



Do Beautiful Stores improve Product Evaluation?

MÜNSTER Mia B.*; KRISTENSEN Tore and GABRIELSEN Gorm

Copenhagen Business School, Denmark

*corresponding author e-mail: mmu.marktg@cbs.dk

doi: 10.33114/adim.2019.02.173

Retail designers often emphasize the importance of creating stores that consumers will find attractive. This paper challenges that commonly held view, presenting empirical results from a field experiment showing that a positive rating of a store interior does not affect the product rating to the degree expected. This paper proposes a method for measuring spillover effects, which ordinarily take place without conscious attention. The method was applied in an experiment where 50 shoppers were asked to rate six fashion products in three differently designed stores. Respondents were asked to rate stores and products from within the stores. Any discrepancy between the in-store ratings can be interpreted as the influence of the store design. Results indicate measurable spillover effects from store design to product preference. Surprisingly, however, only one of the three stores showed a significant correlation between the respondents' highest product rating and store preference.

Keywords: Retail design, store design, retail atmosphere, field experiment, consumer preference

Introduction

Throughout mankind's history, architectural spaces intended to enhance a particular experience account for significant expenditures of both effort and resources (Gehl 2011; Lawson 2001). It is also widely acknowledged that physical surroundings have an effect on not only our feelings and behavior, but also on our experience of other objects or entities in that environment, be they works of art, music, or artifacts (Clark 2006; Damasio 1995; Frijda 1989). Retailers are well aware that store environments have an influence on consumer behavior and emotion, and as a component of a marketing strategy often devote considerable resources to designing stores that display their products. Stores can be designed in countless ways, so what should the retail designer's priorities be, if the goal is to create store environments that attract consumers and sell products?

Philip Kotler (1973) was among the first researchers to acknowledge the effect of store design. He introduced the term 'atmospherics' to denote the intentional control and structuring of environmental cues. Since that time, scholars have undertaken various studies on atmospheric variables in retail spaces (J. A. Bellizzi, Crowley, and Hasty 1983; Donovan et al. 1994; A. M. Kent and Kirby 2009; Spence et al. 2014; Turley and Chebat 2002; Turley and Milliman 2000). In spite of these efforts, we know surprisingly little about how consumers are actually affected by a store design as a whole. Consumers are seldom asked to provide feedback about their experiences of store designs. Even when they are, this feedback is of questionable reliability, since most people are simply not able to articulate how an interior environment affects them, let alone their perception of products within that environment. At least some portion of atmospheric effects take place without conscious attention, and this circumstance casts a shadow over direct interrogation as a method of data collection. Previous research has addressed how context can affect product perception and refers to



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 4.0 International License.

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

this phenomenon as a 'spillover' effect (Hagtvedt & Patrick 2008; Leder et al. 2004) yet scholarship investigating the spillover effect from store design to product perception is meager at best.

This paper describes first of all a study design for measuring the extent to which store preference spills over onto product preference, and then compares this measurement with respondents' explicit impression of the store design. The method is tested in a field experiment conducted in three fashion stores with contrasting interior designs. In the experiment, the store design preferences of 50 shoppers are correlated with their product preferences within the three stores. Data from the experiment shows that a spillover effect from store design onto product preference is in fact measurable. At the aggregate level, product preferences are indeed supported by highly rated interiors. But analysis of ratings taken from within each of the three interiors separately reveals interesting differences. By virtue of repeated individual measurements carried out in the three different settings, we can see that the supporting effect of a highly rated store interior is present in only one of three test stores. No significant connection between store preference and product preference is identifiable in the two remaining environments. These results and their implications lay the groundwork for a theoretical discussion about spillover effects from within the context of retail design.

The following provides a brief introduction to existing studies on retail atmospherics and the perception of designed environments. We discuss literature supporting the contention that interior spaces affect product perception, and also describing the fact that environments are largely experienced unconsciously and how this complicates studying their effects. Following this we describe our methodological approach, document the field experiment, and then present and compare the results. Finally, the paper ends with a discussion of the theoretical, methodological, and managerial implications of these findings, and proposes areas for further research.

