Food Research International 97 (2017) 265–271

Contents lists available at ScienceDirect

Food Research International

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

Cognitive decision strategies adopted by consumers in reminder difference tests: Influence of the authenticity test

M. Stocks[a], D. Shepherd[b], H.-S. Lee[c], D. van Hout[d], M.J. Hautus[a,*]

[a] School of Psychology, Faculty of Science, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[b] Department of Psychology, Auckland University of Technology, 90 Akoranga Drive, Auckland 0627, New Zealand
[c] Department of Food Science and Engineering, Ewha Womans University, Seoul 120-750, South Korea
[d] Unilever Food and Health Research Institute, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands

ARTICLE INFO

Keywords:
Consumer
Difference test
Discrimination
Authenticity test
Reminder paradigm
Decision strategy

ABSTRACT

Discrimination tests are used in food companies to quantify small differences between products. Within the diversity of methods available, some are quicker to conduct, whereas others are more sensitive or statistically powerful. One class of methods includes the reminder tasks in which the reference product is given before tasting the actual test stimuli. During the task, such a 'reminder' can be compared directly to each test stimulus, or alternatively, only serve to prime the memory of the judge without being taken into account in decision-making. Previous research with trained judges provided evidence for the latter process while research with untrained consumers has provided some evidence for the former process. Two studies were conducted with untrained consumers using the A Not-AR and 2-AFCR reminder tasks. Objectives were to determine the decision strategies used in, and the relative sensitivity of the tasks. In addition, the use of an "authenticity test" was explored to see if this has a positive effect on test performance. In the first study, mayonnaise and ice tea with small stimulus differences ($d' < 1$) were used in A Not-AR and 2-AFCR. Results were compared to those from A Not-A and 2-AFC tasks, with and without an authenticity test. It was difficult to draw clear conclusions on the decision strategy used, though the use of an authenticity test increased the sensitivity for these small differences, as it improved the performance of 6 out of 8 tests. In the second study, ice teas with larger stimulus differences (at two levels) were tested using the A Not-AR and 2-AFCR tasks, in comparison to the same-different task. The results showed that consumers use the less optimal strategies and that the authenticity test decreases performance, which is contradictory to the results of the first study. It seems that for very small stimulus differences the authenticity test can improve performance, but with larger differences the authenticity test decreases performance; it seems to confuse the judges.

1. Introduction

The food industry employs sensory difference tests to quantify differences between products and with different formulations. Some tasks are quicker to set up and for judges to complete, and some tasks are more sensitive to stimulus differences than others. It is important to consider these issues in the food industry as it strives to balance the need for accurate and efficient testing against the costs of an effective testing programme. There is increasing interest in tasks that use the reminder paradigm (Bi, O'Mahony, & Lee, 2013; Jeong, Kang et al., 2016; Jeong, van Hout, Groeneschild, & Lee, 2016; Kim, Chae, van Hout, & Lee, 2014; Lee, van Hout, & Hautus, 2007; Shin, Hautus, & Lee, 2016; Stocks, van Hout, & Hautus, 2013). This is where a standard task uses a reminder stimulus as a comparison at the beginning of each test. Examples include the reminder versions of the A Not-A and two-alternative forced-choice (2-AFC) tasks which are named the A Not-A with reminder (A Not-AR) and the 2-AFC with reminder (2-AFCR) tasks, respectively. The reminder stimulus could aid judges in making a correct decision by allowing a recently presented reference stimulus (the reminder) to be compared to the test stimuli (Jeong, Kang et al., 2016; Kim et al., 2014). From a different perspective, the reminder stimulus may be used only to prime memory and the judge may not even compare the reminder directly to the test stimuli (Stocks et al., 2013). These possible uses of the reminder can lead to different performance outcomes.

