Review

Paired Preference Testing: False Preferences and Disruptive Protocols

Yixun Xia, Fang Zhong*, and Michael O'Mahony¹

Department of Food Science and Technology, Jiangnan University, Wuxi, 214122, Jiangsu, China

¹Department of Food Science and Technology, University of California, Davis, CA 95616, USA

Received July 6, 2015
Revised October 5, 2015
Accepted October 5, 2015
Published online February 29, 2016

*Corresponding Author
Tel: +86-138-125-36-912
Fax: +86-510-8519-7876
E-mail: fzhong@jiangnan.edu.cn

pISSN 1226-7708
eISSN 2092-6456

© KoSFoST and Springer 2016

Abstract This review introduces recent changes regarding measurement of preferences brought about by the discovery of false preferences that occur with paired preference testing. Testing protocols are examined critically. False preferences result from a response bias elicited by factors distinct from sensory properties of products being tested that can distort results. Statistical approaches to circumventing this problem are examined. Alternative methods are being developed to eliminate false preferences based on understanding information processing in the brain. This has led to development of disruptive experimental preference testing protocols based on hypothesized use of processing subroutines in the brain to protect the central processing system from overload. These disruptive protocols suppress response bias. Paired preference testing is designed to predict consumer behavior. A paucity of studies addressing whether they do is put into perspective.

Keywords: preference tests, false preferences, response bias, cognitive subroutines, disruptive protocols

Introduction

Testing and operational preferences Measures of consumer acceptance of foods are important for product development and decisions regarding launching of new products in the marketplace. One such measure is the paired preference test (1-3). When used informally, it is simple. Yet, when used formally as a research tool or as a measure of market acceptance, there are concerns that must be addressed. These will be discussed in this article.

Preference responses given by consumers in a preference test in a place set aside for testing do not always correspond to preferences exhibited by consumers in real-life situations. For clarity, preferences measured during a preference test are defined as test preferences, while preferences displayed during the everyday life of a consumer, away from the testing environment, are defined as operational preferences (4). This review will consider research associated with the goal of developing methods for measurement of test preferences that can be used for prediction of operational preferences.

A test preference only gives part of the information required for gaining some insight into whether the test preference can be used for prediction of an operational preference. When a consumer reports a preference for A over B in a preference test, it is then necessary to ask further questions about liking. Consider a case where a consumer, on further questioning, reports liking product A, but disliking product B. It is reasonable to hypothesize that this test

preference may likely predict an operational preference. A consumer is more likely to choose a product that is liked than one that is disliked. On the other hand, if a consumer reports liking both products, yet happens to prefer product A to B at the time of testing, it may be argued that this test preference is less likely to be a predictor of any lasting operational preference. Given a desire for novelty or a change in mood, product B might be chosen over product A on some future occasion. If so, this would suggest an operational no preference. There is also the problem of deciding at what point no preference becomes a preference. If a consumer chooses product A over product B 50% of the time and vice versa, it is reasonable to declare that there is no preference. Yet, at what point does the choice ratio become a preference of 55:45%, or 60:40%, or any other preference? Any definition would be arbitrary as the problem is semantic. The term *preference* infers liking more, without defining how much more with no exact behavioral definition.

To avoid problems regarding lack of definition for terms based loosely on the concept of *liking*, use of more actionable measures of preference have been examined (5-9). Measures of likelihood to buy, or purchase intent, have been compared with measures based on liking for paired preference testing. Items (a) and (b) in Table 1 illustrate typical examples of preference response options concerned with the degree of liking. Purchase intent is a more realistic measure for marketing than liking. Consumers might buy products for the family that they themselves do not like or might not buy for reasons

Table 1. Typical preference response options for *Liking* and *Likelihood to Buy*

Type of preference		Preference Options		
Liking	(a)	Prefer 'A' ¹⁾	No Preference	Prefer 'B'
	(b)	I like 'A' more	I like them just about the same	I like 'B' more
	(c)		I like neither	
Likelihood to buy (Purchase intent)	(d)	I would only buy 'A' and never buy 'B'	Depending on price, availability and mood, I might buy either	I would only buy 'B' and never buy 'A'
	(e)	I would buy 'A' more than 'B'	Depending on price, availability and mood, I might buy either	I would buy 'B' more than 'A'
	(f)	I would buy neither	I don't buy these products	

¹⁾In reality, coded products would not be coded with letters but would be coded using 3-digit random numbers.

of health. Compared with liking options (a) and (b), the purchase intent option (d) elicits a markedly higher proportion of consumers for choosing the central option of having no buying preference. This is not surprising, with such a highly committed preference option. Interestingly, the less committed buying preference options in item (e), also have the same effect of increasing the proportion of consumers choosing the central option of having no buying preference, albeit to a lesser extent. The same effects were found with choosing rather than buying options (7). Items (c) and (f) are sometimes presented as response options for consumers who neither like nor would buy either product, although this option becomes unnecessary if only regular consumers of products are included in testing.

Use of more actionable response options for buying or choosing is appropriate for different consumer situations and behaviors. Yet, the effect of increasing the proportion of consumers who express no preference may be useful in a quest to reduce the number of false preferences. It may also be argued that results of purchase intent testing using option (d) would be more likely to predict operational buying preferences than other, less extreme, options. At this point, it is worth noting that products presented to consumers should not be identified in any way. Letters, such as A or B, should not be used while 3 digit random numbers should be used.

