Food Quality and Preference 78 (2019) 103745

Contents lists available at ScienceDirect

Food Quality and Preference

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

Consumer ratings of situational ('item-by-use') appropriateness predict food choice responses obtained in central location tests

Davide Giacalone[a,*], Sara R. Jaeger[b]

[a] University of Southern Denmark, Odense, Denmark
[b] The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand

ARTICLE INFO

Keywords:
Context
Use situations
Appropriateness
'item-by-use'
Food choice

ABSTRACT

Measures of product performance that effectively predict food choices are highly sought after. A simple method to add value to simple hedonic data is that of item-by-use (IBU) "appropriateness" or "situational fitness", where consumers are presented with a list of possible consumption situations and asked to indicate how well a product fits each of them. In a recent paper [Giacalone & Jaeger, 2019], we presented evidence that IBU ratings obtained in response to verbal and pictorial stimuli consistently predicted consumer choice across a diverse range of food and beverage categories. Here, we extend these results to *tasted* products obtained in four Central Location Tests (CLTs). In all studies, the target products were evaluated by two separate consumer panels, one rating the samples for hedonic and appropriateness (IBU), and the second rating for hedonic and choice likelihood (CL) in response to a target situation. In line with expectations, regression analyses consistently revealed a significant and positive link between IBU ratings and CL, with appropriateness accounting for over 70% of variation in consumer choice. These results confirm that appropriateness predicts food choice even when evaluation depends on sensory modalities such as smell, taste and texture. Taken collectively, these four studies indicate that appropriateness should be considered as an important product performance criterion in CLT evaluations of foods and beverages.

1. Introduction

Situational appropriateness, defined as perceived fit between products and use situations, is an important cognitive-contextual measure of consumer attitudes towards foods and beverages (Schutz, 1994; see Giacalone, 2019 for a recent review). Unlike evaluations of preferences and acceptability, where the emphasis is squarely on subjective experiences, the concept of appropriateness relates to normative aspects of food consumptions (how well a food or a beverage fits the situation in which it is supposed to be consumed) learned through experience and socialization.

Although preferences are certainly an important contributor of food choice, high hedonic ratings do not guarantee that a product will be consumed frequently or that it will be appropriate for the situation that product developers/marketers had in mind (Jaeger & Porcherot, 2017). Accordingly, evaluations of 'item-by-use' (IBU) appropriateness have been proposed as an important adjunct to hedonic testing in central location tests (CLT) to ensure that target products not only have high hedonic value, but also high appropriateness for the usage context(s) they are intended for (Cardello & Schutz, 1996; Jaeger & Porcherot, 2017). The literature indicates that appropriateness may have important behavioral correlates of appropriateness; for instance, appropriateness ratings have been shown to predict actual and expected liking (Cardello, Schutz, Snow, & Lesher, 2000; Jaeger, Roigard, Le Blond, Hedderley, & Giacalone, 2019), frequency of consumption (Sosa, Martínez, Arruiz, Hough, & Mucci, 2005), and willingness to pay (Elzerman, Hoek, van Boekel, & Luning, 2015).

In an effort to paint a more complete picture of the appropriateness construct, an important question is whether situational appropriateness is predictive of consumers' food and beverage choices. The theoretical premise for this question is that consumers differentiate products based on intended usage situations, rather than solely on the basis of individual preferences. Anticipated usage situation is therefore an important frame of reference for choosing products that are perceived to be better means to reach specific goals (Cardello & Schutz, 1996; Giacalone, 2019; Jaeger & Porcherot, 2017). For example, a consumer who needs to choose a bottle of wine may pick a reduced alcohol wine if they recently decided to go on a diet, or pick a specific grape variety that they know will pair well with what they are having for dinner – in either case, appropriateness or fit-for-purpose can effectively represent

* Corresponding author at: Dept. of Technology and Innovation, Campusvej 55, DK-5230, Odense M, Denmark.
E-mail address: dg@iti.sdu.dk (D. Giacalone).

https://doi.org/10.1016/j.foodqual.2019.103745
Received 19 June 2019; Received in revised form 18 July 2019; Accepted 19 July 2019
Available online 23 July 2019

an efficient cognitive metric to compare alternative products and choose the most appropriate alternative, and the eventual choice may not always relate to liking for that product in an absolute sense.

To empirically verify this proposition, we recently conducted a series of 15 studies, spanning a wide range of product categories and consumption contexts, focusing on the link between perceived appropriateness and choice (Giacalone & Jaeger, 2019). In each study, consumers performed a choice task on sets of stimuli (verbal or pictorial) chosen to cover the full range of appropriateness (low to high) in a target context. All studies provided consistent and robust evidence that appropriateness evaluations significantly predict choice responses (in some case accounting for over 65% of variance), and overall indicated that appropriateness ratings can provide a simple yet powerful predictor of consumer choice (Giacalone & Jaeger, 2019).

