Food Quality and Preference 30 (2013) 242–249

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

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

Investigation of bias of hedonic scores when co-eliciting product attribute information using CATA questions

Sara R. Jaeger [a,*], Davide Giacalone [c], Christina M. Roigard [a], Benedicte Pineau [a], Leticia Vidal [b], Ana Giménez [b], Michael B. Frøst [c], Gaston Ares [b]

[a] The New Zealand Institute for Plant and Food Research Ltd., Mt Albert Research Centre, Private Bag 92169, Victoria Street West, Auckland, New Zealand
[b] Departamento de Ciencia y Tecnología de Alimentos, Facultad de Química, Universidad de la República, Gral. Flores 2124, CP 11800 Montevideo, Uruguay
[c] Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, DK-1858 Frederiksberg C, Denmark

ARTICLE INFO

Article history:
Received 12 April 2013
Received in revised form 7 June 2013
Accepted 7 June 2013
Available online 14 June 2013

Keywords:
Hedonic scaling
CATA
JAR
Attribute scaling
Research methodology
Consumer research

ABSTRACT

Sensory and consumer scientists disagree on the practice of concurrently obtaining sensory information in hedonic tests. This is in part due to different mindsets about what consumers are able to do and evidence that such co-elicitation may bias hedonic scores. Check-all-that-apply (CATA) questions have been claimed to have a smaller effect on hedonic scores than other attribute such as just-about-right or intensity scales. In this research, nine studies using consumers as participants examined effects on hedonic product scores when sensory attribute information was co-elicited using CATA questions. The use of CATA concurrently with hedonic was benchmarked against concurrent attribute liking scores, attribute intensity scores and just-about-right scaling. Across a range of product categories (beer, fresh fruit, tea, flavoured water, crackers, savoury dips), only weak and transient evidence of bias of hedonic scores when concurrently using CATA questions was established. This effect was independent on whether samples, on average were moderately liked or moderately disliked, and replicated when samples were assessed partially by the sense of smell only or via full product assessment (appearance, aroma, flavour, taste, aftertaste, mouthfeel). The present research suggests that co-elicitation of hedonic scores and product attribute information using CATA questions may bias the hedonic scores, but not that it certainly will do so. This needs to be recognised, leading to more widespread acceptance that co-elicitation has merit. Investigators should decide on whether or not to co-elicit product attribute information using CATA questions on a case-by-case basis, acknowledging that bias may occur. Further research is needed to understand when/when not bias is likely to occur.

1. Introduction

Quantitative consumer research is often aimed at determining consumers' hedonic reaction to the sensory characteristics of products (Lawless & Heymann, 2010). Consumers are asked to sample a set of products and to indicate how much they like them using a hedonic scale (Lim, 2011). In some instances, hedonic information is supplemented with questions on specific sensory characteristics with the aim of understanding consumer preferences and identifying recommendations for product reformulation (Stone & Sidel, 2004).

One of the major concerns of including questions about specific sensory characteristics is that they can be a source of bias on hedonic scores (Stone & Sidel, 2004). According to Prescott, Lee, and Kim (2011) asking consumers to complete analytical tasks can hinder the utilisation of hedonic information. Based on previous research (Prescott, 1999, 2004; Small & Prescott, 2005), these authors argue that directing consumer attention to multiple attributes may inhibit a cognitive representation of synthetic characteristics, such as overall liking. Besides, survey research has shown that previous questions have the potential to alter a person's perception of a product by making certain aspects more salient and relevant (Strack, 1992).

Studies on the influence of analytical tasks on hedonic ratings have shown contradictory results and scholars are divided on the topic (Moskowitz, Munoz, & Gacula, 2003). Asking consumers to evaluate sensory attributes using intensity scales, attribute liking questions or just-about-right (JAR) scales have been reported to significantly affect hedonic scores and to affect conclusions regarding consumers' preference patterns (Earthy, MacFie, & Hedderley, 1997; Popper, Rosenstock, Schraidt, & Kroll, 2004; Prescott et al., 2011). However, some authors have reported no effect (Gacula, Mohan, Faller, Pollack, & Moskowitz, 2008; Mela, 1989; Vickers, Christensen, Fahrenholtz, & Gengler, 1993). In summary, published studies suggest that the influence of analytical tasks on hedonic

* Corresponding author. Tel.: +64 9 925 7000.
E-mail address: sara.jaeger@plantandfood.co.nz (S.R. Jaeger).

0950-3293/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2013.06.001

scores is methodology and sample dependent, and that it does not always occur.

Check-all-that-apply (CATA) questions are an increasingly popular technique to collect analytical consumer evaluations of food products (Adams, Williams, Lancaster, & Foley, 2007; Ares, Barreiro, Deliza, Giménez, & Gámbaro, 2010; Ares & Jaeger, 2013; Dooley, Lee, & Meullenet, 2010; Plaehn, 2012). The CATA method consists of presenting consumers with a predefined list of terms from which consumers should tick all they find appropriate to describe a focal product. This methodology has been reported to be quick and easy to use for participants (Ares et al., 2010; Driesener & Romaniuk, 2006), providing similar results than trained assessors panels using Quantitative Descriptive Analysis (Ares et al., 2010; Bruzzone, Ares, & Giménez, 2012; Dooley et al., 2010).

Furthermore, it has been proposed that that, when used together with hedonic scales, CATA questions have a smaller effect than other attribute-based question types such as just-about-right or intensity scales (Adams et al., 2007). Tentatively, this may be explained by the fact that CATA questions are not thought to be cognitively demanding and do not encourage respondents to engage in deep processing (Krosnick, 1999; Sudman & Bradburn, 1992). Nevertheless, no published study has yet provided empirical evidence to support the claim by Adams et al. (2007).