To complicate matters, most tasks have more than one decision

* Corresponding author.
E-mail addresses: m.stocks@auckland.ac.nz (M. Stocks), daniel.shepherd@aut.ac.nz (D. Shepherd), hlee@ewha.ac.kr (H.-S. Lee), danielle-van.hout@unilever.com (D. van Hout), m.hautus@auckland.ac.nz (M.J. Hautus).

http://dx.doi.org/10.1016/j.foodres.2017.04.007
Received 25 January 2017; Received in revised form 6 April 2017; Accepted 7 April 2017
Available online 12 April 2017
0963-9969/

strategy that a judge may use. Sometimes the different available strategies for a given task result in a different estimate of sensitivity. Therefore, it is essential to determine the strategy used by the judge to accurately estimate sensitivity and quantify perceptual differences between the products under investigation. Even though signal detection theory (SDT) predicts that sensitivity will be the same on any task for the same decision strategy, judge, and stimuli, this is often not the case in the real world. Some tasks are more sensitive to stimulus differences (have better 'test sensitivity') than others (Choi, Kim, Christensen, van Hout, & Lee, 2014; Jeong, Kang et al., 2016; Kim & Lee, 2012; Kim, Sim, & Lee, 2015, 2016; Kim et al., 2014; van Hout, Hautus, & Lee, 2011). The differences in measured sensitivity may result from factors unaccounted for in SDT. For example, tasks that have more than one stimulus on a test (e.g., 2-AFCR, same-different task) have greater carry-over effects and memory load than a single-stimulus task (i.e., A Not-A) (O'Mahony & Goldstein, 1986; Lau, O'Mahony, & Rousseau, 2004; Stocks et al., 2013; Stocks, van Hout, & Hautus, 2014). These factors reduce apparent sensitivity as measured by SDT.

The most common method used to determine decision strategies adopted by judges in a difference test is a comparative approach. Consider a judge, or group of judges, is performing two tasks, of which the first task is a so-called "baseline" task in which only one decision strategy is possible (or likely), and the second task is amenable to two decision strategies. Estimation of d' for the second task will produce two values; one for each strategy. Using the principle that sensitivity should be similar on both tasks, comparison of those two d' estimates to that obtained for the baseline task can provide an indication of the decision strategy used in the second task.

In the A Not-A task, *A* or *B* is presented to the judge who then indicates which was presented (refer Table 1; Green & Swets, 1988). For the β-strategy, $d' = z(H) - z(F)$ (Green & Swets, 1988, pp. 15–23; Hautus, van Hout, & Lee, 2009), where *H* is the hit rate and *F* is the false-alarm rate. As there is only one decision strategy for this version of the A Not-A task (but see Wichchukit and O'Mahony (2010) or Santosa, Hautus, and O'Mahony (2011), for possible decision strategies for other versions of this task), estimates of d' obtained from the A Not-A task can be considered relatively free of the effects of different decision strategies. A Not-A therefore makes a good candidate for a baseline task in the determination of decision strategies for other tasks.

Consider the A Not-AR task for an example of how this method works. The A Not-AR task extends the A Not-A task by including a reference (reminder) sample (i.e., *A*) presented immediately prior to the standard *A* or *B* sample, with the judge reporting whether *A* or *B* followed the reminder. Here, a judge may use the β-strategy where they effectively ignore the reminder when making a decision and partake in an A Not-A task (though the reminder may still function as a cue to the characteristics of *A*), or the τ-strategy, in which the reminder stimulus is used in decision making and the judge assesses the difference between the reminder sample and the test sample (refer Table 1). For the A Not-AR τ-strategy, $d' = \sqrt{2}(z(H) - z(F))$ (Hautus et al., 2009). The A Not-AR task yields two d' estimates; one for each decision strategy. The estimate most similar to that for the baseline task, in this case the A Not-A d' estimate, would stem from the model based on the decision strategy most likely adopted by the judge in the A Not-AR task.

We now briefly outline the other tasks of interest in this study. For the 2-AFC task, both *A* and *B* are presented in sequence (i.e., $\langle A,B\rangle$ or $\langle B,A\rangle$) and the judge indicates whether *A* occurred first or second. Here, the β- or τ-strategy can be used, and as both lead to the same level of performance $(viz\ d' = (z(H) - z(F))/\sqrt{2})$ it is convenient to refer to them collectively as the β/τ strategy (refer Table 1; Stocks et al., 2013). Therefore, it is not necessary to determine which strategy was adopted and a single model can be fitted to the data, making 2-AFC, like A Not-A, a good task to use as a baseline for comparison.