A limitation of this previous work is that the studies did not employ actual food and beverages, but rather verbal (product names) and pictorial (product images) stimuli, which has often been the case in similar past research (e.g., Giacalone et al., 2015; Giacalone & Jaeger, 2016; Jaeger & Rose, 2008). While visual inspection is an important cue to infer usage appropriateness of products, the question remains whether the findings of Giacalone and Jaeger (2019) generalize to situations where other sensory modalities, such as smell, taste and texture, contribute to consumers' evaluations of appropriateness.

Situated within this context, the aim of the present research is to evaluate whether ratings of IBU appropriateness elicited upon actual product tasting are predictive of consumer choice data obtained under CLT conditions.

2. Materials and methods

2.1. Research overview

To address the research aim, we present results from four consumer studies focusing on snack products (Table 1). Each study employed a between-subject design where two different consumer panels (a and b) evaluated the same set of products.

The first panel (a) rated the appropriateness of the samples for a series of use situations, and the second panel (b) rated the same samples for likelihood of choice in response to a sub-set of the use situations used by first panel.

2.2. Participants

Participants in all studies lived in Auckland (New Zealand) and were recruited by a marketing research provider according to specified criteria that besides age and gender quota included willingness to eat the focal product categories. Sample size, age and gender composition of the participants in each study are shown in Table 1. Participants completed the studies as part of CLTs that investigated consumers' responses to a range of foods and beverages. The research was covered by a general approval from the human ethics committee at The New Zealand Institute for Plant and Food Research Limited (PFR). All participants gave informed written consent and were compensated in cash for their participation.

2.3. Products and use situations

Products were selected among commercial varieties of nuts, salted snacks and sweet foods. All were commercially available in Auckland (New Zealand) and purchased at local supermarkets. Specifically, in Study 1, the nut samples (n = 7) were: honey roasted peanuts, roasted and salted peanuts, tamari roasted almonds, roasted and salted macadamia, walnut pieces, brazil nuts and salt and pepper cashew nuts. In Study 2, the salted snacks (n = 5) were: spicy broad beans, pretzel bows, lentil chips, quinoa chips with tomato and garlic flavor, and kumara chips. In Study 3, the sweet and salty snack foods (n = 6) were: sweet and salty muesli bar, caramel and pecan slice, sweet honey and sea salt peanuts, sweet and salty biscuits with lemon filling, sweet and salty kettle corn (popcorn), and protein nut bar with salted caramel flavor. In Study 4, the salted snacks (n = 6) were: potato chips with tzatziki flavor, puffed wheat snacks with flavoring, chickpea chips, potato chips with salt and vinegar flavor, wholegrain chips with honey and mustard flavor, and salted potato chips. The products were selected by author SRJ with input from 4 to 5 sensory and consumer staff at PFR.

For use situations, the strategy was that these be highly diverse for the assessment by the first panel (a). This was to aid selection, prior to the second panel (b), of a sub-set of use situations that were very discriminative between the focal set of samples (largest Cochran Q values). The lists of use situations were developed by author SRJ with input from 4 to 5 sensory and consumer staff at PFR. The full list of use situations for each study (a) is available as Supplementary material (Appendix S1), whereas Table 1 reports those selected for the second panel (b).

2.4. Procedures

The procedures were very similar across the studies. In the first consumer panel of each study (panel a), the elicited responses were liking ratings (1 = 'dislike extremely', 9 = 'like extremely') and perceived appropriateness for a range of use situations obtained via CATA (check-all-that-apply) questions, with liking ratings always being elicited first. In the second consumer panel of each study (panel b), liking was again elicited, followed by ratings of choice probability (1 = 'very unlikely', 7 = 'very likely') in response to the question: "How likely or unlikely would you be to choose this sample?" for each of the focal use situations.

Since between-subjects designs were used, hedonic responses were collected from both panels to rule out alternative explanations for the results based on liking. In Studies 3 and 4, incomplete designs for sample presentation were used, as dictated by the constraints of the

Table 1
Overview of the studies, with experimental conditions, and use situations included in choice likelihood predictions.

Study	Product category	N samples	N consumers	Age range (y.o.) and gender (% female)	Experimental condition[§]	Use situations in Studies 1b, 2b, 3b and 4b
1a	Nuts	7	144	21–63; 29.9%	H + IBU	1) For health-conscious people; 2) In cooking and baking; 3) To supplement my protein intake; 4) When drinking alcohol.
1b	"	"	103	21–67; 0%	H + CL	
2a	Salted snacks	5	270	20–68, 54%	H + IBU	1) For children; 2) For the whole family; 3) When I am with friends and family; 4) When I want something different.
2b	"	"	155	25–60, 52%	H + CL	
3a[¶]	Sweet and salty foods	6	102–105	20–68, 34%	H + IBU	1) For energy; 2) To indulge myself; 3) When choosing a healthy snack; 4) With a hot beverage.
3b	"	"	133	21–63, 0%	H + CL	
4a	Salted snacks	6	116	20–66, 50%	H + IBU	1) For use with dips.
4b[¶]	"	"	136	19–68, 54%	H + CL	

[§] H = hedonic (9-pt scale); IBU = item-by-use (CATA); CL = Choice likelihood (7-pt scale).
[¶] This study used an incomplete block design.

overall research session. In all studies, additional information about the samples were co-elicited with IBU responses (panel a) and/or choice likelihood responses (panel b). This was directed by the overall objectives of these studies, which extended beyond the aim of the present research.