The No Preference option A concern with the methodology of the paired-preference method is a disagreement about use of a *No Preference* response option. It is argued that a no preference option should not be allowed based on confusing preference testing with difference testing. Difference testing is generally used to determine whether subjects can discriminate between 2 foods that are so similar in flavor and appearance that it is not clear whether they can be distinguished or not. Such testing is used to study the effects of ingredient changes and processing and packaging changes, or shelf life studies where any detected differences would be at threshold levels.

There are several difference testing protocols available. One such protocol is the Paired Comparison, (1,10) often called the 2-Alternative Forced Choice, or the 2-AFC procedure (11). A judge is required to indicate which of 2 samples is stronger for a specified

attribute. The 2-AFC method is a forced-choice test. The judge is forced to choose one of 2 foods and there is no *Neutral* or *No Difference* or *Cannot Discriminate* category. Instead, difference tests are repeated so that if a consumer can discriminate between 2 foods, choices given will be consistent and if the stimuli cannot be distinguished, choices will be inconsistent over repeated testing (Fig. 1). Thus, repetition of the difference test is a method of discovering whether a consumer can or cannot discriminate between 2 foods. The common reason given why a *Don't Know* or *Cannot Tell* response option is not used is that consumers, who would be capable of discriminating between the 2 foods presented to them, sometimes do not feel confident enough to report that they can. Yet, if told to guess, consumers may perform a set of repeated difference tests correctly while thinking they are guessing.

Surprisingly, consumers are not always aware that they can perceive small differences. Traditionally, binomial statistics provide a simple statistical analytical method for data derived from such testing. More recently, Thurstonian analyses (11) have enabled convenient measures of consumer sensitivity to small differences between foods based on term d' , a signal-to-noise ratio parameter derived from communications engineering (12).

The only commonality between paired preference testing and the 2-AFC method is that 2 foods are presented to a consumer. Otherwise, the tests are completely different. For paired preference testing, 2 discriminable products are presented to a consumer who then reports which one is preferred or whether there is no preference. Repeated testing, used for difference testing, is not required because, with preference testing, foods should be discriminable. Thus, a consumer would not even need to re-taste any food sample to give consistent answers, even if repeatedly questioned. This means that immediate repetition of the test is not a viable method of determining whether a consumer has no preference. The sensible option is to allow a third *No Preference* response option (Fig. 1). Yet, some researchers still prefer to omit the *No Preference* option in order to allow use binomial statistical analysis.

There is a further argument. Without a *No Preference* option, data interpretation can be difficult. If 50% of consumers report a preference for product A and 50% report a preference for product B,

THE 'NO PREFERENCE' OPTION

DIFFERENCE TEST				PREFERENCE TEST		
CAN TELL DIFFERENCE		CANNOT TELL DIFFERENCE		HAVE PREFERENCE	NO PREFERENCE	
stronger taste	weaker taste	stronger taste	weaker taste			
A > B		A > B		Prefer A	No Preference	Prefer B
A > B		B > A				
A > B		A > B		Use 'No Preference' option		
A > B		A > B				
A > B		B > A		Prefer A		Prefer B
A > B		B > A				
A > B		A > B		Not use 'No Preference' option		
A > B		B > A				
Consistent responses indicate judge can tell the difference		Inconsistent responses indicate judge cannot tell the difference		If not use 'No Preference' option, immediate repetition would elicit only automatic answers		

Fig. 1. Presentation schemes for 2-AFC difference and paired preference testing protocols illustrating how inconsistent responses to repeated testing indicates that a consumer cannot discriminate between two stimuli, while a *No Preference* response option is required for a consumer to indicate a lack of preference.

then no clear conclusion can be drawn if a *No Preference* option is not provided. One possible interpretation would be that all consumers had no preference and they simply chose products randomly. Yet, it would also be possible that half of the consumers actually did prefer product A while the other half preferred product B. The action to be taken by a food company would be different in each case. The *No Preference* option avoids this problem.

Despite the arguments given above, use of simple binomial analyses for paired preference testing leads to elimination of a *No Preference* option, thus, eliminating or distorting some data obtained. However, if no preference responses actually occur, the 3 main ways of dealing with them (13) are ignoring them, splitting them equally, or splitting them proportionally according to proportions of preference responses. These strategies introduce false alarms, altering the power of the test, and complications occur when the no preference proportion is not negligible (14,15). The conclusion is to abandon the practice of trying to fit data to a binomial analysis and use a more suitable analysis for unaltered data obtained during testing (13).