Background

Studying Store Atmospheres

A number of scholars have studied the effects of individual atmospheric cues in store design. The literature is replete with studies investigating atmospheric variables such as sound (Knofler et al. 2012; Milliman 1982; Morin, Dubé, & Chebat 2007; North, Sheridan, & Areni 2016; Yalch & Spangenberg 2000), color (J. A. Bellizzi, Crowley, & Hasty 1983; J. a. Bellizzi & Hite 1992), odor/scent (Chebat et al. 2012; Herrmann et al. 2013; Spangenberg, Grohmann, and Sprott 2005), lighting (Areni and Kim 1994; Custers et al. 2010; Quartier, Vanrie, and Van Cleempoel 2014), and indoor climate (Frontczak & Wargocki 2011; Zhao, Kim, and Srebric 2015). But atmospheric cues are never experienced in isolation. Recently, scholars have begun to suggest that studies taking a more holistic approach to retail environments should provide a more realistic assessment of their effects on consumers (Baker et al. 2002; Ballantine, Parsons, & Comeskey 2015; A. M. Kent & Kirby 2009; van Rompay et al. 2012; Spence et al. 2014).

Several circumstances stand in the way of studying variables in store environments with any precision. First of all, as mentioned above, consumers do not experience individual cues in isolation, but rather experience the store environment as a whole (Ballantine, Jack, and Parsons 2010; Ballantine, Parsons, and Comeskey 2015; Mattila & Wirtz 2001). Moreover, store atmospheres are created as backgrounds for merchandise, and the influence of the atmosphere is largely experienced without conscious attention. Consumers do not enter a store to evaluate its interior, or to consciously identify qualities that attract their attention (Newman and Foxall 2003). The effect of the retail interior can therefore be difficult to verbalize and recall (Donovan & Rossiter 1982). Consumers come to stores to shop, and their attention is normally focused on the products, and not on the interior design. For these reasons, the use of direct interrogation as a method for studying retail atmospheres is problematic at best. To counteract the effects of quickly fading memory, measurements in any such study should be taken as closely as possible to the time and place of the shopping experience, and preferably from within the store itself, so that the influence of the environment is still in effect (Donovan & Rossiter 1982). Control of multiple variables also presents a challenge for researchers. Some studies approaching atmospheric cues holistically do exist, but most are made in artificial settings where variables are easier to control. Research shows, however, that people behave and react differently in an artificial setting than they do in a 'real life' setting (Frijda 1989; Groeppel-Klein 2005; Lynch, Jr. 1982; Tversky 2008). A subject's simple awareness of being in a laboratory or other artificial setting shifts his or her consciousness, and therewith any response he or she might give. In reality, a consumer's experience of a store

consists of a blending of sensory effects, a commingling that is difficult if not impossible to achieve in artificial settings. This seems a strong argument for studying customer experiences in real shops, despite the problems presented by their complexity.

A small group of scholars have indeed braved these challenges and executed holistic studies of store atmospheres in real shopping environments. Ballantine et al. (2015) made on-site interviews to analyze atmospheric cues in store environments, while both Kirby and Kent (2010) and Petermans, Kent, and Van Cleempoel (2014) used photo-elicitation, where respondents are shown photographs of a store environment, to better understand how atmospheric cues are interpreted by consumers. This method provides a visual representation of the environment as a whole, and therefore to some extent represents the actual store atmosphere. Studies like these provide a good beginning to the study of consumers' store experiences from a more integrated perspective. Still, a significant problem is common to the method of these studies, and one that potentially clouds their results. This problem is that direct interrogation necessarily triggers respondents' reflection upon their own reactions, which colors the nature of their responses. In addition, interrogation presupposes respondents' ability to express how a store design affects them, when in fact this ability varies considerably among individuals. Actual experiences of interior spaces are immediate and take place without much conscious reflection. We therefore believe that we should not limit our investigations to respondents' conscious expression, but incorporate unconscious effects, which the respondent might not be able to articulate. We will expand upon these ideas, and how we approach this conundrum, in the following sections.