In all studies (both panels), consumers were seated in standard sensory testing booths (white lighting, positive air flow, 20–22 °C). Samples were evaluated in blind conditions, each only labeled with a random 3-digit code. Sample presentation was always monadic and balanced for sample position and first order carry-over effects, following a Williams' Latin square or an incomplete counterbalanced design (depending on whether a full or incomplete design was used – cf. Table 1). In the CATA questions for situational appropriateness, the order of terms varied across and within participants in accordance with Ares et al. (2015). Water and plain crackers were available for palate cleansing (encouraged, but not enforced).

2.5. Data analysis

A stepwise analytical approach, similar to that employed in Giacalone and Jaeger (2019), was applied to the data. Preliminary analyses were performed to ensure that the data were fit-for-purpose. First, differences between the samples in terms of IBU appropriateness were evaluated by means of Cochran's Q test, whereas one-way analysis of variance (ANOVA) was used to ascertain that samples were different in terms of choice likelihood and liking. In Study 3 only, Fisher's exact test for count data was used in lieu of Cochran's test to estimate differences in appropriateness between the samples (the reason being that Cochran's Q test requires a full balanced design, whereas in this study the IBU data came from two different consumer groups, and also because the number of consumers evaluating each sample was slightly different – cf. Table 1).

To achieve the aim of the research, i.e. evaluate whether ratings of IBU appropriateness elicited upon actual product tasting would be predictive of consumer choice data obtained under CLT conditions, results were first visually explored by plotting mean choice likelihood ratings corresponding to differences in appropriateness (% citation) on a context-by-context level. An ANOVA model using IBU data (CATA counts, from the first panel (a)) as the independent variable, and choice likelihood data (7-pt scale, from the second panel (b)) as the response variable was conducted on a context-by-context basis. Subsequently, a linear regression model linking appropriateness (% citation) and mean choice likelihood was conducted, using data from all contexts and all studies. The usual outputs for linear regression (direction and strength of the linear relationship, and % variance of the choice data accounted by appropriateness) were retained for interpretation. For comparison purposes, an identical regression model using mean liking (9-pt scale) as the independent variable was also carried out.

All analyses were performed using R (R Core Team, 2017), with statistical significance set at $\alpha = 5\%$.

3. Results

3.1. Preliminary analyses

Preliminary analyses of between-product differences in terms of appropriateness (Cochran's Q test), liking and choice likelihood (ANOVA), revealed significant differences between products for all contexts and all studies. A summary of these analyses is reported in Tables 2 and 3, which show that products spanned a relatively large range for both IBU appropriateness and choice likelihood, thus confirming that the experimental conditions were appropriate to study variation in the target measures.

Table 2 also shows that all products were in the positive range of the hedonic scale, as no products scored less than 5.6 (of 9), which fitted with the use of commercially available products. Additional results (i.e., at the level of individual products) are available in the Supplementary material.

Table 2
Analyses of between-product differences with respect to key experimental variables in Studies 1a to 4a. The range column reports minimum and maximum mean values on the 9-pt scale (Liking) or % citation (item-by-use (IBU) appropriateness). The last column reports the test statistic used to quantify between samples differences (Cochran's for IBU data and ANOVA for Liking).

Study	Variable	Context	Between-sample differences	
			Range	Test statistic[¶]
1a	Liking	(None)	6.2–7.7	$F_{(7,1168)} = 15.3$
	Appropriateness	For health-conscious people	8.3%–81.3%	$Q_{(7)} = 260.6$
	Appropriateness	In cooking and baking	7.6%–90.3%	$Q_{(7)} = 249.7$
	Appropriateness	To supplement my protein intake	18.8%–63.2%	$Q_{(7)} = 117.6$
	Appropriateness	When drinking alcohol	17.4%–86.8%	$Q_{(7)} = 360.9$
2a	Liking	(None)	5.8–7.2	$F_{(4,1345)} = 27.5$
	Appropriateness	For children	17%–53%	$Q_{(4)} = 135.2$
	Appropriateness	For the whole family	30%–71.1%	$Q_{(4)} = 117.5$
	Appropriateness	When I am with friends and family	48.5%–74.4%	$Q_{(4)} = 51.5$
	Appropriateness	When I want something different	30.4%–78.1%	$Q_{(4)} = 117.9$
3a	Liking	(None)	6.4–7.6	$F_{(5,610)} = 10.9$
	Appropriateness	For energy	16.7%–82.4%	(None)[§]
	Appropriateness	To indulge myself	14.7%–85.3%	(None)[§]
	Appropriateness	When choosing a healthy snack	2.9%–71.6%	(None)[§]
	Appropriateness	With a hot beverage	7.8%–74%	(None)[§]
4a	Liking	(None)	5.6–6.5	$F_{(5,690)} = 4.7$
	Appropriateness	For use with dips	19%–82.8%	$Q_{(5)} = 143.2$

[§] Exact test.
[¶] $p < 0.001$ in all cases.