In view of the growing practice in academia and industry of using consumers for concurrent attribute assessment (Varela & Ares, 2012), there is an increasing need for studies investigating possible biases of CATA questions on consumer hedonic ratings. The present paper explores the influence of CATA questions on hedonic ratings through nine consumer studies. First, it benchmarks the use of CATA concurrently with hedonic against concurrently obtained attribute liking scores, attribute intensity scores and just-about-right scaling (Studies 1–3). In the second part of the manuscript the degree of biasing effect of different conditions of concurrent hedonic and CATA elicitation is examined (Studies 4–8). The studies were conducted in multiple countries using multiple product categories. Between-subjects designs were used in all studies (see Tables 1 and 2 for an overview), and participants took part in a single research session only.

2. Methodology

2.1. Study 1: liking only vs. liking + intensity ratings vs. liking + CATA (flavoured water)

Study 1 compared concurrent use of CATA questions and questions about attribute intensity ratings with regards to producing a bias in liking ratings.

2.1.1. Participants

Consumers (N = 190, 63% female) completed Study 1 as part of a larger study on sensory odour acuity and food preferences. They were aged 18–60 years old and self-identified as Caucasian. All participants lived in Auckland, New Zealand. Participants gave voluntary written informed consent and were compensated in cash. According to chi-square test, across the three experimental treatments (A, B and C), consumers were balanced with respect to age (χ^2 = 3.86, p = 0.42) and gender (χ^2 = 5.97, p = 0.051), in order to infer that possible differences could be ascribed to the experimental design.

2.1.2. Samples

Consumer responses were collected in response to the aroma of a non-cyclic sulphur-containing carbon-based compound (99% purity). Two samples (2431 and 694 ppb) were prepared by diluting in water a stock solution of the compound. Consumers received 10 ml of each sample in wine glass covered by watch glassed and labelled with a 3-digit code. Samples were poured 1–2 h prior being presented to consumers.

2.1.3. Experimental design, CATA lexicon and data collection

Participants attended a single research session, and a between-subjects design was used. In Treatment A, consumers were asked to smell the flavoured solutions and rate overall liking for the two samples using a labelled 9-pt hedonic scale where the endpoint anchors were 1 = 'dislike extremely' and 9 = 'like extremely'. In Treatment B, consumers first rated overall liking, then rated the intensity of the following four attributes: *mussel*, *seafood*, *canned/cooked vegetable*, *seaside/marine*. Ratings were collected on a category scale with anchors 0 = 'I cannot smell anything'; 1 = 'Extremely weak'; and 9 = 'Extremely strong'). In Treatment C consumers evaluated liking first, and then completed a CATA task with the following 15 terms: *cabbage*, *cooked vegetables*, *canned asparagus*, *heavy*, *light*, *mussel*, *salty*, *seafood*, *seaweed*, *sickening*, *sulphuric*, *seaside*, *sweet vomit*, *smoked fish*, *steamed fish*. These terms were developed in pilot work with Plant & Food Research staff. Sample presentation order was counter-balanced. Data collection took place in standard sensory booths under controlled temperature and airflow conditions. Artificial white lighting was used.

The participants in this study also completed Studies 2 and 6–8. The allocation of participants to experimental treatments was such that a person always completed Treatment A or Treatment B or Treatment C (see Tables 1 and 2). Treatment A was always the 'Hedonic only' treatment and the participants in this group were at no time presented with a task where attribute information was concurrently elicited with hedonic scores. Conversely, the participants who always completed Treatments B or C were always providing hedonic and attribute specific information about the focal samples. We used this allocation of participants to experimental treatments to retain participants in 'stable mindsets'.

2.2. Study 2: liking only vs. liking + attribute liking vs. liking + CATA (salmon dip)

Study 2 compared the use of CATA questions to the use of attribute liking questions concurrent with hedonic scaling.

Table 1
Overview of studies in Part 1 and summary of results.

Study ID	Samples summary	Experimental treatment	Results summary from mixed linear model	
Study 1	Flavoured water	(A) Hedonic only	$F_{Exp.Tr.}$ = 0.64	p = 0.20
	2 Samples	(B) Hedonic + attribute intensity	F_{Sample} = 0.95	p = 0.33
	Aroma only assessment	(C) Hedonic + attribute CATA	$F_{Exp.Tr. \times Sample}$ = 0.01	p = 0.99
Study 2	Salmon dip	(A) Hedonic only	$F_{Exp.Tr.}$ = 0.07	p = 0.93
	1 Sample	(B) Hedonic + attribute liking	F_{Sample} is N/A	N/A
	Aroma only assessment	(C) Hedonic + attribute CATA	$F_{Exp.Tr.\times Sample}$ is N/A	N/A
Study 3	Strawberries	(A) Hedonic + attribute JAR	$F_{Exp.G.}$ = 0.25	p = 0.61
	6 Sample	(B) Hedonic + attribute CATA	F_{Sample} = *14.04*	p < *0.0001*
	Full sample assessment		$F_{Exp.Tr.\times Sample}$ = 0.51	p = 0.46

Table 2
Overview of studies in Part 2 and summary of results.

Study ID	Samples summary	Experimental treatment	Results summary of linear mixed model	
Study 4	Beer	(A) Hedonic only	$F_{Exp.Tr.} = 3.58$	$p = 0.02$
	4 Samples	(B) Hedonic + CATA	$F_{Sample} = 3.28$	$p = 0.02$
	Full sample assessment	(C) CATA + Hedonic	$F_{ExpTr \times Sample} = 0.87$	$p = 0.51$
Study 5	Rice crackers (flavoured)	(A) Hedonic only	$F_{Exp.Tr.} = 3.91$	$p = 0.02$
	6 Samples	(B) Hedonic + CATA term order 1	$F_{Sample} = 11.41$	$p < 0.0001$
	Full sample assessment	(C) Hedonic + CATA term order 2	$F_{Exp.Tr. \times Sample} = 1.44$	$p = 0.16$
Study 6	Green tea (unsweetened)	(A) Hedonic only	$F_{Exp.Tr.} = 0.60$	$p = 0.56$
	2 Samples	(B) Hedonic + delayed CATA	$F_{Sample} = 10.24$	$p = 0.0015$
	Full sample assessment	(C) Hedonic + CATA	$F_{Exp.Tr. \times Sample} = 1.30$	$p = 0.27$
Study 7	Seaweed rice cracker	(A) Hedonic only	$F_{Exp.Gr.} = 0.28$	$p = 0.76$
	1 Sample	(B) Hedonic + CATA (28 terms)	F_{Sample} is N/A	N/A
	Full sample assessment	(C) Hedonic + CATA (7 terms)	$F_{Exp.Tr. \times Sample}$ is N/A	N/A
Study 8	Seafood dip	(A) Hedonic only	$F_{Exp.Tr.} = 0.84$	$p = 0.44$
	1 Sample	(B) Hedonic + CATA (4Q)	F_{Sample} is N/A	N/A
	Full sample assessment	(C) Hedonic + CATA (1Q)	$F_{Exp.Tr. \times Sample}$ is N/A	N/A
Study 9	Crackers (plain)	(A) Hedonic + CATA (1Q)	$F_{Exp.Gr.} = 0.84$	$p = 0.44$
	3 Samples	(B) Hedonic + CATA (3QM)	$F_{Sample} = 2.74$	$p = 0.07$
	Full sample assessment	(C) Hedonic + CATA (3QS)	$F_{Exp.Tr. \times Sample} = 0.51$	$p = 0.73$