Beauty and Judgment of Store Designs

Most people will be able to recall the sensation of entering an architectural space that incites a bodily reaction, or perhaps even feelings, like happiness, excitement or fear. Most people can also recall a particular store where the atmosphere created a particular feeling, without being aware of what it might be that gave rise to that feeling. It is well known that variables like figural goodness, contrast, repetition, symmetry, and prototypicality influence aesthetic judgments (Kubovy 2000; Leder et al. 2004; Martindale & Moore 1988). It is also well known that aesthetic judgments are individual and depend on a person's memory, knowledge, experience and personal taste (Leder et al. 2004; Reber, Schwarz, & Winkielman 2004). Thus, the impact of a store design is highly context-dependent, and inevitably associated with a consumer's personal experiences of the individual store (van Rompay et al. 2012).

In order to investigate whether a consumer's appreciation for a store design affects their perception of the products in it, we need to be clear about what we mean. In the title of the paper we use the word 'beautiful,' which can mean many things. Studies of beauty are rare, perhaps because the concept is so difficult to define precisely, changing as it does according to time and place. A preponderance of the beauty studies that have been carried out are associated with Gestalt theorists such as Rudolf Arnheim (1974) and Ramachandran and Hirstein (1999). Aesthetic experiences are often supposed to be experienced intuitively. Asking people to articulate such experiences will usually receive a reply reflecting a rationalized version of what is assumed to be 'good taste.' For this reason, researchers rarely ask respondents to comment on 'beauty' or 'good taste' per se. Instead, most researchers focus on attributes like figural goodness, pleasantness, liking, or preference. The argument is that eliciting such simple judgments makes it possible to identify the basic processes or modifying variables that underpin aesthetic experiences (Reber, Schwarz, & Winkielman 2004). Judgments based on the question 'how much do you like this interior?' are therefore believed to be closely related to how beautiful or aesthetically pleasing an interior is in the eyes of the respondent. For our purposes, we will define beauty in the following way: a subjective experience of pleasure produced by products/interiors and not mediated by intervening reasoning. This formulation hews closely to definitions of aesthetic experience used by Kubovy (2000); Martindale and Moore (1988); Reber et al. (2004) and Leder et al. (2004). This study is therefore designed around the question 'how much do you like this store/product?' which was used to elicit and determine respondents' expressed preferences.

Searches of existing literature did not reveal studies where preference for store design is studied in relation to product preference.

Objectives

With the abovementioned literature as a background, we propose studying retail atmospheres from a practical and process-oriented perspective, using actual customers' in-store preferences and behavior as indicators. Research investigating atmospheric cues in this manner is limited, as mentioned above, and where it does

exist, employs a predominantly qualitative approach. We contend that atmospheric cues are largely perceived unconsciously, and that qualitative methods using direct interrogation as a method of data collection are therefore inadequate. To address this deficiency, this paper proposes a method for discovering the extent to which expressed store preference spills over onto product preference, without bringing respondents' conscious attention to the store interior.

Tools for studying a store interior's effect on consumers should be of great interest to both retailers and design managers, both of whom have obvious interest in determining whether their designs are properly matched with their targeted market segment. The uncertainty surrounding how a design is perceived by consumers is often pointed out in design and marketing literature; designers, it is usually argued, are the ones who should both know and be able to realize the consumer's needs. John Heskett (2005), for example, acknowledges the conflicting imperatives between a company and the users of its products, and describes how the role of design needs to be understood as providing a bridge between them. Philip Kotler (1973) underscores the relevance of the designer's understanding of the consumer by drawing a distinction between intended and perceived atmosphere. Kotler defines the intended atmosphere as the set of sensory qualities that the designer of the store means to invoke, while the perceived atmosphere is the sensory qualities experienced by the consumer. As a further complication, perception can of course vary significantly from one consumer to the next. Further still, designers are employed by clients who typically bring their own perceptions and intentions to bear on the situation (Haug & Münster 2015; T. Kent & Stone 2007). Considering all these complicating factors and interests together, it becomes obvious that a reliable method for understanding how designs are experienced subjectively would be of indisputable value to decision makers at every level in the field of retail.

In order to delimit the study, we will focus our attention on the variables controlled by the designer: namely, layout, furniture, decoration, and display (Haug & Münster 2015). Using this method, we are able to investigate whether different store characteristics affect individual consumers' product preferences and determine the extent to which a highly rated store design affects ratings of products inside the store. A strong correlation between store design preferences and merchandise preferences would indicate that interviews might in fact be a sufficient method of data collection; a weaker correlation would indicate, conversely, that perhaps this data is not as trustworthy as we'd like it to be.