Table 3
Analyses of between-product differences with respect to choice likelihood in Studies 1b to 4b. The range column reports minimum and maximum mean values (7-pt scale). The last column reports the test statistic used to quantify between samples differences and the η^2 for the product effect.

Study	Context	Between sample differences		
		Range	Test statistic[¶]	η^2
1b	For health-conscious people	2.7–5.6	$F_{(6,817)} = 39.9$	0.23
	In cooking and baking	2.4–5.5	$F_{(6,817)} = 29.9$	0.18
	To supplement my protein intake	3.2–4.9	$F_{(6,817)} = 8.4$	0.06
	When drinking alcohol	2.4–5.8	$F_{(6,817)} = 64.3$	0.32
2b	For children	2.9–4.7	$F_{(4,770)} = 18.6$	0.09
	For the whole family	3.9–5.2	$F_{(4,770)} = 10.6$	0.05
	When I am with friends and family	4.4–5.1	$F_{(4,770)} = 4.0$	0.02
	When I want something different	4.2–5.3	$F_{(4,770)} = 5.6$	0.03
3b	For energy	3.8–6.3	$F_{(5,791)} = 54.5$	0.26
	To indulge myself	4.0–6.3	$F_{(5,791)} = 30.4$	0.16
	When choosing a healthy snack	2.7–5.7	$F_{(5,791)} = 62.4$	0.28
	With a hot beverage	3.2–6.2	$F_{(5,791)} = 83.6$	0.35
4b	For use with dips	2.5–5.3	$F_{(5,822)} = 36.8$	0.18

[¶] $p < 0.001$ in all cases except Context "When I am with friends and family" in Study 2b ($p = 0.003$).

3.2. Predicting choice from appropriateness

To address the main aim of this research – that is, whether IBU appropriateness is predictive of product choice – we start by looking at Table 3. Although the ANOVA models in Table 3 nominally estimate product effects, appropriateness levels (citation frequencies) are