Notes: in study 9 experimental treatment B used 3 CATA questions of mixed modality (3QM), whereas experimental treatment C used 3 CATA questions of separate sensory modality (3QS). The terms were unchanged and identical to the 21 terms used in experimental treatment A.

2.2.1. Participants

Participants were the same group of individuals who took part in Study 1 of this research.

2.2.2. Samples

A teaspoon of a commercially available salmon dip (*Turkish Kitchen (Auckland, New Zealand), "Flavour it dip", Smoked salmon & dill with cashew™*), served in 30 ml custard cups and labelled with a 3-digits code, was used as test stimulus.

2.2.3. Experimental design, CATA lexicon and data collection

Participants attended a single research session, and a between-subjects design was used. Treatment A served as control group in which consumers were instructed to smell the sample and subsequently rate overall liking for the two samples using a 9-pt hedonic scale (see Study 1). In Treatment B, consumers first rated overall liking, then liking for four attributes (two related to appearance and two to aroma): *colour intensity*, *visual amount of dill*, *salmon aroma*, *smoky aroma*, using a 9-pt hedonic scale identical to that in Study 1. In Treatment C consumers evaluated liking first, and then completed the CATA task in the same ballot. The CATA terms were organised within two columns: Aroma (containing the terms *cooked fish*, *dill/herb*, *salmon*, *smoky*, *sweet*, and *tangy*) and Appearance (*Artificial*, *oily*, *salmon colour*: *strong*, *salmon colour*: *weak*, *small lumps*, *smooth*). These terms were developed through pilot work with Plant & Food Research staff.

2.3. Study 3: liking + jar vs. liking + CATA (strawberries)

Study 3 compared the use of CATA questions to just-about-right scales (JAR) when implemented concurrently with hedonic scaling.

2.3.1. Participants

Consumers were randomly recruited among shoppers in a supermarket in Montevideo (Uruguay) based on their stated strawberry consumption (at least once during the week before the study) and their interest in participating (cash incentives were not used). The 120 consumers who took part (58% female: 42% male; aged 18–75 years old), were randomly divided in two groups of 60 people: Treatment A and Treatment B. Consumers in both treatments completed the study in the same session. Consumers did not differ in key demographics across treatments (Gender: $\chi^2 = 1.98$, $p = 0.16$, Age: $\chi^2 = 0.80$, $p = 0.37$).

2.3.2. Samples

Six strawberries cultivars were used, of which three were commercially available in Uruguay (Yurí, Yvahé and Guenoa) and three were promising new cultivars developed by Instituto Nacional de Investigación Agropecuaria (INIA, Uruguay) (L20.1, L.53.3 and K50.5). All cultivars were grown under the same environmental and management conditions. Fruits were hand harvested from plants growing in a commercial greenhouse in Salto (Uruguay). Only mature, fully coloured unblemished fruit that felt firm to the touch were picked and then immediately packed into plastic, vented, lidded containers and stored under air at 2 °C for 1 day prior to the consumer study. For each cultivar, two strawberries were presented to the consumers in closed plastic containers labelled with 3-digit random numbers, at room temperature. Samples were presented monadically according to a balanced design (Latin square design) and mineral water was available for rinsing between samples.

2.3.3. Experimental design, CATA lexicon and data collection

A between-subjects design was used. In Treatment A, consumers had to try the strawberries, rate their overall liking using a 9-pt hedonic scale (same anchors as Study 1) and then evaluate six sensory attributes using 5-pt just-about-right scales (1 = 'not enough', 3 = 'just about right', 5 = 'too much'): *red colour*, *sweetness*, *sourness*, *strawberry odour*, *strawberry flavour*, and *hardness*. In Treatment B consumers evaluated liking first, and then completed the CATA task with the following 18 terms: *not much red colour*, *moderate red colour*, *intense red colour*, *not very sweet*, *sweet*, *too sweet*, *not very sour*, *sour*, *very sour*, *no strawberry odour*, *moderate strawberry odour*, *intense strawberry odour*, *no strawberry flavour*, *moderate strawberry flavour*, *intense strawberry flavor*, *soft*, *firm*, and *hard*. These terms were based on published data (Péneau, Brockhoff, Escher, & Nuessli, 2007) and previous studies with trained assessors and consumers (Ares, Barrios, Lareo, & Lema, 2009; Lado, Vicente, Manzzioni, & Ares, 2010) and known from a previous study to be the main attributes responsible for differences in the sensory characteristics of strawberry cultivars in Uruguay (Lado et al., 2010). Although the terms 'too much strawberry odour' and 'too much strawberry flavour' are not common in JAR scales, they were included in the study because some consumers tended to use them in previous studies. Data collection took place in standard sensory booths under controlled temperature and airflow conditions.

2.4. Study 4: liking only vs. liking + CATA yes/no vs. CATA yes/no + liking (beer)

Study 4 tested whether the presence of CATA items on the ballot affects hedonic scores, and specifically what is the impact of elicitation of CATA responses prior to vs. subsequent to the hedonic question.