False preferences reported for putatively identical stimuli A potential problem with false preferences was reported for paired preference testing (16). Two cigarettes, X and Y, were mailed to a large number of consumers for comparison of a variety of attributes, including better flavor, easier draw, better aftertaste, and slower burning. Consumers were asked to indicate preferences. Of the 1,787 replies, 40% reported a preference for cigarette X, 20% reported no preference, and 40% a preference for Y. Yet, X and Y were putatively identical as cigarette samples had been taken from the same production run. Therefore, any preferences expressed for one or the other would have been due to factors other than sensory

characteristics. These factors will be called *extraneous factors* and defined as factors that elicit a response bias to report preferences that are not a result of input from sensory attributes of products under assessment. Extraneous factors may be derived from experimental conditions of preference testing which, in turn, initiate an internal response bias in the brain that demands reporting of these preferences. In addition to biasing responses to putatively identical pairs of products, such bias has potential to bias regular preferences recorded for target pairs of different products. If 80% of consumers can respond to extraneous factors when presented with a putatively identical pair, how many of these consumers might respond to extraneous factors when assessing a target pair? The problem is that there is no obvious way of knowing. The potential for misinformation should not be ignored.

An international research effort was initiated that involved laboratories in the USA, Korea, Central and South America, and East Asia. Initially, interest was in identification of variables associated with a putatively identical pair. It was soon found that a 40-20-40% frequency distribution (16) was not general. Response frequencies varied with products being tested, experimental conditions, types of consumers tested, and types and numbers of response options used in testing (5,8,15,17-24). Frequencies of *No Preference* varied, but most were in the 20-35% range. Yet, in all cases, the majority of consumers indicated a preference rather than no preference.

Why do consumers report preferences for a placebo pair? The question arises as to why consumers exhibit a response bias that causes them to report a preference when confronted with a placebo pair consisting of 2 putatively identical food samples? When questioned, consumer responses implied a preconception existed that a preference

should be given in a preference test (21,23).

In addition to consumers having a preconception that initiated a response bias that a preference must be given, there is another possibility that stimuli in the putatively identical pair did not elicit identical sensory input. Variability in momentary sensory inputs to the brain may have been dependent on neural noise and central processing. Such variability in momentary sensory input is important for sensory difference testing and has been explained based on central processing, invoking Signal Detection theory/Thurstonian modeling (11,12,25-29). Further variation in sensory input is explained based on effects of sensory adaptation to residual stimuli from prior tastings and stimulus dilution by saliva that affects performance, depending on the tasting sequence (30-33). Therefore, slight differences in sensory input to the brain were probably responsible for elicitation of preferences with a putatively identical pair of stimuli. This possibility is called the Input Hypothesis. A further variation in sensory input might be due to product variability. Even so, it would be surprising if all products tested in placebo pairs experienced bad enough quality control that noticeable differences in flavor were evident in samples coming from the same package or bottle.

This raises a question central to Signal Detection Theory/Thurstonian modeling of how big a difference is required between 2 sensory inputs for a consumer to accept and report that products are different? In Thurstonian modeling terms, this is called a tau-criterion. If consumers use a tau-criterion appropriate to difference testing, it is not surprising, given the effects described above, that they would perceive or think they perceived differences in sensory input from products in a placebo pair, and then infer a preference. Yet, confronted by a large number of consumers reporting such preferences, the question becomes one of whether the tau-criterion that is suitable for difference testing is also suitable for preference testing.

A consumer drinking 2 sips of a beverage does not usually remark that one sip tasted better than the other, even though they might have reported a difference under controlled conditions of discrimination testing. It would seem that use of a difference testing tau-criterion is too sensitive for preference testing. The tau-criterion probably needs to be larger for elicitation of operational preferences than for differences. Put simply, a difference between 2 products required to elicit a reported operational preference needs to be bigger than a difference that would merely elicit a reported difference. This has been demonstrated experimentally by comparison of detection and consumer rejection thresholds for unpleasantly reformulated products (34-36). Another way of stating this is that, for preference testing, the difference required for a consumer to be confident enough to report that 2 products are different is larger than the difference required for reporting that 2 sensations are different.

Consequently, for predictions of operational preferences, the tau-criterion would need to be larger than for difference testing. As a result, there would be a reduction in the proportion of consumers reporting no preference for a placebo pair. Yet, the problem remains

of how this can be achieved. Also, there is a further problem that the Input Hypothesis does not seem to correspond with subjective reports. However, a pressing question becomes how to deal with bias that affects judgments made for a putatively identical pair? One answer is provided by statistical analysis.

Statistical comparison between identical and target pairs for assessment of whether preferences are statistically significant: placebo pairs and identity norms

Data obtained from paired preference testing for a target pair of products are proportions of consumers who prefer product A, who prefer product B, and who have no preference. Equivalent responses to a putatively identical pair are due to factors other than sensory inputs elicited by products in a pair. In essence, these factors represent responses elicited in paired preference testing when consumers indicate no preference. Thus, these factors can be used as a control. If response frequencies for a target pair are compared with response frequencies to a putatively identical pair, then a significant difference would signal the presence of real preferences in the target pair. In this context, the putatively identical pair is called a placebo or a placebo pair because of a resemblance to a placebo in a drug trial (17,18). The frequency of preferences obtained using a placebo pair has been called the *identity norm* (37).

The identity norm varies with products being tested, experimental conditions, types of consumers tested, and types and numbers of response options allowed in testing (5,8,15,17-24). Accordingly, consumers assess both a target pair and a corresponding placebo pair so that response frequencies can be compared (9,17-19,22). If response frequencies for the target pair are significantly different from the placebo pair (the identity norm), it can be concluded that a significant proportion of consumers had preferences elicited by sensory properties of the target pair.