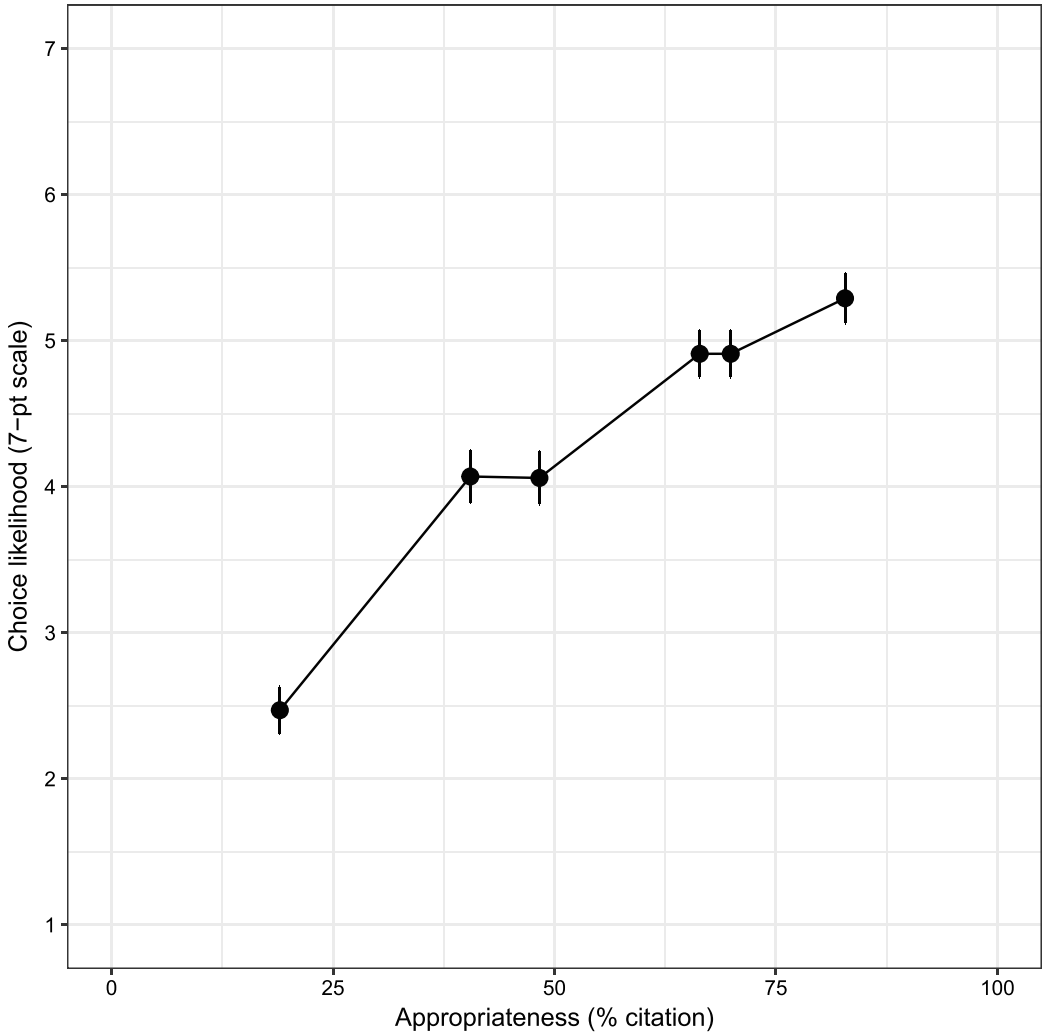

Fig. 1. Choice likelihood ratings (n = 136,7-pt scale, 1 = Very unlikely, 7 = Very likely) by level of IBU appropriateness (% citation, n = 116) for Study 4 (Context: "For use with dips"). Confidence intervals (95%) around the means are also shown.

conflated with products in these datasets. Hence, a significant product effect also implies a significant difference in mean choice likelihood between (at least two) different levels of IBU appropriateness. The significant effects consistently obtained across all use situations unambiguously indicate that IBU appropriateness affects choice likelihood. Appropriateness was also found to account for a substantial proportion of variance in choice likelihood (η^2_{Mean} = 17%). However, consistent with the findings reported in Giacalone and Jaeger (2019), effect sizes were found to vary between studies and between contexts (η^2_{Min} = 2%, η^2_{Max} = 35%). Study 2 showed the weakest effects as appropriateness accounted for less than 10% of the variance in all four contexts, most likely because the range of differences in appropriateness for the products in this study was narrower than in the other three studies (Table 2).

To look at the direction of the effects, we first explored the results visually by plotting mean choice likelihood associated with different IBU levels on a context by context basis. Fig. 1 shows an example of this analysis for the context "for use with dips" used in Study 4, and shows the typical trend observed across all studies which was a linear increase in choice likelihood with increasing level of appropriateness.

Fig. 2 displays the relationship between evaluations of IBU appropriateness and choice likelihood across all contexts and all studies, with each data point corresponding to a specific product (n = 78). Visual inspection of this figure strongly indicates a linear relationship, and, accordingly, these two measures were highly correlated ($r_{(76)} = 0.84$, $p < 0.001$), with IBU appropriateness explaining over 70% of variation in the choice data. Conversely, a corresponding plot (Fig. 3) using mean liking ratings to predict choice likelihood shows a complete lack of correlation between these two measures ($r_{(76)} = -0.05$, $p = 0.649$).

The result that appropriateness and liking were not significantly correlated in this meta-analysis is possibly due to the narrow hedonic range covered, supporting recent remarks of Jaeger and collaborators stating that IBU appropriateness is most useful when working with products that are liked (i.e., above the neutral point on the hedonic scale) and that do not vary much in terms of liking ratings (Jaeger et al., 2019).

4. Discussion and conclusion

The primary goal of this research was to investigate the relationship between situational appropriateness and product choice for food and beverages. Evidence across all four studies consistently indicated that perceived appropriateness significantly predicts consumer choices for food and beverages. Importantly, with respect to informing CLT practices, these results fully replicate those obtained with visual and verbal stimuli (Giacalone & Jaeger, 2019), and show that this relationship extends to tasted food and beverages stimuli. At an overall level, perceived appropriateness was found to predict over 70% of the variance