2.4.1. Participants

Participants (N = 129, 44% female) were a group of university students (all aged 18–30) at the Faculty of Science, University of Copenhagen, Denmark. Participants did not receive an incentive for participation.

2.4.2. Samples

Four commercially available beers, all brewed by the same commercial brewery, were used as stimuli. They were selected to provide a sufficient span with regards to style and sensory characteristics (*Harboe Pilsner*©, *Harboe Classic*©, *Harboe Juleøl*©, and *Harboe Bjørnebryg*© (Skælskør, Denmark) – indicated respectively as samples A, B, C, D). The beers were stored at room temperature prior to serving. Evaluations took place in a central hallway outside a canteen of the Faculty of Science. Each consumer was monadically served about 5 cl of each beer sample in a plastic cup. Samples were labelled with 3-digit numbers, and served in a randomised order using a Latin square design.

2.4.3. Experimental design, CATA lexicon and data collection

A between-subjects design was used. In Treatment A consumers (n = 42) only gave hedonic ratings. In Treatment B (n = 46) consumers gave hedonic ratings for the same beers, and concurrently completed a CATA questionnaire (on a beer-by-beer basis). Consumers in Treatment C (n = 41) performed the same task as consumers in Treatment B, but with an inverted question order, such that on their ballot the CATA questions appeared before the hedonic rating. For all groups, hedonic responses were elicited with the question "How much do you like this beer?", and rated on a 9-pt hedonic scale with anchors 1 = 'Not at all', 5 = 'Neutral' and 9 = 'Like very much'. The list of CATA descriptors listed 10 terms: *transparent*, *sparkling*, *yellow*, *rye-bread*, *flowery*, *sweet*, *malt*, *acidic*, *alcohol*, and *caramel*. These were developed in pilot work with students and staff at the Department of Food Science at the University of Copenhagen. The CATA question was administered with the wording "Do you find this descriptor in the sample?" and answered in a "yes/no" dichotomous response. This approach differs slightly from the classical formulation (check-all-that-apply), and is known to increase the likelihood that respondents go through the whole list (Rasinski, Mingay, & Bradburn, 1994), hereby reducing satisficing response strategies (Krosnick, 1991).

2.5. Study 5: liking only vs. liking + CATA term order 1 vs. liking + CATA term order 2 (flavoured crackers)

Study 5 tested whether the order in which CATA items were listed on the ballot affects hedonic and compared two fixed CATA term orders to hedonic only elicitation.

2.5.1. Participants

Consumers (N = 145, 57% female) completed Study 5 as part of a larger study on sensory odour acuity and food preferences. The recruitment criteria were identical to Study 1, and with the exception of two people, participants did not take part in Study 1. Across the three experimental treatments (A, B, and C), consumers were balanced with respect to age ($\chi^2_2 = 1.16$, $p = 0.56$) and gender ($\chi^2_4 = 1.59$, p = 0.81), in order to infer that possible differences could be described to the experimental design.

2.5.2. Samples

Six commercially available flavoured rice crackers were used (Pams Original, Fantastic Barbecue, Fantastic Chicken, Peckish™ Thins Salt & Vinegar, Trident® Seaweed, and Sakata® Wholegrain Smokey Barbecue. Two crackers were placed in a 60 mL lidded plastic cup and labelled with a 3-digit code.

2.5.3. Experimental design, CATA lexicon, and data collection

A between-subjects design was used. In Treatment A, consumers (n = 56) gave only rated overall liking (9-pt labelled scale where 1 = 'dislike extremely' and 9 = 'like extremely'). In Treatment B, consumers (n = 43) first rated overall liking, and then completed a CATA task with 14 items: vinegar, hard, smoky, salty, visible flavouring, sweet, bland, spicy, savoury, crisp, seeds/grains, seaweed, soy sauce and uneven surface. Consumers in Treatment C (n = 46) completed the Treatment B task but used a CATA ballot where items were listed in a different order: uneven surface, hard, savoury, seeds/grains, crisp, spicy, bland, salty, visible flavouring, smoky, soy sauce, seaweed, sweet and vinegar. The two presentation orders of the terms were randomly selected.

Sample presentation order was balanced according to a Latin square design. Data collection took place in standard sensory booths under controlled temperature and airflow conditions. Artificial white lighting was used. Water was available for rinsing between samples.

2.6. Study 6: hedonic vs. concurrent hedonic/analytic evaluation vs. delayed analytic (green tea)

Study 6 tested whether the presence of CATA items on the ballot has an effect on liking ratings, and also whether the mere expectation of having to perform an analytical task is a sufficient condition to produce a bias of hedonic scores.

2.6.1. Participants

Participants were the same group of individuals who took part in Study 1 of this research.

2.6.2. Samples

Two types of commercially available green teas (*Chanui Fine Leaf Tea*™, and *Dilmah All Natural Green Tea – Pure Green*™) were used as stimuli. The samples looked identical and were served (20 ml, at room temperature) in 60 ml plastic cups with lids, labelled with a 3-digit code.

2.6.3. Experimental design, CATA lexicon and data collection

A between-subjects design was used to compare three experimental treatments. In Treatment A, consumers tasted the green teas and were only asked to rate overall liking for the two samples using a 9 point hedonic scale. In Treatment B, hedonic and analytical evaluations were split in time. Consumers started by rating overall liking for the two samples in a monadic order. Then, they received the same samples one more time and were asked to evaluate them on a separate ballot with nine CATA terms. Prior to hedonic assessment participants were told that they would receive each sample twice and that they the second time would be asked to evaluate them for a set of attributes. These were then read aloud and a CATA only ballot shown. In Treatment C consumers evaluated overall liking first, and then completed the CATA task in the same ballot. The nine sensory terms identical for treatments B and C were selected on the basis of bench top testing with sensory staff at Plant & Food Research. The terms were: *dry*, *floral*, *metallic*, *hay-like*, *off-flavour*, *green tea*, *sweet*, *bitter*, and *grassy/vegetable*.

Sample presentation order was balanced according to an experimental design. Data collection took place in standard sensory booths under controlled temperature and airflow conditions. Arti-

ficial white lighting was used. Water was available for rinsing between samples.