For presentation of 2 product pairs, it is important that the physical treatment and presentation of the target pair (A and B) and the placebo pair (placebo A and placebo B), even though the placebos are identical, are exactly the same. For example, if A and B are kept in separate packages, then placebo A and placebo B must be kept in separate packages. Results of whether the placebo pair should be presented before or after the target pair are conflicting or show no discernable effect (9,17,22,38). As a precaution, counterbalancing the order of presentation of the placebo and target pair is advisable, as well as the usual counterbalancing of items within pairs, which involves arrangements so that all possible order effects are balanced. The complexity involving counterbalancing schemes for products is outlined in Fig. 2.

In addition to inspection of preference and no preference frequencies, the preference trend can be represented by the single number d' , which is a value used to denote the degree of perceived difference in signal detection theory (11). When used with preference testing, it can be conceptualized as a hedonic d' , which describes the degree of preference rather than the degree of difference (14,15,39)

Fig. 2. Presentation scheme for pairs of target stimuli A and B and 2 placebo pairs indicating variations in the order of stimulus presentation that are required to balance order effects.

indicating which sample in a target pair consumers tend to prefer. Use of d' also has the advantage of allowing comparisons between test protocols that do and do not use the *No Preference* option (15).

Recently, there has been renewed interest in statistics associated with these computations (40-42). However, this statistical approach merely considers whether the d' value for a test pair and a placebo pair are significantly different, from which it can be deduced that there is a preference among some consumers for a target pair and responses to the target pair do not correspond to every consumer having no preference. However, consumers who show a response bias based on reporting preferences for a placebo pair are still present in samples used to assess a target pair. Therefore, they still have potential to distort results for the target pair.

There is an alternative statistical approach to elimination of the effects of bias based on analogy with Signal Detection Theory (11). Rather than simply noticing that d' values for target and placebo pairs are significantly different and inferring that an, albeit significant, proportion of consumers must have reported a preference for the target pair, it would be better to have a d' value that directly measures the difference between a placebo pair (noise) and a test pair (signal+noise). This would give a direct measure of the signal strength of the degree of preference, with the effect of elimination of biased consumers. A maximum likelihood computation has recently been developed for this (43). Having used a placebo pair as a control, the earlier problem remains that the Input Hypothesis for explanation of why consumers report preferences for a placebo pair does not seem to correspond with subjective reports.

An alternative approach based on how the brain organizes information processing Consumers give a variety of answers when

questioned about why they give preference responses for a placebo pair (23). Some consumers have a strong preconception that, because it is a preference test, products must be different and a preference is required. There are variations on this theme. Consumers report that they wish to please the experimenter by reporting a preference, while others wish to show the experimenter that they were really sensitive and could detect a difference. The feeling that preferences should be given was so strong that some consumers even apologized for not having a preference. One study described a subject whose response bias was so strong that, when presented with 2 clearly identical color swatches of material to judge a color preference, with color numbers and color names clearly visible, was suspicious enough to smell the swatches to see whether this was the reason for a difference (21). A comparison between American and Korean consumers indicated that Koreans are more prone to report preferences with putatively identical stimuli. Later work (8) confirmed the effect, but to a lesser extent. From all reports, no consumers reported that products in a placebo pair tasted different.

Based on reports regarding why consumers reported preferences for a placebo pair, an alternative approach was taken for understanding the nature of response bias that initiated preference responses. Casual observation (23) where a consumer might be tested more than once has noted that response bias is labile and not fixed. Whatever the mechanism, the result was a dichotomy of either a bias is operating or not operating. Thus, it would be sensible for consumers to assess a target pair of products when a response bias was not operating. The only way of determining this for a consumer is when a *No Preference* response is given for a placebo pair. Accordingly, if a target pair is tested immediately after, or even immediately before, a no preference result for a placebo pair,

Fig. 3. Representation of how placebo pairs indicate whether response bias is operative or inoperative and affects the decision to collect data from target pairs.

Fig. 4. Representation of how central processing in the brain is shielded from becoming overloaded, achieved by attenuation of sensory input and by formation of motor and cognitive processing subroutines.

response bias would probably not be operating when a target pair was assessed (Fig. 3). However, a problem arose when this approach was first applied (8). The proportion of consumers who chose the *No Preference* option for the placebo pair was small for liking preferences, ranging from 21-32%, and larger for buying preferences, ranging from 24-50%. In all cases, these numbers are unacceptably low for this approach to be used in industrial research. The same would have been true for previous studies that had consumers assessing placebo pairs had they used this approach (1,5,8,15,17-19,21-24) with values ranging from 20-35%. The problem is how to arrange for more consumers to have no operative response bias. For this, it is necessary to consider brain function.