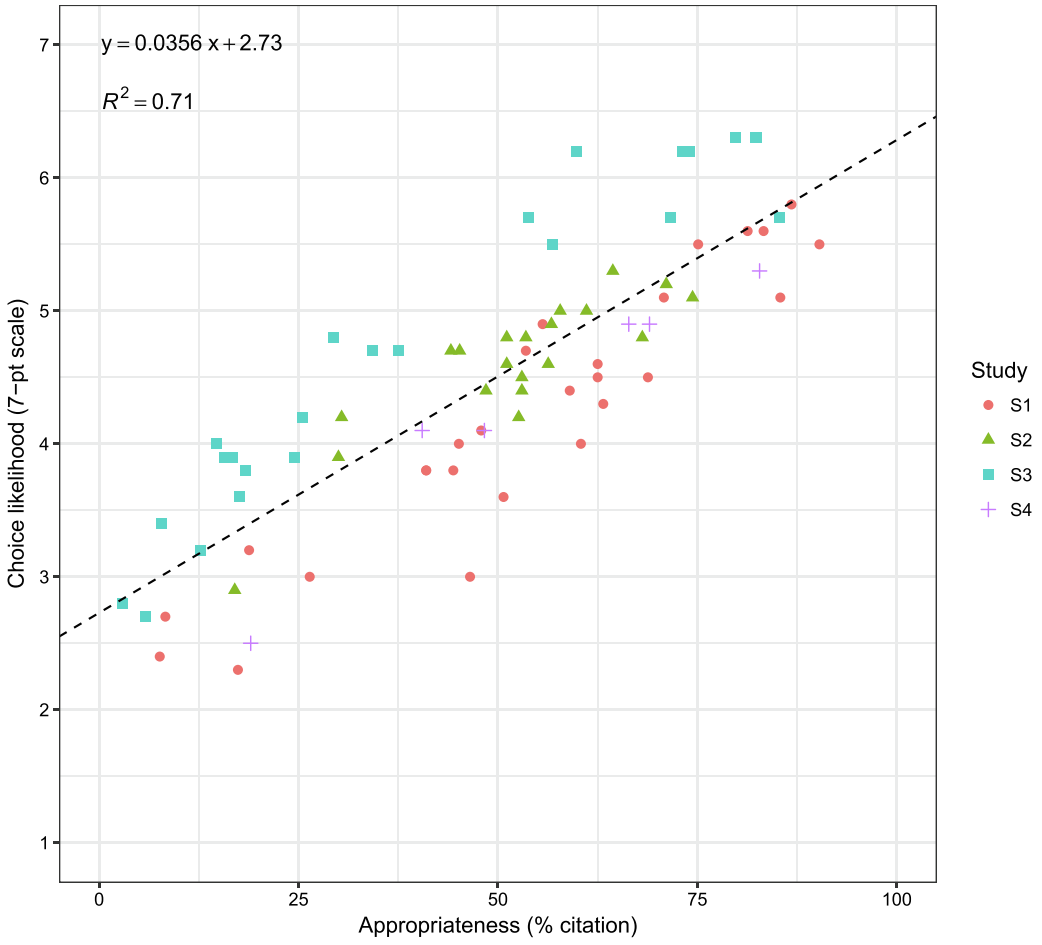

Fig. 2. Scatterplot displaying interrelationships between item-by-use appropriateness (% citation) and choice likelihood (7-pt scale, 1 = Very unlikely, 7 = Very likely). Each data point corresponds to a specific product, and the plot combines data from all contexts and all studies (n = 78). Line of best fit and the associated regression equation are also shown.

in the choice data, making a compelling argument for considering it as an important product performance criterion. The IBU method provides a simple and efficient tool for its assessment under CLT conditions and, we therefore echo previous calls to its wider use in sensory and consumer science (Cardello & Schutz, 1996, Jaeger & Porcherot, 2017; Giacalone, 2019).

Effect sizes were found to vary between different studies and different situations, and this variation was related to the range of differences in IBU appropriateness covered in that study. This is consistent with the data presented in Giacalone and Jaeger (2019), to which we refer the reader for a theoretical account of these results. From an applied perspective, this begs the question of when the inclusion of IBU evaluations would be most beneficial. In this regard, an interesting result was that, in contrast to appropriateness, liking failed altogether to predict choice likelihood within the same datasets. Although admittedly products did not vary in terms of liking as much as they did in terms of appropriateness, this does not make this result any less important as consumer product tests are often conducted with a focus on commercial products (with relatively high sensory quality), and with the goal to differentiate between relatively close alternatives. Evidently, a product deemed highly appropriate for a given situation may still be rejected if it is not liked, so clearly both measures are ultimately relevant to understand and predict consumer food choices. Recent research on appropriateness and liking has indicated that the relationship between the two is non-linear, whereby low appropriateness often couple with *dis*liking, whereas a decoupling occurs for *liked* products, which can then be differentiated on the basis of appropriateness (Jaeger et al., 2019). The lack of correlation between liking and appropriateness found in this paper (where all samples were on average liked and similarly liked) is in line with this interpretation, which was also suggested in early work by Lähteenmäki and Tuorila (1995, 1997). We, therefore, reiterate with greater confidence the suggested implication for product testing put forth in Jaeger et al. (2019) that "the inclusion of appropriateness characterization will be most useful when working with a set of products that are more liked, and more similarly liked, whereas it would be of limited value when working with products with large expected differences in degree of liking/disliking" (p. 11).

From an applied perspective, the ability to predict product success must ultimately encompass a combination of product performance indicators. Therefore, we expect that situational appropriateness evaluations be increasingly used in conjunction with hedonic and other perceptual product responses, like expectations, emotions, and well-being, as already demonstrated in recent papers (e.g., Jaeger et al., 2017; Spinelli et al., 2019), and how to optimally combine these measures should be an important research agenda. Additionally, it should be emphasized that appropriateness evaluations in response to the product itself (e.g., in a blind evaluation) may differ from those elicited in response to extrinsic product elements, requiring careful consideration in the choice of stimuli. For example, in a product development context the classical IBU approach where one consumer evaluates several usage contexts may be more suited to early stage product development, for instance to confirm that the product fits the usage intended by the product developers (Giacalone, 2019). It may also help identify early on whether consumers can readily envision how