2.7. Study 7: liking only vs. liking + short CATA vs. liking + long CATA (seaweed cracker)

Study 7 sought to investigate whether potential bias of hedonic scores during co-elicitation with CATA questions was dependant on the length of the CATA question.

2.7.1. Participants

Participants were the same group of individuals who took part in Study 1 of this research.

2.7.2. Samples

The sample was a commercially available seaweed cracker (*Fantastic Snacks' Seaweed Flavor Rice Cracker*™), served in 130 ml custard cups and labelled with 3-digit code.

2.7.3. Experimental design, CATA lexicon and data collection

A between-subjects design was used. Consumers in Treatment A served as control group in which tasted the sample and rated liking using a 9-pt hedonic scale (same anchors as Study 1). In Treatment B, consumers were ask to taste the sample, rate overall liking and then complete a CATA task with 28 item: *soy sauce*, *fish*: *weak*, *rice*, *salty*: *low*, *sticky*, *greasy*, *seaweed*: *weak*, *thick*, *dry*, *brittle*, *crunchy*, *sweet*: *high*, *rancid*, *smooth*, *aftertaste*, *toasted colour*, *salty*: *high*, *flavorsome*, *tasteless*, *oiled/fried flavor*, *toasted flavor*, *sweet*: *low*, *crisp*, *fish*: *strong*, *off-flavor*, *seaweed*: *strong*, *thin*, *hard*. Consumers in Treatment C tasted the sample, rated overall liking and completed a CATA task with 7 items: *grassy*, *seaweed*, *crunchy*, *toasted colour*, *flavorsome*, *off-flavor*, *hard*. The CATA terms were generated in pilot work with Plant & Food Research staff.

2.8. Study 8: liking only vs. liking + CATA by modality vs. liking + single CATA (seafood dip)

Study 8 explored whether possible bias of hedonic scores would be linked to implementation of the CATA question, and specifically whether a different effect would be obtained when CATA items are organised by sensory modality as opposed to appearing a single item list.

2.8.1. Participants

Participants were the same group of individuals who took part in Study 1 of this research.

2.8.2. Samples

One sample was used: a seafood dip (*Country Goodness' Seafood Fiesta Flavored sour cream dip*™). Consumers received a teaspoon of the dip in a 130 ml custard cup, labelled with a 3-digits code, together with a water cracker (*Arnott's Original Water Cracker*™).

2.8.3. Experimental design, CATA lexicon and data collection

A between-subjects design was used. Consumers in Treatment A tasted the sample and gave overall liking ratings. In Treatment B, consumers tasted the sample, rated liking and then completed a CATA task in which items where organised under four separate headings: *Appearance* (*white colour*, *smooth*, *lumpy*), *Aroma* (*garlic*, *tangy*, *cheese*), *Taste* (*off-flavour*, *seafood/shellfish*, *chives/parsley*), *Texture/Mouthfeel* (*grainy*, *soft*, *creamy*). Consumers in Treatment C tasted the sample and completed a single-question CATA with 12 items: *garlic*, *creamy*, *tangy*, *parsley/chives*, *off-flavour*, *savoury*, *shellfish*, *flavoursome*, *seafood*, *oily/fatty*, *mussel*, *cheese*. The CATA terms were generated in pilot work with Plant & Food Research staff.

2.9. Study 9: hedonic + 1 CATA question vs. hedonic + 3 mixed sensory modality CATA questions vs. hedonic + 3 separate sensory modality CATA questions (Crackers)

Study 9 explored whether CATA questions of similar length that featured CATA terms from only one or from multiple sensory modalities would lead to different bias on hedonic scores.

2.9.1. Participants

Consumers ($N = 120$), who completed the study as part of a larger study on bakery products, were recruited from the consumer database of Departamento de Ciencia y Tecnología de Alimentos (Montevideo, Uruguay) based on their consumption of bakery products and crackers, as well as their interest and availability to participate in the study. They were aged 18–60 years old (64% female). Participants gave written informed consent and were compensated with a small gift. Consumers were randomly divided into three experimental groups of 40 participants, which completed the study in the same session. Key consumer demographics did not differ across groups (Gender: $\chi^2 = 0.89$, $p = 0.35$, Age: $\chi^2 = 0.14$, $p = 0.71$).

2.9.2. Samples

Three commercially available samples of plain crackers were evaluated. The samples were purchased from local supermarkets. One cracker of each sample was served in a plastic plate labelled with a 3-digit code.

2.9.3. Experimental design, CATA lexicon and data collection

A between-subjects design was used. The experimental design defined three treatments. One group of participants (Treatment A, $n = 40$) rated their overall liking using a 9 point hedonic scale, and then received a single CATA question featuring 21 attributes: *hard*, *toasted colour*, *greasy*, *salty*, *big*, *adhesive*, *dry*, *toasted flavour*, *thin*, *heterogeneous colour*, *crunchy*, *sour*, *tasteless*, *homogeneous colour*, *soft*, *off-flavour*, *thick*, *small*, *aftertaste*, *brittle*, and *oily flavour*. The second group of participants (Treatment B, $n = 40$) rated their overall liking and answered three CATA questions comprising 7 of the 21 terms, each featuring multiple sensory modalities. The third group of participants (Treatment C, $n = 40$) rated their overall liking and answered three CATA questions by modality, as follows: "Check all the terms you consider appropriate to describe the appearance of this cracker", comprising the attributes: *toasted colour*, *big*, *thin*, *heterogeneous colour*, *homogeneous colour*, *thick* and *small*; and the same question for texture (*hard*, *greasy*, *adhesive*, *dry*, *crunchy*, *soft*, *brittle*) and flavour (*salty*, *toasted flavour*, *sour*, *tasteless*, *off-flavour*, *aftertaste* and *oily flavour*). These terms were generated using available literature (Vázquez, Curia, & Hough, 2009) and previous qualitative consumer studies.

Samples were assessed monadically according to a balanced random design (Williams' design). Samples could be tasted more than once. The test took place in standard sensory booths, under white lighting, controlled temperature (23 °C) and airflow conditions.