A systems analysis of the brain considering strategies for protecting against central processing overload The problem of how to

increase numbers of consumers with no operative response bias became a problem for an approach based on experimental psychology and required a systems analysis of information processing in the brain (Fig. 4). The act of seeing, for example, requires so much information processing that the central processing system of the brain that deals with conscious thought has evolved systems for self-protection from central processing overload. One system is selective attention in which not all sensory input results in conscious awareness. For example, if a room full of people is asked what they saw on the way to work, many could give a description. If asked what they smelled on their way to work, few could respond. Attention to visual input is attended to while smell information is received, but does not necessarily become part of consciousness. Another system is sensory adaptation. A particular sensory input that remains at a constant strength becomes redundant. To protect from processing

overload, the intensity of a redundant input is decreased, sometimes vanishing completely. Yet, should the input change, the intensity is restored. The brain is programmed to monitor changes in input rather than register repeated and, thus, redundant input. Accordingly, consider a person who, upon entering a room, notices a slight smell. If volatiles in the room have reached equilibrium, the nose will respond to a continually constant input. As this is redundant information, the intensity of input reaching consciousness will gradually be reduced and the smell may even appear to vanish completely. Yet, if this person then leaves the room and re-enters, the input will no longer be constant. The perception of the smell will be restored.

There is also evidence of the central processor self-protecting from central processing overload with output activities. With practice, motor activities that, at first, require conscious attention, tend to become automated. Tying a necktie or shoelace, riding a bike, pressing a correct pedal sequence, changing gears in a car, all tend to become automated. These processes can be thought of as being transferred from central processing to a processing subroutine to avoid overloading the central processor. Thus, it is often difficult to perform or explain such activities consciously. It can be difficult to describe how to tie a necktie or the sequence of pedals that need to be pushed when changing gears.

Just as the brain has developed motor subroutines to relieve the central processor of processing overload, the brain has also developed cognitive subroutines. Selection of vocabulary for a second language is, at first, a conscious activity, but becomes automatic with practice. It can be conceptualized as having been transferred to a cognitive subroutine. A second candidate is Kahneman's fast thinking (44), which can be best described as thinking on auto-pilot, or thinking without using the level of attention that would be engaged with central processing. Questions that are asked frequently, with slight

variation, become stereotypical of a class of questions. With practice, these questions can be answered automatically. Answers given can be envisioned as the product of fast thinking. An automatic answer is provided by a cognitive subroutine (Table 2).

Consider a consumer being asked a question that resembles, yet is slightly different from, a stereotypical question, the answer to which is controlled by a cognitive subroutine. Under fast thinking, the consumer may simply give the answer to the question controlled by the subroutine rather than an actual answer for the question being asked (Table 3). Any teacher who marks exams will be familiar with this phenomenon. More importantly, this phenomenon has potential to generate misleading information when using consumer acceptance measures.

Use of brain modeling for development of disruptive protocols to increase the proportion of consumers whose response bias is not operative at the time of testing These ideas can be applied to preference testing. When presented with 2 products and asked a preference question, consumers could be considered as acting automatically by switching to a familiar binary choice subroutine that includes response bias for giving a preference when a preference is requested. Consequently, with a placebo pair, consumers would choose a product without detailed examination and with no mechanism to cause the response bias to be rendered inoperative. Accordingly, as previously demonstrated (5,7,8), only a minority of consumers would choose the *No Preference* option.

On the other hand, if a preference test protocol was disrupted so that the protocol did not resemble a regular paired preference task, the novelty would be different enough for the brain to use central processing rather than automatically switching to a binary choice subroutine. A protocol based on this hypothesis is called a Disruptive Protocol. Because the choice task is referred to central processing,

Table 2. Examples of fast thinking responses provided by a cognitive subroutine for stereotypical questions

Question	Fast thinking answer provided by a cognitive subroutine	Correct answer that would probably be provided by central processing
1. A farmer has 17 sheep. All but 9 die. How many are left?	8 sheep	9 are left
2. It takes 5 machines 5 min to make 5 puddings. How long does it take 100 of the same machines to make 100 puddings?	100 min	5 min
In a lake is a patch of weeds. Every day the patch doubles in size. It takes 46 days to cover the whole lake. How long did it take to cover half the lake?	26 days	45 days

Table 3. Examples of alternative question substitution by a cognitive subroutine under fast thinking when a real question is not answered

Actual question asked	Question substituted by a cognitive subroutine	Honest answer that would have been given if the consumer had been thinking carefully using central processing
1. How excited did you feel when you saw this toothpaste packet? Please use the rating scale	How much did I like the toothpaste packet? I have never been excited by a toothpaste packet	
2. How polite did you feel when you ate this cookie? Please use the rating scale	How much did I like the cookie?	Cookies never made me feel polite
3. Why did you prefer this food?	What attribute did I notice?	I cannot describe why

Fig. 5. Diagram showing how the proportion of *No Preference* responses for a placebo pair can vary depending on whether a binary choice subroutine that elicits response bias is engaged or not engaged.

the binary choice subroutine would be inoperative and response bias would be eliminated. Consequently, consumers would probably pay proper attention to a task. Accordingly, this possibility is called the Attention Hypothesis. Consumers who pay proper attention would be stimulated more by sensory attributes of products under consideration than by extraneous factors of a response bias. Consequently, reacting to the fact that a placebo pair was identical, consumers would be more likely to choose the *No Preference* option (Fig. 5). Therefore, whether consumers were reacting to a response bias would be indicated by a choice of preference or no preference for a placebo pair. If a consumer responded with no preference, this would indicate that the response bias was inoperative and a target pair, if presented immediately before or after the placebo, was probably being assessed using central processing. Based on Kahneman's (44) approach, the consumer would be using slow rather than fast thinking.