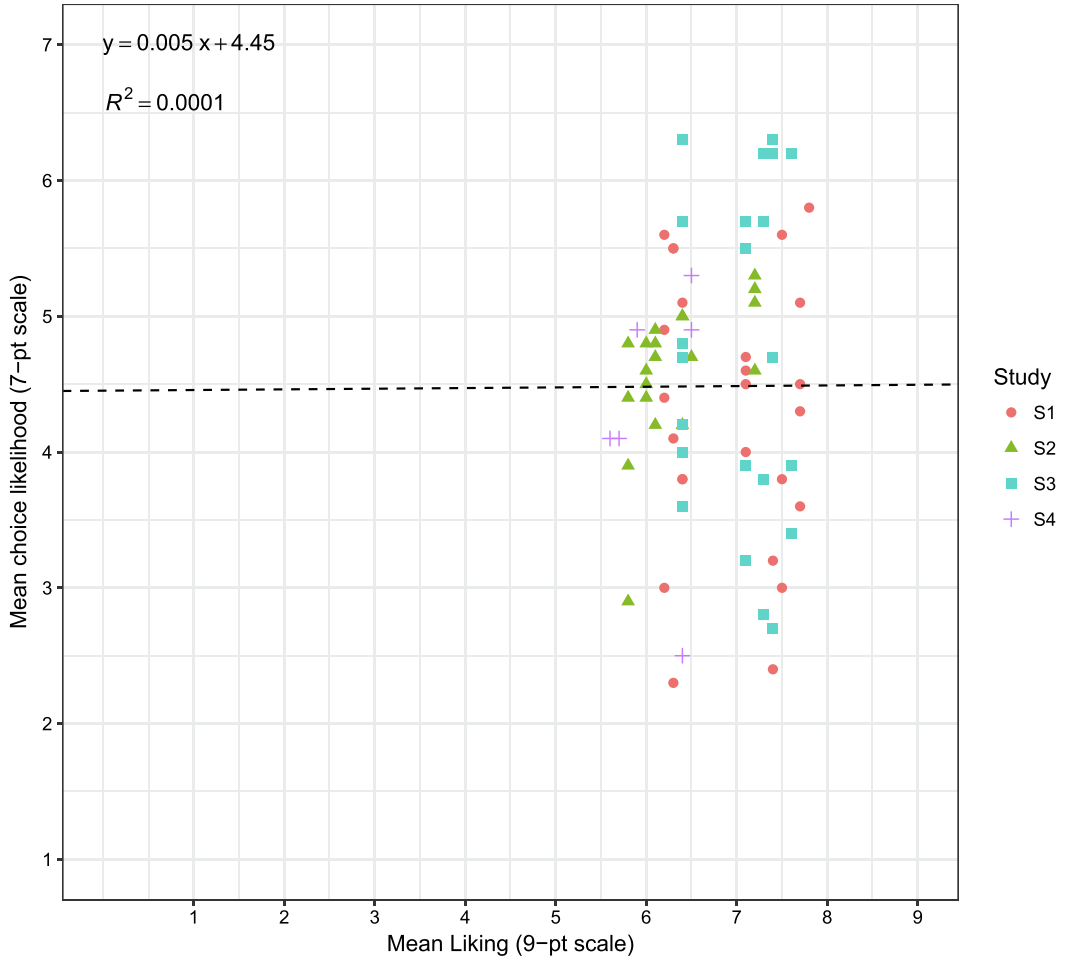

Fig. 3. Scatterplot displaying interrelationships between liking ratings (9-pt scale, 1 = Dislike extremely, 9 = Like extremely) and choice likelihood (7-pt scale, 1 = Very unlikely, 7 = Very likely). Each data point corresponds to a specific product, and the plot combines data from all contexts and all studies (n = 78). Line of best fit and the associated regression equation are also shown.

to incorporate a product in their existing dietary habits, which is often a non-trivial barrier in case of novel product (Giacalone & Jaeger, 2016). Conversely, closer to market launch (once a few target usage contexts have been identified, and where blending intrinsic and extrinsic product elements is important) it may be more meaningful to use more complex approaches such as conjoint analyses, where usage contexts can be specified as design factors with other factors of interest, such as price, packaging, information, etc. (see Jaeger & Rose, 2008 for a discussion on how to implement this).