2.10. Data analysis

For each study, linear mixed modelling was performed to uncover significant differences in hedonic ratings across experimental treatments. For those studies with only one sample (Studies 2, 7 and 8) the linear mixed model included experimental treatment as fixed effect and consumer (within experimental treatment) as random effect. For the other studies (Studies 1, 3–6 and 9), in which more than one sample were considered, the linear mixed model included experimental treatment, sample and their interaction as fixed effects, and consumer (within experimental

treatment) as random effect. A 5% significance level was considered in the analyses. When effects were significant, honestly significant differences were calculated using Tukey's test. All analyses were carried out in R, Version 2.11.1.

3. Results[1]

3.1. Influence of CATA question on hedonic scores and comparison with other attribute scaling methodologies

Table 1 reveals that overall liking scores elicited concurrently with CATA questions did not significantly differ from those elicited concurrently with attribute intensity scales (Study 1), attribute liking questions (Study 2) or just-about-right scales (Study 3). In Studies 1 and 2 the inclusion of attribute intensity scales, attribute liking questions or CATA questions did not lead to a systematic shift on overall liking scores ($p > 0.07$). Besides, in Study 1 CATA questions and attribute liking did not cause a significant change in rank order of samples, as seen by the non-significant interaction effect ($p = 0.99$). For completeness the average liking scores by experimental treatment and product is shown in Table 3.

3.2. Degree of biasing effect of different conditions of concurrent hedonic and CATA elicitation

Part 2 of this research included Studies 4–9 and results regarding the influence of concurrent use of CATA questions on hedonic scores in one instance revealed evidence of bias. Specifically, significant main effects of sample and experimental treatment on hedonic scores were established in Study 5. Pair-wise comparisons revealed that the mean hedonic score across the six rice cracker samples was significantly different between Treatments A and C ($M_{Hedonic\ only} = 6.4$, $M_{Hedonic+CATA\ order\ 2} = 5.9$, $p = 0.0009$), but not between Treatments A and B ($M_{Hedonic\ only} = 6.4$, $M_{Hedonic+CATA\ order1} = 6.3$, $p = 0.44$). The effect on hedonic scores was transient in the sense that with CATA terms listed in one order hedonic bias was observed, whereas bias did not occur when CATA terms were listed in a different order. The interaction between sample and experimental treatment was not significant (Table 2), suggesting that the influence of the position of the CATA ballot did not lead to changes in consumers' preference patterns.

In Study 4 a significant effect on hedonic scores due to experimental treatment was also established. However, the difference was found between the two treatments where CATA was elicited concurrently with overall liking, with the Hedonic After treatment giving significantly lower ratings than the Hedonic First treatment ($M_{\text{Hedonic After}} = 4.47 < M_{\text{Hedonic First}} = 5.11$, adj. $p = 0.017$). No significant differences were found between overall liking scores in the Hedonic Only treatment ($M_{\text{Hedonic Only}} = 4.74$), versus either of the two treatments with concurrent rating of CATA, and this outcome did not change when the two latter groups were combined ($t(514) = 0.33$, $p = 0.74$). Although participants in the Hedonic After treatment did produce lower ratings than those in the Hedonic Only treatment, this difference did not reach statistical significance ($M_{\text{Hedonic After}} = 4.47$, $M_{\text{Hedonic Only}} = 4.74$, $p = 0.24$). As per Study 4, the interaction between sample and experimental treatment was not significant (Table 2), suggesting that the influence of the position of the CATA ballot did not lead to changes in consumers' preference patterns.

Taken together the results from Part 2 of this research revealed that CATA questions only caused weak and transient evidence on co-elicited hedonic scores. The number of CATA questions, the length of the CATA questionnaire and considering mixed modality or single modality CATA questions did not seem to bias hedonic scores. Besides, asking consumers to complete a hedonic task for all samples and then answer the CATA question yielded the same results than asking them to complete one task after another for each sample. The one feature that distinguished Study 5 from the other Studies in Part 2 was the inclusion of a larger number of samples (6 vs. 1–4). Product category or degree of liking of the products did not appear to influence the results.

4. Discussion and conclusions

In Part 1 of this research three studies were conducted to examine whether concurrent use of attribute intensity scales, attribute liking questions or just-about-right scales influenced hedonic scores. We found no evidence of hedonic bias and these results support those reported by Mela (1989), Vickers et al. (1993) and Gacula et al. (2008) when working with intensity scales, attribute liking or just-about-right scales, respectively. Six studies were conducted in Part 2 to examine if concurrent use of CATA questions to obtain sensory product characterizations influenced hedonic scores. The evidence of bias was weak and transient. Bias was established in one of six studies and only when the 14 CATA terms used to characterise the products were listed on one of the two tested orders (Study 5). In Study 4 hedonic bias was not established *per se*, but differences were observed between overall liking scores when CATA questions were asked before and after hedonic scores. Previously, Ares and Jaeger (2013) conducted three consumer studies and found no evidence of concurrent use of sensory CATA questions resulting in bias of hedonic scores.

In their seminal work on CATA questions for sensory product characterization, Adams et al. (2007) stated that CATA questions do not produce a large bias on liking scores. Our results largely support this claim and the suitability of CATA questions for concurrent elicitation of consumers' sensory and hedonic responses to food products. Reasons why bias of hedonic scores are unlikely to occur when CATA questions are used concurrently with the hedonic question may be linked to the characteristics of CATA questions. When completing a CATA question consumers have to check all the terms they consider appropriate for describing the product from a list that contains terms that are both applicable and not applicable to describe it. Therefore, consumers do not need to strongly focus their attention on each of the terms, which could minimize "priming effect" and the activation of information that can become more accessible, even without consumers' awareness, when rating overall liking (Strack, 1992). For this reason, the extent to which consumer attention is directed towards specific attributes when evaluating their overall liking may be minimised. However, we cannot ignore that some evidence of bias was established. The work by Prescott et al. (2011) is in support hereof, and, moreover, imply that bias always be observed if co-elicitation of attribute information is performed. These authors suggested that asking analytical questions prevents consumer from constructing a synthetic representation of the product, which affects overall liking scores.