Using this approach, a disruptive protocol was used by Xia *et al.* (24) and compared with the regular protocol of presentation of a placebo and a test pair with standard response options of indicating a preference or no preference. The new protocol was called the Reversed Hidden Demand Characteristics (RHDC) protocol. The protocol resembled the regular protocol using a placebo and a test pair, but with different instructions. Consumers were given a placebo pair and asked to indicate which of the pair was preferred or whether there was no preference. As the consumer did this, the experimenter caused a disruption by telling consumers that she knew that they would like the 2 stimuli just about the same because everybody else did. Many consumers reported no preference. Consumers who did report a preference were asked by the experimenter whether the 2 stimuli were very different and consumers reported that they were not. Accordingly, the experimenter asked whether that meant they

actually liked them about the same and many consumers agreed. There were some consumers who still did not agree and it was for these individuals that the attempt to render the response bias inoperative had failed. However, with this protocol, the proportion of consumers with an inoperative response bias and who were able to assess a target pair ranged from 80-90%. For a substantial majority of consumers, the protocol rendered the response bias that is usually induced in regular preference testing to be inoperative. With an inoperative response bias, a truer assessment of products in a target pair could then be achieved. However, it was also possible that the RHDC protocol did not simply render the response bias inoperative, but replaced it with a bias to always choose the *No Preference* option. Had this been so, there would have been many no preferences for the target pair of stimuli. This did not happen.

A second approach was used by Calderón *et al.* (45) in triadic preference testing that disrupted the usual procedure by presenting a consumer with 3 stimuli rather than 2. This consisted of a placebo pair and a different stimulus for comparison. Consumers were given 4 response options. They could like 2 stimuli about the same (this must be the placebo pair), while they could like the third stimulus either more or less. They could also like all 3 stimuli about the same. Consumers who gave any of these 3 responses would have reported liking the placebo pair about the same and, consequently, would have an inoperative response bias so that preference assessments for the target pair could be treated as unbiased data. Consumers who chose the fourth possible response option of liking each stimulus differently would have indicated a preference for the placebo pair and, consequently, would have had an operative response bias at the time of testing. With this protocol, 76.5 - 94% of consumers had an inoperative response bias. Based on these preliminary experiments, a placebo pair can be used as an

instrument for indicating when a response bias was inoperative and can be used for sorting consumers who are and are not responding with a response bias.

Re-assessment of disruptive preference test protocols The disruptive protocols described above were designed using the Attention Hypothesis as a guide based on subjective reports of consumers (21,23). Nevertheless, the action of disruptive protocols can be explained according to the Input Hypothesis in which the problem was that a tau-criterion suitable for difference testing was used when assessing products for preference. For predicting operational preferences, the tau-criterion would need to be larger. Triadic preference testing (45) provides an opportunity for this adjustment to be made because the consumer knowingly experiences both the degree of difference within the placebo and target pairs during testing. A similar explanation can be used for the RHDC protocol (24). Telling consumers that everyone reports no preference for a placebo pair probably encourages consumers to adjust the tau-criterion so that placebo stimuli are not considered different enough to illicit a preference. Therefore, both hypotheses can be used to explain how disruptive testing works. Which hypothesis is correct or whether both are correct is a matter for future research.

The question becomes what is the advantage of having more *No Preference* responses for a placebo pair? For statistical analysis, the greater the difference between response frequencies for target and placebo pairs, the more powerful the test. A placebo pair with responses concentrated in the *No Preference* category is more likely to be different from a target pair, which usually has responses concentrated in preference categories.

A further point concerns use made of a placebo pair by Calderón *et al.* (45) and Xia *et al.* (24) where a placebo pair was used as a means of separating consumers using central processing from consumers merely using a familiar binary choice subroutine when required to give a preference when a preference was expected. Because binary choice subroutine-using consumers demonstrated that they were not paying proper attention to the sensory characteristics of products being assessed, data derived from them could be regarded as not relevant to the measurement at hand and could be eliminated. This is controversial because, for data analysis, sensory testing professionals do not like to exclude data. Yet, if target consumers were females and some males happened to be sampled as well, exclusion of male-derived data because it was not relevant to the measurement at hand, would appear justified. In the same way, it was argued that data derived from consumers who had demonstrated inattention to the sensory attributes of products being tested could be considered as irrelevant and, therefore, excluded.

The weakness of this argument stems from the fact that the purpose for development of disruptive protocols (24,45) was to devise a test that was no longer than a regular protocol using both placebo and target pairs. Thus, judgment of whether consumers

were using a biased sub-routine was made using a single tasting of a placebo pair. Nevertheless, a single test is probably not a large enough behavioral sample for making this decision, especially if it leads to elimination of data. Response bias may have been operating at the time of testing, but may usually be inoperative. Thus, the consumer would have been incorrectly categorized. The number of tests to apply is an experimental question.

The same argument applies using the Input Hypothesis. Thurstonian arguments stress the variability of sensory input so that sometimes input may fall within the tau-criterion and sometimes outside. A single preference test is a poor predictor of operational preferences because first impressions are poor predictors of final liking and choice (46,47).