Research Methodology

Experimental approach

The experiment was designed to investigate whether an interior, that a respondent has expressed a preference for, has a positive influence on the respondent's perception of the products in that interior. In order to determine this, various measurements from each participant were needed. The experiment was therefore designed to collect data indicating (1) each participant's preference for products rated in a neutral setting and (2) each participant's preference for products rated from within the different interiors, (3) each participant's preference for the interiors.

Ratings were measured for each individual. Preferences depend on individual factors like cultural and biological background, education, and personal experience (Kubovy 2000; Leder et al. 2004; Martindale & Moore 1988; Reber, Schwarz, & Winkielman 2004; Thurstone 1928), which makes aggregated data less useful. Each consumer is unique, in other words. Despite the fact that many marketing scholars study consumers by grouping them together at the aggregate or market-segment level, we contend that considering consumers as individuals will grant access to potentially latent information. The experiment was therefore designed first to take individual ratings, and then analyze this data to identify patterns and structures in the population (Krackhardt 1992; Wright 1997). The first part of the study was designed to measure the influence that the interior has on the product when presented in the store — the so-called spillover effect. The second part of the study was designed to measure an explicit preference, or lack thereof, for the store interiors. Comparing each respondent's expressed preference for a store with his or her highest rated product will tell us whether the product is indeed more highly rated in an interior that the respondent likes.

To measure the unconscious effects that interiors might have on products, it was deemed necessary to avoid drawing attention to the interior. To avoid undue reflection, respondents were asked to rate products while physically present in three different test stores. Respondents were therefore influenced by the interior

and its various cues, without knowing that the stores were the real subject of the questioning, and not the products.

Ratings were collected using a touch screen tablet which showed photos of the products taken in the respective environments, such that the hanger and background was visible in the photos (Photo 4). Preference for the products was indicated on a visual analogue scale (VAS) using rated, paired comparisons (Wright 1997). Moving the cursor left or right from center, the respondent indicates which of two products he or she prefers, and the relative strength of that preference.



Photo 4. For each pair of products, the respondent is asked to indicate his or her preference by moving the cursor on the visual analogue scale to a position that reflects the degree of preference to the products shown.

To add a control product rating, respondents were then taken to what we will call a neutral zone, outside the influence of the three test interiors. Here respondents were allowed to examine the products more closely, before rating each product again. Product ratings were again collected on a touch screen showing a photo of the product and a visual analogue scale running from 'I don't like at all', on the left, to 'I like very much', on the right. Respondents were asked to indicate a preference for the products by moving the cursor left or right of center. Any discrepancy between in-store product rating and product ratings given in the neutral zone is, we contend, attributable to the influence of the interior: the spillover effect (Model 1).

In the second part of the study, the same respondents were asked to observe the store interiors and rate them according to their personal taste. These ratings were once again collected on a touch screen tablet, which showed a photo of the store from the perspective where the respondent was standing with the researcher. Above the photo was the question: 'How much do you like this store?' and below, a visual analogue scale with 'I don't like at all' at the left endpoint and 'I like very much' on the right. Once again, the respondent was asked to indicate how much he or she liked the store design by moving the cursor left or right of center.



Model 1. In the first part of the study the products are rated inside three different test stores and in a neutral zone. In the second part of the study the design of the test stores is rated.

Comparison of results from part 1 and part 2 should indicate the level of correspondence between a respondent's expressed preference for an interior and that interior's spillover effect, if any, on the respondent's product preferences (Model 1).

Stimulus Selection

Before carrying out the experiment, field studies and interviews with retailers and retail designers were conducted in order to locate suitable test stores. The three test stores selected were in close proximity to one another, which was considered ideal. This would allow the researcher to guide respondents through the different environments without encountering additional, unintended atmospheres, which might influence respondents. The experiment was conducted in a European mall in 2015. The first 20 tests took place in February, the next 20 in March-April, and the last 10 in May. Respondents were actual customers who had entered the stores to view, try on, and potentially purchase the displayed fashion items, and were recruited upon leaving the store. The researcher explained that the experiment involved rating different products in the store.