The 2-AFC with reminder task (2-AFCR) contains a reference sample (i.e., *A*) presented prior to a standard 2-AFC task (i.e., $\langle A,A,B\rangle$, and $\langle A,B,A\rangle$), with judges indicating whether *A* was first or second after the reminder. For the β-strategy, the reminder is only used as a cue to aid memory and is effectively ignored in decision making, and sensitivity is the same as for the 2-AFC task, while for the τ-strategy the judge separately compares the first and second samples with the reminder (Stocks et al., 2013). As for 2-AFC, sensitivity is the same for both the β- and τ-strategies (i.e., the β/τ-strategy) and again $d' = (z(H) - z(F))/\sqrt{2}$ (Hautus et al., 2009). However, there is an additional strategy available for the 2-AFCR task, the comparison of distances strategy (COD strategy), also a differencing strategy like τ, in which the judge ignores the direction of difference between the reminder and each test sample and makes judgments based on the magnitude of the differences (refer Table 1; Ennis, 1993). This strategy is based on how similar samples are to each other rather than the identities of the samples. It is equivalent to the strategy generally assumed to be used by judges in the fixed-reference duo-trio (procedurally identical to 2-AFCR) and balanced-reference duo-trio tasks. Sensitivity is a little more complicated to calculate in 2-AFCR for this strategy from *F* and *H*, and not reported here (see Hautus, Shepherd, & Peng, 2011b for details).

Finally, the same-different task requires a judge to respond either "same" or "different" when presented with a pair of samples: $\langle A,A\rangle$, $\langle B,B\rangle$, $\langle A,B\rangle$, or $\langle B,A\rangle$. A judge may use the β-strategy, in which each sample is considered independently, or the τ-strategy, in which the judge assesses the difference between the two samples on a test and compares this difference to a criterion, frequently referred to as the τ-criterion (refer Table 1; Lee, van Hout, Hautus, & O'Mahony, 2007). The majority of research using the same different task with simple stimulus differences has found that judges use the τ-strategy (Hautus & Collins, 2003; Irwin, Hautus, & Francis, 2001; Irwin, Hautus & Stillman, 1992). Furthermore, this task has been recommended as an appropriate overall sensory difference test for consumer judges and it has been broadly accepted that consumer judges use the τ-strategy (O'Mahony-& Rousseau, 2003).

In addition to determining the decision strategies used by consumers in the reminder tasks, the current research considers the sensitivity of consumers and whether it can be enhanced by including a test that elicits emotional responses, the so-called authenticity test (Mojet & Koster, 1986; cited in Frandsen, Dijksterhuis, Brockhoff, Nielsen, & Martens, 2003). In this authenticity test consumers are, prior to the discrimination task in the test itself, presented with an upsetting story about the product that they care about. The assumption is that this would lead to affective responses which would make them more sensitive to the product differences.

Research on the efficacy of authenticity tests has produced mixed results, with some showing positive results (Frandsen, Dijksterhuis, Brockhoff, Nielsen, & Martens, 2007; Kjearulff, 2002), and others not. For example, Forde, O'Riordan, and Williams (2006) used Vegemite® (an iconic Australian food) versus a spiked version of Vegemite®. Judges were Australian citizens who were informed that "only real Australians can distinguish the true taste of Vegemite®" from an

Table 1
Characteristics of the five difference tests employed.

Test	Sample sequence	Response	Strategies
A Not-A	*A* or *B*	1 = *A*, sure … 6 = *B*, sure	β
A Not-AR	***A****A* or ***A****B*	1 = *A*, sure … 6 = *B*, sure (after the reminder)	β or τ
2-AFC	*AB* or *BA*	1 = *A* was 1st, sure … 6 = *A* was 2nd, sure	β/τ
2-AFCR	***AA****B* or ***A****BA*	1 = *A* was 1st, sure … 6 = *A* was 2nd sure (after the reminder)	β/τ or COD
Same-different	*AA* or *AB* or *BB* or *BA*	1 = Same, sure … 6 = Different, sure	β or τ

Note: Bold *A* indicates the reminder sample; β/τ indicates either the β- or τ-strategy for difference tests where both strategies result in the same level of performance.