A final question concerns whether any consumer-derived data should be eliminated, as with disruptive testing (24,45). If the goal of preference testing is merely to predict test preferences in future tests, data would not be eliminated. Conditions and biases occurring in these tests should remain constant. However, the situation is different for prediction of operational preferences.

Elimination of data would appear to depend on whether the Attention Hypothesis or the Input Hypothesis was followed. With the Attention Hypothesis, there are arguments for elimination of consumer-derived data for individuals who do not pay sufficient attention. For the Input Hypothesis, there would not seem to be a strong argument for elimination of data for consumers who report preferences for a placebo pair as these are simply consumers with an overly sensitive tau-criterion. This would merely be accepted as one of the many characteristics of the consumer, which may or may not change when confronted with the target pair. It would not provide grounds for exclusion.

Given different conclusions, future research should investigate which hypothesis is true or whether both are operative. Until then, it is wise to analyze data both from consumers who do and do not report preferences for a placebo pair separately and decide how to treat these data after inspection.

Finally, it is important to note that the purpose of testing is prediction of behavior and, as such, procedures require validation. This will require consumer follow-up after testing to determine whether the testing protocol predicted consumer behavior correctly. At present, research involving development of testing protocols that are efficient and bias free and appropriate for validity testing is underway.

Disclosure The authors declare no conflict of interest.

References

1. Lawless HT, Heymann H. Sensory Evaluation of Food-Principles and Practices. Springer, New York, NY, USA. p. 596 (2010)
2. Resurreccion AVA. Consumer Sensory Testing for Product Development. Aspen Publishers, Inc., Chicago, IL, USA. p. 254 (1998)

