Food Quality and Preference 39 (2015) 8–15

Contents lists available at ScienceDirect

Food Quality and Preference

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

The triadic preference test

Eduardo Calderón [a,b], Alondra Rivera-Quintero [a], Yixun Xia [a], Ofelia Angulo [b], Michael O'Mahony [a,*]

[a] Department of Food Science and Technology, University of California, Davis, CA 95616, USA
[b] Unidad de Investigación y Desarrollo de Alimentos, Instituto Tecnológico de Veracruz, Veracruz 91860, Mexico

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form 7 May 2014
Accepted 7 May 2014
Available online 12 June 2014

Keywords:
Preference testing
Triadic preference test
Placebo pairs
Extraneous factors
Placebo test 'screening'

ABSTRACT

A triadic preference test was developed as an alternative to the paired preference test. The three stimuli in the test comprised a putatively identical placebo pair and a different stimulus. This was in contrast to the regular paired preference test that utilizes a placebo pair. Such a test requires the presentation of two pairs of stimuli: a putatively identical placebo pair and a test pair. The triadic preference test only requires one triad. With the regular test, the majority of consumers respond to the placebo pair with a preference response. It is generally assumed that these consumers are responding to extraneous factors: those factors that elicit a preference response that are different from the sensory attributes of the food under assessment. As an attempt to minimize the possibility of responses to extraneous factors when assessing the test pair, it has been suggested to only use those consumers who chose the 'No Preference' option for the placebo pair. However, this form of 'screening' is not viable because the resulting 'screened' sample size is greatly reduced to approximately one third. However, in the present study, with the triadic preference test, the resulting 'screened' sample size ranged 76.5–94% of the total. Thus, this form of 'screening' against consumers who demonstrated response to extraneous factors for the placebo pair, was now feasible.

Introduction

Preference and acceptance tests are important for product development and decisions regarding the launching of new products on to the market. A simple test of preference is the paired preference test (Lawless & Heymann, 2010; Resurreccion, 1998; Stone & Sidel, 2004). With the 'No Preference' option included, the data obtained from such a test specify the proportion of consumers who prefer each product and the proportion who have no preference, along with the sample size, to give an idea of the power of the test. Sometimes, a hedonic d' is used as a single overall measure to indicate the strength of preference for one product over the other (Alfaro-Rodriguez, Angulo, & O'Mahony, 2007; Alfaro-Rodriguez, O'Mahony, & Angulo, 2005; Angulo & O'Mahony, 2005). Accompanying the preference test, questions regarding liking/disliking for each product should be included, to give further insight into the reasons for the measured preferences. For example: whether a consumer liked both products but happened to like one more than the other or whether a consumer had a preference because one product was liked while the other was disliked, etc.

As a casual test, the paired preference test is a simple tool for preliminary experimentation. Yet, for more formal testing, there are problems that need to be addressed. One such problem is a tendency to report preferences when the stimuli are putatively identical. Ennis and Collins (1980) mailed two cigarettes (call them 'X' and 'Y') to a large number of consumers' homes for comparison on a variety of attributes like: better flavor, easier draw, better aftertaste, and slower burning. Finally, they were asked for their preferences and 40% reported preference for cigarette 'X', 20% reported 'No Preference' and 40% preference for 'Y'. Yet, 'X' had been taken from the same production run as 'Y'; they were essentially the same cigarette. This experiment was repeated for four different brands of cigarettes with consumer sample sizes ranging 412–488 (total 1787), giving remarkable agreement between each brand. Because these preferences are not systematically related to the properties of the products in the test that are of relevance, it is a matter for concern.

The question becomes why consumers should respond to a putatively identical pair of stimuli with a preference response. It may be assumed that the response was not the result of the assessment of the sensory properties of the putatively identical stimuli. The sensory input elicited by the attributes of the putatively identical pair, would be close enough to 'identical' to be deemed as being elicited by the same product. In which case, it would appear to be due to something else, which here will be called 'extraneous

* Corresponding author. Address: Sensory Innovation Lab, Department of Food Science and Technology, University of California, Davis, CA 95616, USA. Tel.: +1 530 756 5493; fax: +1 530 756 7320.
E-mail address: maomahony@ucdavis.edu (M. O'Mahony).

http://dx.doi.org/10.1016/j.foodqual.2014.05.016
0950-3293/

factors': those factors that elicit a preference response that are not a result of input from the sensory attributes of the foods under assessment. Such extraneous factors could have consequences for the main preference tests that are being performed with the pair of different products that are of interest. This will be called the 'test' pair. If eighty percent of consumers can respond to 'extraneous factors' when presented with the putatively identical pair, how many of these consumers might respond only to extraneous factors when assessing the test pair? The problem is that there is no obvious way of knowing. The potential for misinformation is not to be ignored.

Ennis and Collins's (1980) 40–20–40 frequency distribution for putatively identical stimuli does not seem to be general. For reasons as yet unresolved, other authors (Alfaro-Rodriguez, Angulo & O'Mahony, 2005, 2007, 2008; Alvarez-Coureaux, Aguilar, O'Mahony, & Angulo, 2010; Angulo, Okayama, Nakamura, Yuen, & O'Mahony, 2009; Chapman, Grace-Martin, & Lawless, 2006; Chapman & Lawless, 2005; Kim, Lee, O'Mahony, & Kim, 2008; Marchisano, Lim, Cho, & Suh, 2003; Sung, Lee, O'Mahony, & Kim, 2011) found different frequencies, the numbers varying with the products being tested, the experimental conditions, the types of consumers tested and the types and numbers of response options allowed in the test. The frequencies vary a great deal but most appear to be in the 20–35% range. Yet, in nearly all cases, the majority of judges indicate preferences rather than no preference.