Overall, our results suggest that concurrent sensory product characterization by CATA questions has the potential to bias hedonic scores. However, it should be taken into account that the observed influence of CATA questions on overall liking scores could have occurred by chance due to the numerous statistical tests performed. Further, we acknowledge possible bias of the results due to learning effects arising from the same group of participants completing five of the reported studies.

[1] It is beyond the scope of this paper to report the product-specific attribute information generated in the focal studies. Interested readers may contact the authors for further details.

Table 3
Mean and standard deviations for liking ratings obtained across the studies. Ratings were collected on a 9-pt hedonic scale (1 = 'dislike extremely' and 9 = 'like extremely'). Treatment A is the control group (liking only) for all studies except Study 9.

Study ID	Sample	Treatment A	Treatment B	Treatment C
Study 1 (Flavoured water)	Sample 1	4.3 ± 2.2	4.7 ± 2.0	4.4 ± 2.0
	Sample 2	4.6 ± 2.0	4.8 ± 1.7	4.7 ± 1.7
Study 2 (Salmon dip)	Sample 1	7.0 ± 1.9	7.1 ± 1.8	7.0 ± 1.3
Study 3 (Strawberries)	Sample 1	6.0 ± 2.5	6.3 ± 2.5	N/A
	Sample 2	5.8 ± 2.5	5.3 ± 2.6	N/A
	Sample 3	5.9 ± 2.3	5.9 ± 2.5	N/A
	Sample 4	4.8 ± 2.6	4.5 ± 2.7	N/A
	Sample 5	5.8 ± 2.4	5.6 ± 2.5	N/A
	Sample 6	5.9 ± 2.2	5.9 ± 2.4	N/A
Study 4 (Beer)	Sample 1	4.7 ± 2.1	5.4 ± 1.8	4.0 ± 1.8
	Sample 2	5.1 ± 2.2	5.3 ± 1.8	5.0 ± 1.7
	Sample 3	5.0 ± 2.0	5.0 ± 2.0	4.5 ± 2.1
	Sample 4	4.1 ± 2.7	4.6 ± 2.5	4.3 ± 2.8
Study 5 (Rice crackers)	Sample 1	7.1 ± 1.8	7.3 ± 1.0	6.7 ± 1.5
	Sample 2	6.7 ± 1.7	6.0 ± 2.0	6.6 ± 1.8
	Sample 3	6.8 ± 1.6	6.0 ± 1.8	5.7 ± 2.0
	Sample 4	5.9 ± 2.0	6.1 ± 2.1	5.5 ± 1.8
	Sample 5	6.0 ± 2.2	6.2 ± 1.8	5.4 ± 2.2
	Sample 6	6.0 ± 2.3	6.0 ± 2.1	5.2 ± 2.1
Study 6 (Green tea)	Sample 1	4.0 ± 2.1	3.8 ± 2.0	4.3 ± 1.8
	Sample 2	3.2 ± 1.8	3.6 ± 1.9	3.6 ± 1.5
Study 7 (Seaweed cracker)	Sample 1	7.2 ± 1.2	7.1 ± 1.4	7.0 ± 1.4
Study 8 (Seafood dip)	Sample 1	6.6 ± 2.2	7.0 ± 1.4	7.1 ± 1.6
Study 9 (Plain crackers)	Sample 1	6.6 ± 1.8	6.9 ± 1.4	6.5 ± 1.6
	Sample 2	6.3 ± 2.0	6.5 ± 2.0	6.5 ± 1.9
	Sample 3	5.9 ± 2.0	5.8 ± 1.7	5.9 ± 1.8

The reasons why bias occurred in some studies and not the others are unclear at present, but may be linked to the differences between the studies. Gacula (in Moskowitz et al., 2003) mentioned that lack of ability to generalise results of single studies have also been observed when comparing intensity scaling by trained panellists and consumers and he noted that this may be due to effects linked to product categories. In summary, further research is warranted into the conditions where bias occurs and until such time we advocate a pragmatic perspective whereby investigators make informed decisions about whether or not to co-elicit CATA attribute information and if choosing to do so acknowledging that hedonic scores may be biassed.

In a broader perspective it seems that concurrent elicitation of attribute information in hedonic testing (using intensity scales, attribute liking questions, just-about-right scales or CATA questions) has the potential to bias hedonic scores. In Part 1, although we did not establish evidence of hedonic bias when using intensity scales, attribute liking questions or just-about-right scales, others have reported this. The inconsistency among available research findings thus suggests that the effect of analytical questions may be owing to multiple factors, which can be classified as relating to the product, the test ballot/testing conditions, and/or the test participants. While most of previous research, including the present work, has focused on the first two sources of bias, future research could fruitfully investigate whether and how inter-individual differences play a role. The bias of analytical questions, as well as other biases, has been explained within the paradigm of effort avoidance (Krosnick, 1991; Kool, MacGuire, Rosen, & Botvink, 2010) which posits that anticipated cognitive demands have a disruptive effect on the hedonic experience. However, research in cognitive psychology show that individuals are known to vary greatly in the way they respond to effortful cognitive activities. Thus, the issue might be productively explored by using existing psychographic scales to account for inter-individual variation in cognitive effort avoidance, such as the need for cognition scale (Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1984) and the BIS/BAS scale (Carver & White, 1994). It would also be of interest to test other motivation-related factors that can have a moderating factor on effort aversion, both externally – such as the presence of a monetary incentive – and inter-individually – such as product usage and involvement.

In conclusion, existing results demonstrate that the mere presence of analytical does not consistently bias liking ratings, and thus the notion that co-elicitation, in isolation, is sufficient to modify hedonic response should be refuted. Tentatively, a bias is more likely to arise from the interaction effect of two or more co-occurring factors. Research into the experimental conditions that are/are not associated with bias is needed. Another relevant issue to consider is how the inclusion of non-sensory terms in CATA questions affect hedonic scores.

Author contributions

SRJ, GA and DG jointly conceived the research, analysed the data and wrote the paper. All other authors contributed to data collection.