3. Stone H, Sidel JL. *Sensory Evaluation Practices*. 3rd ed. Elsevier, Oxford, UK. p. 377 (2004)
4. Wichchukit S, O'Mahony M. Paired preference tests: 'Liking', 'Buying' and 'Take Away' preferences. *Food Qual. Prefer.* 21: 925-929 (2010)
5. Angulo O, Okayama T, Nakamura T, Yuen R, O'Mahony M. Use of purchase preference options to increase 'no preference' frequencies in placebo preference tests. *J. Sens. Stud.* 24: 258-268 (2009)
6. Wichchukit S, O'Mahony M. 'Liking', 'Buying', 'Choosing' and 'Take Away' preference tests for varying degrees of hedonic disparity. *Food Qual. Prefer.* 22: 60-65 (2011)
7. Weiss BH, O'Mahony M, Wichchukit S. Various paired preference tests: Experimenter effect on 'Take Home' choice. *J. Sens. Stud.* 25: 778-791 (2010)
8. Sung YE, Lee H-S, O'Mahony M, Kim K-O. Paired preference tests: Use of placebo stimuli with liking and buying preferences. *J. Sens. Stud.* 26: 106-117 (2011)
9. Calderón E, Angulo O, O'Mahony M, Wichchukit S. Liking and Take Away preferences for Mexican consumers: Cross-cultural comparison with Thais for psychological style. *J. Sens. Stud.* 30: 77-84 (2015)
10. Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. John Wiley & Sons, New York, NY, USA. p. 479 (1966)
11. Lee H-S, O'Mahony M. Sensory difference testing: Thurstonian models. *Food Sci. Biotechnol.* 13: 841-847 (2004)
12. Wichchukit S, O'Mahony M. A transfer of technology from engineering: Use of ROC curves from signal detection theory to investigate information processing in the brain during sensory difference testing. *J. Food Sci.* 75: R183-R193 (2010)
13. Ennis JM, Ennis DM. A comparison of three commonly used methods for treating no preference votes. *J. Sens. Stud.* 27: 123-129 (2012)
14. Angulo O, O'Mahony M. The paired preference test and the 'No Preference' option: Was Odesky correct? *Food Qual. Prefer.* 16: 425-434 (2005)
15. Alfaro-Rodriguez H, O'Mahony M, Angulo O. Paired preference tests: *d'* values from Mexican consumers with various response options. *J. Sens. Stud.* 20: 275-281 (2005)
16. Ennis DM, Collins J. The distinction between discrimination and splitting in paired testing. Report # 80-233, Philip Morris Research Center, Richmond, VA, USA. p. 50 (1980)
17. Alfaro-Rodriguez H, Angulo O, O'Mahony M. Be your own placebo: A double paired preference test approach for establishing expected frequencies. *Food Qual. Prefer.* 18: 353-361 (2007)
18. Alfaro-Rodriguez H, Angulo O, O'Mahony M. Paired preference tests: '50:50' and 'alternating' no preferences. *J. Sens. Stud.* 23: 765-779 (2008)
19. Alvarez-Coureaux Y, Aguilar P, O'Mahony M, Angulo O. Assessment of preference with controls for response bias operating in the test situation: A practical example using omega-3 enriched wholegrain breads with Ecuadorian consumers. *J. Sens. Stud.* 25: 659-671 (2010)
20. Chapman KW, Lawless HT. Sources of error and the no-preference option in dairy product testing. *J. Sens. Stud.* 20: 454-468 (2005)
21. Chapman KW, Grace-Martin K, Lawless HT. Expectations and stability of preference choice. *J. Sens. Stud.* 21: 441-455 (2006)
22. Kim H-S, Lee H-S, O'Mahony M, Kim K-O. Paired preference tests using placebo pairs and different response options for chips, orange juices and cookies. *J. Sens. Stud.* 23: 417-438 (2008)
23. Marchisano C, Lim J, Cho H-S, Suh D-S, Jeon S-Y, Kim K-O, O'Mahony M. Consumers report preference when they should not: A cross-cultural study. *J. Sens. Stud.* 18: 847-856 (2003)
24. Xia Y, Rivera-Quintero A, Calderón E, Zhong F, O'Mahony M. Paired preference tests with reversed hidden demand characteristics. *J. Sens. Stud.* 29: 149-158 (2014)
25. Ennis DM. Relative power of difference testing methods in sensory evaluation. *Food Technol.* 44: 114-118 (1990)
26. Ennis DM. The power of sensory discrimination methods. *J. Sens. Stud.* 8: 353-370 (1993)
27. Frijters JER. The paradox of discriminatory nondiscriminators resolved. *Chem. Sens. Flav.* 4: 355-358 (1979)
28. O'Mahony M, Rousseau B. Discrimination testing: a few ideas, old and new. *Food Qual. Prefer.* 14: 157-164 (2002)
29. O'Mahony M, Masuoka S, Ishii R. A theoretical note on difference tests: Models, paradoxes and cognitive strategies. *J. Sens. Stud.* 9: 247-272 (1994)
30. O'Mahony M, Odert N. A comparison of different sensory difference testing procedures: Sequential sensitivity analysis and aspects of taste adaptation. *J. Food Sci.* 50: 1055-1058 (1985)
31. O'Mahony M, Goldstein L. Tasting successive salt and water stimuli: the roles of adaptation, variability in physical signal strength, learning, supra- and subadapting signal detectability. *Chem. Sens.* 12: 425-436 (1987)
32. Kim H-J, Jeon S-Y, Kim K-O, O'Mahony M. Thurstonian models and variance I: Experimental confirmation of cognitive strategies for difference tests and effects of perceptual variance. *J. Sens. Stud.* 21: 465-484 (2006)
33. Santosa M, O'Mahony M. Sequential sensitivity analysis for same-different tests: Some further insights. *J. Sens. Stud.* 23: 267-283 (2008)
34. Prescott J, Norris L, Kunst M, Kim S. Estimating a "consumer rejection threshold" for cork taint in white wine. *Food Qual. Prefer.* 16: 345-349 (2005)
35. Lee Y-M, Prescott J, Kim K-O. PROP taster status and the rejection of foods with added tastants. *Food Sci. Biotechnol.* 17: 1066-1073 (2008)
36. Saliba AJ, Bullock J, Hardie WJ. Consumer rejection threshold for 1,8-cineole (eucalyptol) in Australian red wine. *Food Qual. Prefer.* 20: 500-504 (2009)
37. Ennis DM, Ennis JM. Accounting for no difference/preference responses or ties in choice experiments. *Food Qual. Prefer.* 23: 13-17 (2012)
38. Chapman KW, Lovelace E, Cardello A, Lawless HT. Preference for one of two identical stimuli: Expectations, explicit instructions and personal traits. *J. Sens. Stud.* 25(Supplement s.1): 35-53 (2010)
39. Wilke KD, Cochran C-YC, Chambers E. Multiple preference tests can provide more information on consumer preferences. *J. Sens. Stud.* 21: 612-625 (2006)
40. Christensen BBH, Lee H-S, Brockhoff PB. Estimation of the thurstonian model for the 2-AC protocol. *Food Qual. Prefer.* 24: 119-128 (2012)
41. Jesionka V, Rousseau B, Ennis JM. Transitioning from proportion of discriminators to a more meaningful measure of sensory difference. *Food Qual. Prefer.* 32: 77-82 (2014)
42. Christensen RHB, Ennis JM, Ennis DM, Brockhoff PB. Paired preference data with a no-preference option - Statistical tests for comparison with placebo data. *Food Qual. Prefer.* 32: 48-55 (2014)
43. Bi J, Lee H-S, O'Mahony M. A Thurstonian model and statistical inference for the 2-Alternative choice test with both test pairs and placebo pairs. *J. Sens. Stud.* 30: 10-20 (2014)
44. Kahneman D. *Thinking Fast and Slow*. Farrar, Straus & Giroux, New York, NY, USA. p. 499 (2011)
45. Calderón E, Rivera-Quintero A, Xia Y, Angulo O, O'Mahony M. The triadic preference test. *Food Qual. Prefer.* 39: 8-15 (2015)
46. Köster EP, Couronne T, Léon F, Marcelino AS. Repeatability in hedonic sensory measurement: A conceptual exploration. *Food Qual. Prefer.* 14: 165-176 (2002)
47. Kremer S, Shimojo R, Holthuysen N, Köster EP, Mojet J. Consumer acceptance of salt-reduced "soy sauce" bread over repeated in home consumption. *Food Qual. Prefer.* 28: 484-491 (2013)