The safest approach at the present time would appear to be to require consumers to assess the 'test' (different) pair and also the putatively identical pair, to give a measure of the effect of the extraneous factors. This was adopted and required giving consumers two pairs of products to assess in a given test (Alfaro-Rodriguez, Angulo, & O'Mahony, 2007, 2008; Alvarez-Coureaux, Aguilar, O'Mahony, & Angulo, 2010; Kim, Lee, O'Mahony, & Kim, 2008; Sung, Lee, O'Mahony, & Kim, 2011). The response frequencies elicited by the test pair of stimuli could be compared with the responses frequencies elicited by the putatively identical pair, later called the 'identicality norm' (Christensen, Ennis, Ennis, & Brockhoff, 2014) to determine whether they were significantly different. If they were, it could be concluded that the preference responses elicited by the sensory input from the attributes of the test pair, were not solely a response to extraneous factors in the testing situation. Because the putatively identical pair was being used in a way similar to a placebo in drug testing, it was called the 'placebo' pair (Alfaro-Rodriguez, Angulo, & O'Mahony, 2007).

Two lines of research emerged from the potential problems of 'extraneous factors' present in the paired preference test, first reported by Ennis and Collins (1980). The first we will call statistical. The safest approach, as mentioned above, was for consumers to taste both a placebo pair and a test pair of stimuli, so that comparisons could be made between the two sets of response frequencies. Chi-squared comparisons were used to determine whether there was a significant preference indicated in the test pair. In this way, consumers were their own control. There was some investigation (Sung, Lee, O'Mahony, & Kim, 2011) regarding the use of the related samples Bowker test (Bowker, 1948) in this analysis, although it was not adopted.

Comparison between preference tests using different measurement protocols (e.g. with or without the 'No Preference' response option) is problematical when the data are represented as frequencies of response. Instead, it is convenient to use a hedonic variant of the fundamental measure, d', derived from Thurstonian modeling (O'Mahony, Masuoka, & Ishii, 1994; O'Mahony & Rousseau, 2002; Lee and O'Mahony, 2004) and from Signal Detection Theory (Green & Swets, 1966; Macmillan & Creelman, 1991). Values of d' were used by Alfaro-Rodriguez, O'Mahony, and Angulo (2005) to compare protocols which did and did not use 'No Preference' options. For d' values using tests without a response option, tables for the 2-AFC test (Ennis, 1993) can be used, while for tests with a response option, the computation for the 2-AC test is used (Braun, Rogeaux, Schneid, O'Mahony, & Rousseau, 2004). More importantly, for tests with a 'No Preference' option, values of d' for the test pair and the placebo pair have been compared to determine whether they were significantly different (Alfaro-Rodriguez, Angulo & O'Mahony, 2007; Sung, Lee, O'Mahony, & Kim, 2011). Recently, there has been a renewed interest in the statistics associated with this computation (Christensen, Lee, & Brockhoff, 2012; Christensen et al., 2014; Ennis & Ennis, 2012a, 2012b; Jesionka, Rousseau, & Ennis, 2014).

The second line of research was based in experimental psychology and it is this line of research with which this paper is concerned. With this line of research it is important to keep in mind the goal of preference testing. The preferences measured in sensory tests are called 'test preferences'. The goal of a 'test preference' is not to predict preference behavior under test conditions but to predict preference behavior under 'everyday real life' conditions. These are called 'operational preferences' (Wichchukit & O'Mahony, 2010).

Whereas the statistical line of research examined how better to interpret the data from a test pair of products by statistical comparison with the data from placebo pairs, the goal of the experimental psychology approach was different. It was to investigate the variables involved with the preference responses elicited with the placebo pair, with the long term goal of increasing the proportion of 'No Preference' responses. The goal was to design a preference testing protocol which did not elicit responses based on extraneous factors. Freed of this tendency, it is reasonable to hypothesize that consumers would give a more valid picture of consumer preference. Test preference would be better at predicting operational preferences.

The variables associated with the various experimental protocols that tended to vary the proportion of 'No Preference' responses to the placebo pair were investigated. For example, Marchisano et al. (2003) noted that if more than one level of 'No Preference' was available as a response, a larger proportion of 'No Preferences' was elicited. They also noted that Koreans were particularly reluctant to give 'No Preference' responses. Angulo, Okayama, Nakamura, Yuen, and O'Mahony (2009) and Sung, Lee, O'Mahony, and Kim (2011) noted that 'buying' preferences induced more 'No Preference' responses than 'liking' preferences. Alfaro-Rodriguez, Angulo, and O'Mahony (2008) found that the tendency to give 'No Preference' responses to the placebo pair was not stable; sometimes a consumer had it while at other times it might be absent.

Yet, none of these studies solved the problem of how to eliminate the possibility of response to extraneous factors in the test pair. Alfaro-Rodriguez et al. (2007) and Sung et al. (2011) approached the problem by not only considering the data from all the consumers tested but also by examining the response frequencies in the test pair from the sub-set of consumers, who had reported 'No Preference' with the placebo pair. The advantage here was that only the responses of consumers, who had demonstrated a tendency not to report preferences elicited by extraneous factors, were considered. In other words, the placebo pair acted as a form of 'screening' tool by not selecting consumers who had demonstrated a tendency to report preferences elicited by extraneous factors. Naturally, it was not expected that giving a 'No Preference' response to a placebo pair would be a stable attribute of a consumer; it would be expected to come and go. The disadvantage of this approach was the small number of 'No Preference' responses to the placebo pair. The sample size of these 'screened' consumers was reduced, often to around one third of the total. This is unacceptably small and renders this form of screening impractical. Yet, if a high proportion of consumers passed the placebo 'screening' tool, this approach would be feasible.