Test Stores

Test environments were similar in size (between 31-39 m²), room height (4,5 meters), light intensity/quality, sound, and odor, in order to remove the influence of variables that are not always controlled by the retail designer (Haug & Münster 2015). Quantity of products displayed, and price level were also comparable between all three test stores. As mentioned above, it was deemed important to use test stores which were in close proximity to one another, in order to reduce the influence of external factors during the experiment. The test site was in fact a single men's store consisting of three separate in-shops, each specifically designed for three different men's fashion brands. Most importantly for our purposes, the design of each of the three in-shops was easily distinguishable from the other two. Furniture, wall coverings, floor materials, decoration items, brand images, hangers, and collections; each of these had its own separate identity in each of the shops. We will hereafter refer to the test environments as Shop A, Shop B and Shop C (Photos 1-3).

Shop A had a modern, industrial, bright appearance. All surfaces were white, and metal is a consistently used material. Shop B was characterized by a vintage look using different natural surfaces: brick-

walls, wood, and metal with a distressed or patinated look. Finally, shop C had a classical, elegant look resembling a formal men's wear department from previous decades.

The neutral zone was located in continuation of the fitting room area, which was shared by the three stores. While it is implicit in our premises that no interior space is completely neutral, what is meant by this is an interior whose atmosphere is outside the purview of this experiment. The interior of the area in question was, however, monochrome and dark, with no decorations or other products on display.



Photo 1 (left) shows Shop A, photo 2 (in the middle) shows Shop B, and photo 3 (right) shows Shop C.

Test products

Six men's tops were selected as test items. All items were priced from 40 to 55 Euros, which also corresponded to the mean price level for similar items in these stores. The same six products were presented in all three test stores and in the neutral zone, so it was necessary to use several copies of each product. Products were presented on front sticks on wall furniture, such that the front of the product faced the customer, and all six products were placed so that they could easily be seen from the center of the store, where the interview would take place. None of the test products were a part of the collections featured in the shops at the time of the experiment, but they were placed among the current collections as though they were. All items were presented on specially made hangers belonging to the stores where they were displayed.

Participants

Of the 50 respondents, 31 were men and 19 were women, aged from 15 to 49 years.

Study

Pre-tests

Five pre-tests were conducted at the test site, in order to determine whether the test was comprehensible to participants, and to fine-tune the procedure. Evaluation of the pre-tests resulted in some slight adjustments to the procedure before proceeding to the 50 actual tests, which are included in this paper. In the pre-test, the six product combinations were presented in the same order in all three shops. Respondents understood the procedure easily enough but were confused by the realization that they were seeing the exact same products in each of the environments. In order to ameliorate this unsettling effect, we changed the order of the combination of products from store to store, to reduce this predictability. We also decided to inform participants at the outset of the interview that they would be presented with different fashion items, and that some of the items would be presented several times.

Procedure for Product Ratings (Part 1)

In-store Product Ratings

As a warm-up to the rating procedure, each respondent was shown a test sheet to familiarize them with the task. On a test screen, each respondent was asked to indicate his or her preference between two products on a visual analogue scale.

First, the researcher walked the respondent into shop A. While standing in the middle of the store, the researcher presented the respondent with two products at a time, asking the respondent to compare them (Photo 4). The researcher then presented the touch screen with photos of two products in the setting and asked the respondent to indicate his or her preference by moving the cursor to a position that reflected a degree of preference. The six paired comparisons were (product left, product right) $\sim \{(1,2), (2,3), (3,4), (4,5), (5,6), (6,1)\}$; this comprises the statistical design set. Next, the researcher escorted the respondent into shop B, where the procedure was repeated, this time with the products in a different order. The same procedure was then repeated a third time in shop C, with yet another ordering of products. The statistical design set was identical in all three shops; the comparisons were merely shown in a different order in each location. Data collected in this phase will be referred to hereafter as 'Product Ratings' or 'In-Store Ratings'.



Photo 4. In the stores, the research assistant presents the respondent with two products at a time and asks him to compare them.