Acknowledgements

Members of the Sensory & Consumer Science team at Plant & Food Research, particularly SokLeang Chheang, David Jin and Denise Hunter are thanked for help in planning and collection of data in Studies 1–2 and 5–8. Financial support for Studies 1–2 and 5–8 was received from The New Zealand Ministry for Business, Innovation & Employment and Plant & Food Research.

Study 4 was supported by the Danish Agency for Science, Technology and Innovation (through the consortium *Danish Microbrew – Product innovation and quality*) and by the Danish Ministry of Economic and Business Affairs (through the project *Local Foods in Denmark*). Additional support for Study 4 was provided by the Faculty of Science, University of Copenhagen, Denmark. The help of Maria Tougaard Andersen, Signe Maria Udengaard Christensen, Marlene Schou Grønbeck, Marie Wolsing Laugesen, and Jonas Astrup Pedersen with the data collection is thankfully acknowledged.

Instituto Nacional de Investigación Agropecuaria (Uruguay) is thanked for providing the strawberry samples and carrying out

data collection in Study 3. Comisión Sectorial de Investigación Científica (Universidad de la República, Uruguay) is thanked for financial support for Study 9.

References

Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check all-that-apply response compared to traditional scaling of attributes for salty snacks. *Poster presented at the 7th Pangborn Sensory Science Symposium.* Minneapolis, MN, USA (12-16 August). Delegate Manual.

Ares, G., Barreiro, C., Deliza, R., Giménez, A., & Gámbaro, A. (2010). Application of a check-all-that-apply question to the development of chocolate milk desserts. *Journal of Sensory Studies, 25*, 67–86.

Ares, G., Barrios, S., Lareo, C., & Lema, P. (2009). Development of a sensory quality index for strawberries based on correlation between sensory data and consumer perception. *Postharvest Biology and Technology, 52*, 97–102.

Ares, G., & Jaeger, S. R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *Food Quality and Preference, 28*, 141–153.

Bruzzone, F., Ares, G., & Giménez, A. (2012). Consumers' texture perception of milk desserts. II – Comparison with trained assessors' data. *Journal of Texture Studies, 43*, 214–226.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319–333.

Dooley, L., Lee, Y., & Meullenet, J. (2010). The application of check-all-that-apply (CATA) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. *Food Quality and Preference, 21*, 394–401.

Driesener, C., & Romaniuk, J. (2006). Comparing methods of brand image measurement. *International Journal of Market Research, 48*, 681–698.

Earthy, P. J., MacFie, H. J. H., & Hedderley, D. (1997). Effect of question order on sensory perception and preference in central location trials. *Journal of Sensory Studies, 12*, 215–237.

Gacula, M., Jr., Mohan, P., Faller, J., Pollack, L., & Moskowitz, H. R. (2008). Questionnaire practice. What happens when the JAR scale is placed between two "overall" acceptance scales? *Journal of Sensory Studies, 23*, 136–147.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinich, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139*, 665–682.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213–236.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537–567.

Lado, J., Vicente, E., Manzzioni, A., & Ares, G. (2010). Application of a check-all-that-apply question for the evaluation of strawberry cultivars from a breeding program. *Journal of the Science of Food and Agriculture, 90*, 2268–2275.

Lawless, H. T, & Heymann, H. (2010). *Sensory evaluation of food. Principles and practices*. New York, NY: Springer.

Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food Quality and Preference, 22*, 733–747.

Mela, D. J. (1989). A comparison of single and concurrent evaluation of sensory and hedonic attributes. *Journal of Food Science, 54*, 1098–1100.

Moskowitz, H. R., Muñoz, A. M., & Gacula, M. C. (2003). *Viewpoints and controversies in sensory science and consumer product testing*. Trumbull, Connecticut: Blackwell Publishing.

Péneau, S., Brockhoff, P. B., Escher, F., & Nuessli, J. (2007). A comprehensive approach to evaluate the freshness of strawberries and carrots. *Postharvest Biology and Technology, 45*, 20–29.

Plaehn, D. (2012). CATA penalty/reward. *Food Quality and Preference, 24*, 141–152.

Popper, R., Rosenstock, W., Schraidt, M., & Kroll, B. J. (2004). The effect of attribute questions on overall liking ratings. *Food Quality and Preference, 15*, 853–858.

Prescott, J. (1999). Flavour as a psychological construct: Implications for perceiving and measuring the sensory qualities of foods. *Food Quality and Preference, 10*, 349–356.

Prescott, J., Lee, S. M., & Kim, K. (2011). Analytic approaches to evaluation modify hedonic responses. *Food Quality and Preference, 22*, 391–393.

Prescott, J. (2004). Psychological processes in flavour perception. In A. J. Taylor & D. Roberts (Eds.), *Flavour perception* (pp. 256–277). London: Blackwell Publishing.

Rasinski, K. A., Mingay, D., & Bradburn, N. M. (1994). Do respondents really "mark all that apply" on self-administered questions? *Public Opinion Quarterly, 58*, 400–408.

Small, D. M., & Prescott, J. (2005). Odor/taste integration and the perception of flavour. *Experimental Brain Research, 166*, 345–357.

Stone, H., & Sidel, J. L. (2004). *Sensory evaluation practices*. San Diego, USA: Elsevier, Academic Press.

Strack, F. (1992). "Order effects" in survey research: activation and information functions of preceding questions. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 23–34). New York, NY: Springer-Verlag.

Sudman, S., & Bradburn, N. M. (1992). *Asking questions*. San Francisco, CA: Jossey-Bass.

Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International, 48*, 893–908.

Vázquez, M. B., Curia, A., & Hough, G. (2009). Sensory descriptive analysis, sensory acceptability and expectation studies on biscuits with reduced added salt and increased fibre. *Journal of Sensory Studies, 24*, 498–511.

Vickers, Z. M., Christensen, C. M., Fahrenholtz, S. K., & Gengler, I. M. (1993). Effect of the questionnaire design and the number of samples taste on hedonic ratings. *Journal of Sensory Studies, 8*, 189–200.