Food Quality and Preference 41 (2015) 84–95

Contents lists available at ScienceDirect

Food Quality and Preference

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

The use of immersive technologies to improve consumer testing: The role of ecological validity, context and engagement in evaluating coffee

Ronald G. Bangcuyo, Kacey J. Smith, Jamie L. Zumach, Alex M. Pierce, Gretchen A. Guttman, Christopher T. Simons *

The Department of Food Science & Technology, The Ohio State University, 2015 Fyffe Rd., Columbus, OH 43210-1007, United States

ARTICLE INFO

Article history:
Received 13 August 2014
Received in revised form 18 November 2014
Accepted 25 November 2014
Available online 4 December 2014

Keywords:
Context
Engagement
Hedonics
Immersive technology
Consumer testing

ABSTRACT

Traditionally, sensory consumer testing is completed in isolated sensory booths where the influence of non-product (e.g., environmental) attributes is controlled. However, these highly controlled environments strip away meaningful contextual (visual, auditory and olfactory) information important in forming consumer perceptions, liking and behaviors. Moreover, boredom and lack of panelist engagement associated with typical testing paradigms can result in uninformative or misleading consumer data. We utilized an immersive environment depicting a virtual coffeehouse, replete with visual, auditory and olfactory cues found regularly in this setting, and compared liking scores for five coffees to those obtained from the same individuals in a traditional testing environment in which contextual information was absent. We found significant differences in preference order and liking for coffees evaluated by the same people in the two settings and showed hedonic data collected in the virtual coffeehouse to be more discriminating and a more reliable predictor of future coffee liking unlike data collected in traditional sensory booths. Additionally, we found consumers to be more engaged in the testing when evaluating coffees in the virtual coffeehouse, an outcome that likely also contributed to improved data quality. These results suggest that during product evaluations, extrinsic contextual information is processed simultaneously with the intrinsic product attributes to influence hedonic assessments and shape reward outcomes. Further, these results suggest that methodological changes to current testing strategies have the potential to improve the reliability of consumer data providing food and consumer product companies significant savings on product development costs and failed launches.

Introduction

Billions of dollars are spent annually on failed product launches in the Food and Beverage industry (Costa & Jongen, 2010; Stewart-Knox & Mitchell, 2003) and nearly 60% of all new product launches fail in the grocery sector (Costa & Jongen, 2010) with some estimates as high as 85% (Costa & Jongen, 2010; Redmond, 1995). Multiple factors contribute to this failure rate including technical limitations and poor marketing and/or pricing strategies, however, the inability of consumer sensory and hedonic information to reliably predict consumer decisions (Deliza & MacFie, 1996; Koster & Mojet, 2007; Rosas-Nexticapa, Angulo, & O'Mahony, 2005) is believed to be a major contributor (Koster & Mojet, 2007). Current testing methodologies typically sequester consumers into isolated sensory booths where control against panelist bias and confounding non-product influences can be maintained. However, such conditions are devoid of context and lack ecological validity resulting in data having little predictive power or worse, that are misleading. The situation is further exacerbated by testing conditions and paradigms used in traditional sensory and consumer testing facilities that fail to adequately engage consumer panelists. As a consequence, boredom and lack of attention result in noisy or non-discriminating responses that add little value to the accumulated data.

In real-world settings, contextual information available during the assessment and consumption of foods and beverages is complex and multifaceted. Preparation circumstances, consumption setting and the presence or absence of others are all features that consumers typically process when making evaluations about a meal. These conditions are rich with dynamic visual, auditory, tactile and olfactory stimuli which serve as a vivid source of contextual information. Consequently, these various streams of information create complex contextual cues that shape subsequent expectations, perceptions, hedonic assessments and behaviors exhibited by a consumer. Prior studies have indicated that

* Corresponding author. Tel.: +1 614 688 1489; fax: +1 614 292 0218.
E-mail address: simons.103@osu.edu (C.T. Simons).

http://dx.doi.org/10.1016/j.foodqual.2014.11.017
0950-3293/

numerous contextual factors impact liking and food-related behaviors including ambiance (Delarue & Boutrolle, 2010; Stroebele & De Castro, 2004), eating locations (Bell, Meiselman, Pierson, & Reeve, 1994; Delarue & Boutrolle, 2010), number of people present (Sommer & Steele, 1997), ambient temperatures (Westerterp-Platenga, 1999), sounds (Ferber & Cabanac, 1987), lighting (Kasof, 2002), food accessibility (Meyers, Stunkard, & Coll, 1980), time of consumption (De Castro, 1987), stated meal (e.g., breakfast, lunch or dinner; Delarue & Boutrolle, 2009) and the color (Clydesdale, Gover, Philipsen, & Fugardi, 1992), temperature (Zellner, Stewart, Rozin, & Brown, 1988) and aroma (Yeomans, 2006) of food.

In addition to the lack of ecological validity, boredom and lack of panelist engagement is another contributing factor to poor predictability of consumer data. Engagement is a complex construct that depends on the aesthetic appeal, novelty and usability of a system such that attention and interaction are maximized (O'Brien, 2008). In the sterile environment of traditional sensory and consumer testing facilities, many consumers fail to attend to the test questions and, instead, provide the same (or nearly the same) hedonic score to all samples. These non-discriminating panelists limit the utility of hedonic data and, unfortunately, may comprise a substantial proportion of the sample population (Cleaver & Wedel, 2001; Koster, 2009).