In closing, we note three limitations of this work that should be addressed in future research. The first is that all analyses presented here have used aggregated data, whilst future studies would be advised to take into consideration inter-individual differences. Here, we draw attention to the fact that consumers have been shown to vary in a systematic manner with respect to the degree to which they follow cultural rules on what is appropriate to consume in different situations (Jaeger et al., 2019), and that psychographics and behavioral variables, such as product familiarity, have been suggested to affect perceived appropriateness and the way this affects choice (Giacalone, & Jaeger, 2016, 2019). The second limitation is that all data were obtained on a laboratory setting. In general, choice data obtained in laboratory experiments have been shown to have good external validity (e.g., Chang, Lusk, & Norwood, 2009), and in Giacalone and Jaeger (2019) we also established that these effects were robust across experimental protocols (likelihood vs. discrete choice). Nevertheless, in the future it would be interesting to relate appropriateness ratings to, say, retail sales data on the same products to really understand to which degree the results presented here generalize to consumer marketplace behavior. The last limitation is that whilst results were remarkably consistent across four studies, all employed relatively similar products (sweet and salty snacks). Additional studies on different product categories not considered here (e.g., beverages) are advised to further generalize these results, as are studies specifying (or quantifying) *a priori* the degree of sensory differences between the products in order to pinpoint more accurately when the IBU method is most beneficial as an adjunct to liking.

Acknowledgements

Financial support was received from two sources: i) The New Zealand Ministry for Business, Innovation & Employment, and ii) the New Zealand Institute for Plant and Food Research Ltd. Staff at the Sensory & Consumer Science Team at Plant and Food Research are thanked for help in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodqual.2019.103745.

References

Ares, G., Reis, F., Oliveira, D., Antúnez, L., Vidal, L., Giménez, A., ... Jaeger, S. R. (2015). Recommendations for use of balanced presentation order of terms in CATA questions. *Food Quality and Preference, 46*, 137–141.

Cardello, A. V., & Schutz, H. G. (1996). Food appropriateness measure as an adjunct to consumer preference/acceptability evaluation. *Food Quality and Preference, 7*, 239–249.

Chang, J. B., Lusk, J. L., & Norwood, F. B. (2009). How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agricultural Economics, 91*, 518–534.

Cardello, A. V., Schutz, H. G., Snow, C., & Lesher, L. L. (2000). Predictors of food acceptance, consumption and satisfaction in specific eating situations. *Food Quality and Preference, 11*, 201–216.

Elzerman, J. E., Hoek, A. C., van Boekel, M. J., & Luning, P. A. (2015). Appropriateness, acceptance and sensory preferences based on visual information: A web-based survey on meat substitutes in a meal context. *Food Quality and Preference, 42*, 56–65.

Giacalone, D. (2019). Situational appropriateness in food-oriented consumer research: Concept, method, and applications. In H. Meiselman (Ed.). *Context: The effects of environment on product design and evaluation* (pp. 111–140). Amsterdam: Elsevier.

Giacalone, D., Frøst, M. B., Bredie, W. L., Pineau, B., Hunter, D. C., Paisley, A. G., ... Jaeger, S. R. (2015). Situational appropriateness of beer is influenced by product familiarity. *Food Quality and Preference, 39*, 16–27.

Giacalone, D., & Jaeger, S. R. (2016). Better the devil you know? How product familiarity affects usage versatility of foods and beverages. *Journal of Economic Psychology, 55*, 120–138.

Giacalone, D., & Jaeger, S. R. (2019). Perceived situational appropriateness as a predictor of consumers' food and beverage choices. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.01743.

Jaeger, S. R., & Porcherot, C. (2017). Consumption context in consumer research: Methodological perspectives. *Current Opinion in Food Science, 15*, 30–37.

Jaeger, S. R., Cardello, A. V., Chheang, S. L., Beresford, M. K., Hedderley, D. I., & Pineau, B. (2017). Holistic and consumer-centric assessment of beer: A multi-measurement approach. *Food Research International, 99*, 287–297.

Jaeger, S. R., Roigard, C. M., Le Blond, M., Hedderley, D. I., & Giacalone, D. (2019). Perceived situational appropriateness for foods and beverages: Consumer segmentation and relationship with stated liking. *Food Quality and Preference, 78*, 103701.

Jaeger, S. R., & Rose, J. M. (2008). Stated choice experimentation, contextual influences and food choice: A case study. *Food Quality and Preference, 19*, 539–564.

Lähteenmäki, L., & Tuorila, H. (1995). Consistency of liking and appropriateness ratings and their relation to consumption in a product test of ice cream. *Appetite, 25*, 189–198.

Lähteenmäki, L., & Tuorila, H. (1997). Item-by-use appropriateness of drinks varying in sweetener and fat content. *Food Quality and Preference, 8*, 85–90.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing https://www.R-project.org/.

Schutz, H. G. (1994). Appropriateness as a measure of the cognitive-contextual aspects of food acceptance. In D. M. H. Thomson (Ed.). *Measurement of food preferences* (pp. 25–50). London: Blackie Academic.

Sosa, M., Martínez, C., Arruiz, F., Hough, G., & Mucci, A. (2005). Degree of appropriateness and frequency of consumption of mayonnaise, ketchup, mustard and similar sauces in Argentina. *Food Quality and Preference, 16*, 667–674.

Spinelli, S., Dinnella, C., Ares, G., Abbà, S., Zoboli, G. P., & Monteleone, E. (2019). Global Profile: Going beyond liking to better understand product experience. *Food Research International, 121*, 205–216.