old/new 2AFC test on a single-item identification. In other words, people can *sometimes* skip the comparison and base the decision on a single-item familiarity judgment. In a strong relational judgment, it is very unlikely to be able to determine which one is the target by simply evaluating that item without comparing it with the other. Here is an example of a strong relational judgment. Two items located at two points on an attribute continuum are compared, and the task is to choose the one with a higher or lower value on the attribute dimension (Banks, 1977; Jou, 2010; Moyer & Dumais, 1978). For example, to determine “Which city is further to the west, Reno or Kansas City?” one cannot make a decision by evaluating only the location of one city without knowing the location of the other city. Put differently, one cannot predict the location of the second city on the basis of the knowledge of the location of the first city except when one of the two cities is located at the east or west end of the spatial dimension. In this sense, the information contained

in one item is not redundant with the information contained in the other item. Again, in that sense, the relational judgment is a *strong relational* judgment. On the other hand, in an old/new 2AFC recognition, if one can recognize the first item as old, then the second item has to be new, and moreover, the recognition of the first item as old does not necessarily require comparing it with the second item. That is, in an old/new 2AFC recognition test, the information in one item is at least somewhat redundant with the information in the other item, and hence the successful recognition of one item can make computing of the interitem difference unnecessary. In that sense, this type of relational judgment is a *weak relational* judgment. McKenzie, Wixted, Noelle, and Gyurjyan (2001) expressed the same view about the old/new 2AFC. They indicated that because the two items in an old/new 2AFC are mutually exhaustive and exclusive, if one item can be identified as old, the status of the other item is determined (i.e., new), and vice versa. The logic is that once the first item is successfully recognized as “old”, checking the second item is *in principle* unnecessary (again, although this does not mean that people always do not check the second item). In the case where the recognition is based on a single item, the old/new 2AFC is practically reduced to a one-item yes/no recognition. Also, due to this redundant relation between the two items, if one fails to determine whether the first item is old or new, one can have a “second” chance for solving the problem, i.e., by looking at the other item. For example, if one can determine that the second item is new with certainty, then one chooses the first item as old even if one does not recognize it. Thus, because of the redundant relation between the two items, evaluating the second item can increase the probability of making a correct choice. However, the increased accuracy comes at a cost, that is, evaluating the second item takes additional time. A tendency toward probabilistically skipping some portion of comparison process in a 2AFC recognition test may serve the goal of achieving an overall optimal task performance by trading off a little accuracy for a faster speed (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Selmecky & Dobbins, 2013). This tendency or bias can “contaminate” or comprise the decision process in the old/new 2AFC recognition as a process which is believed to be strictly based on the relative familiarity.

According to the mutually exhaustive and exclusive view of an old/new 2AFC, the decision task on two items becomes easier in a 2AFC than in a YN test because of the mutual constraints the two items impose on each other (Macmillan & Creelman, 1991; McKenzie et al., 2001). In a single-item YN recognition, there are four possible outcomes for two items presented as two probes, yes/yes, yes/no, no/yes, no/no. When two items are presented in an old/new 2AFC, there can be only two outcomes, yes/no and no/yes. Although some studies reported the equivalence of a 2AFC and a YN test as a measure of recognition memory (Kroll, Yonelinas, Dobbins, & Frederick, 2002 using pictures; Green & Moses, 1966 using nonsense syllables), many other studies reported otherwise. Using Deese-Roediger-McDermott paradigm materials, Westerberg and Marsolek (2003) found that the overall recognition performance was better with a 2AFC than with a YN test. Weinstein, McDermott, and Chan (2010) and Jou and Flores (2012) found that false recognition rates of the critical lure words in the Deese-Roediger-McDermott paradigm were substantially reduced in a 2AFC relative to a YN recognition test. Similarly, Deffenbacher, Leu, and Brown (1981) reported that face recognition was better in a 2AFC than in a corresponding YN test. Likewise, Yonelinas, Hockley, and Murdock (1992) reported that their subjects judged the YN test to be more difficult than the 2AFC. These findings were consistent with the *information redundancy* hypothesis of decision in the old/new 2AFC.

RT should reflect the amount of information retrieved and computed before the decision is made. A shorter RT for a left choice would imply that some portion of the decisions is based on an absolute judgment on the left item rather than a fully executed comparative judgment. The function of the tendency toward probabilistically skipping an evaluation of the right item or the comparison process (when the left item is recognized as old) is to speed up the decision process. In cases

Download English Version:

<https://daneshyari.com/en/article/919636>

Download Persian Version:

<https://daneshyari.com/article/919636>

[Daneshyari.com](https://daneshyari.com)