One potential way to provide relevant context and simultaneously improve consumer engagement is through the use of emerging immersive technologies that create virtual environments specific for a particular occasion. Within the business sector, immersive technologies have been employed to investigate consumer behavior at the point of purchase (for review see Daugherty, Li, & Biocca, 2005). Such efforts have allowed companies to identify how extrinsic product properties such as packaging, shelf placement, pricing and advertising influence the consumer purchase decision. Incorporating immersive technologies into consumer hedonic paradigms might also improve the ecological validity and level of engagement of testing conditions by restoring relevant contextual information during new product testing.

Presently, we investigate the use of immersive techniques in consumer hedonic testing of coffee. In particular, we sought to understand the role of complex, multifaceted contextual information in shaping reward and hedonic responses and to determine the impact of such testing conditions on consumer engagement and data quality. As such, we asked consumers of coffee to evaluate samples in a traditional sensory environment versus a virtual coffeehouse replete with the visual, auditory and aromatic cues associated with this setting. We hypothesized that incorporating important contextual variables associated with coffee consumption would improve the sensitivity, power and reliability of consumer hedonic data as manifested by (1) a different preference order for products evaluated in immersive compared to traditional conditions, (2) more significant differences among products when evaluated in an immersive environment compared to a traditional testing environment and (3) more stable liking results when the same panelists repeat the test following *ca.* a 1-month hiatus, respectively. We further hypothesized that using immersive environments in hedonic testing of coffee would be more engaging as indicated by subjective assessments of the testing environments.

Materials and methods

Participants

Fifty coffee consumers (23 male and 27 female) ranging in age from 19 to 61 years, who regularly (at least once per week) purchase and/or consume coffee from a coffeehouse were recruited for this study. One person was dropped from the study during phase 1 due to incomplete data and three additional participants failed to return for phase 2. All participants were enrolled underwritten informed consent approved by The Ohio State University Institutional Review Board. Subjects were asked to refrain from eating, drinking or smoking for at least 2 h prior to the start of the experiment. Each subject participated in two experimental sessions (referred to as replication 1 and replication 2) approximately 1-month apart. Each experimental session lasted approximately 30-min. At the conclusion of each experimental session, participants received a $10 gift card.

Stimuli

Whole bean coffees were purchased, placed in a freezer (−18 °C) and kept sealed until used. Coffee samples included Starbucks Pike Place Blend (Seattle, WA), Stauf's Ethiopia Moka Harras (Columbus, OH), Maxwell House (Kraft, Glenview, IL) and Douwe Egberts (DE) Java Coast Espresso (Amsterdam, The Netherlands). These coffees were selected as they represented a range of price points and, we hoped, quality. However, to increase the likelihood that the coffees spanned the hedonic continuum, we served the DE coffee at regular and 1.5 strength (DE1.5) with the belief that at least one of those coffees would be disliked by each consumer as being too strong or too weak. Within 1-h prior to the start of the experiment, coffees were removed from the freezer and ground to completion using a commercial grade conical burr grinder (Capresso 560 Infinity Burr Grinder, Jura AG, Niederbuchsiten, Switzerland) on a grind setting of "medium". Two grinders were used and cleaned between uses. Fifty grams of each ground coffee, except DE1.5 for which 75 g was used, was weighed out, put into paper coffee filters and placed into one of four coffee makers (Sunbeam, Boca Raton, FL); each coffee maker was dedicated to brewing only one of the coffee samples. Two-thousand-one-hundred grams of filtered water was added to the reservoir of each coffee maker. The coffee makers were turned on and hot water was filtered through the grounds into a glass carafe. At the conclusion of the brewing cycle, the coffees were poured into 2.5 liter, glass-lined vacuum-insulated air pots (Choice, Lancaster, PA) where they were maintained at a constant temperature of ~81 °C for the duration of the experiment. Fresh pots of coffee were brewed as needed typically every 2.5 h.

Procedure

Each experimental session consisted of two trials—one taking place in the traditional sensory booth and the other in the virtual coffeehouse. To minimize extraneous confounding influences unrelated to our contextual variables of interest, panelists completed the test in one environment and then went immediately to the other environment and completed the test a second time; the environment in which samples were evaluated first was counterbalanced across subjects. At the onset of the first trial, each panelist answered a demographic questionnaire. When completed, subjects were given the five coffees simultaneously and in random order and asked to rate the acceptability of each sample using the 9-point hedonic scale. The verbal instructions to the panelists, as well as the written instructions on each sample ballot, indicated that coffees should be evaluated in the order presented and that, after evaluating the coffee and completing the ballot, they should not go back and re-taste or look at previous hedonic scores. Filtered water was provided for rinsing between coffee samples. Following this, the coffee samples were removed and subjects completed the 21-item Engagement Questionnaire. The Engagement Questionnaire was derived based on prior testing instruments developed by O'Brien and Toms (2010) and Witmer and Singer (1998). From these surveys, we identified 21 questions that measured the level