Results and Discussion - In-store product ratings

The paired comparisons method allows us to determine not only how many respondents prefer, for example, product 6 over product 5, but also reveals the relative strength of the preference for each single respondent. Accordingly, separate preference structure for the six products and for the three stores, were calculated for each respondent. Product ratings were converted into numbers, one for each comparison, using the following method: The distance from the middle of the scale to the mark noted by the respondent was measured, positively to the right and negatively to the left. The observations are denoted $\gamma_{i,j}$, $(i,j) \in D$. It is assumed that the numerical score will increase with the strength of preference for one product over the other product, and that equal but upper site preferences would correspond with equal but upper site ratings. For each subject and each room, the six comparison ratings, $\gamma_{i,j}$, $(i,j) \in D$ were combined into a metric rating scale. Therefore, for each subject and for each room, there exists six γ 's; $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ and γ_6 , corresponding to the six products, so that the expected value, $E(\gamma_{i,j})$, of $\gamma_{i,j}$ has the form $E(\gamma_{i,j}) = \gamma_j - \gamma_i$. The estimation of the γ 's is performed using the least squares method. Preference-scores for the six products are calculated as: scores for product $i = \exp \{\gamma_i\}$.

Because respondents did not rate products individually, but as paired comparisons from within each space, the experiments and data from each of the test sites can be considered independent of the other two. It was therefore decided, that the importance of maintaining a natural flow from one space to the next outweighed the minimal risk of a mere-exposure effect, where a respondent might come to like a product or products more by being shown it multiple times.

Unqualified Product Ratings

In order to determine each respondent's product ratings when not influenced by the test atmospheres, respondents were then escorted to the neutral zone, where they rated each product individually. Chronologically, the neutral setting phase took place after the store ratings to allowed as much time as possible to pass between the two rating phases. In the neutral setting, ratings were collected using a VAS with the parameters 'I don't like at all' on the left, and 'I like very much' on the right. These ratings will be referred to in what follows as 'Unqualified Product Ratings.'

Results and Discussion - Unqualified Product Ratings

Each respondent's unqualified product ratings were converted into numbers by measuring the distance from the middle of the scale (0) to the mark noted by the participant. Positive scores from 1 to +50 to the right of zero and negative scores -1 to -50 to the left of zero. Products with high scores were considered to be products for which the respondent had a high degree of preference or liking.

Consumers buy products they like. To give focus to the results, it was decided to isolate the highest-rated product for each respondent. This was determined by locating each respondent's highest unqualified product rating. This product will hereafter be referred to as the respondent's 'Favorite Product'. Having established each respondent's favorite product, it was possible to compare the three in-store ratings for that product. Fig. 1 shows the ratings given for the favorite products in all three stores. The results show that favorite products on average were rated rather homogeneously in the three stores: The favorite products were rated highest in Store B (mean score 19.8), lowest in Store A (mean score 18.0), and slightly higher in Store C (mean score 18.3).

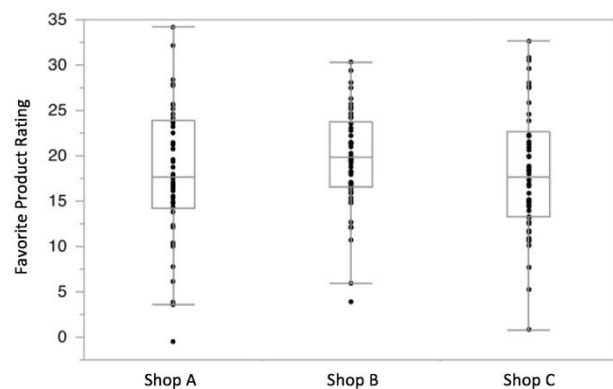


Figure 1. This Box-Plot covers each respondent's Favorite Product rating in store A, B and C.

However, when directing the focus on the one store where the very highest in-store rating of the favorite product was given, a more nuanced picture comes to view. Fig. 2. shows the highest in-store rating for each favorite product. Considering favorite product ratings in all three environments, shop A had the lowest impact on product ratings. In contrast, when isolating the store environment where the favorite products were rated the very highest, differences occurred: The mean score given in store A is now the highest (25.8) but is based on only 9 responses. The mean score in store C is almost as high (25.2), and is based on 13 responses, while the mean score in store B has gone from highest to lowest (21.7), but is based on a much larger sample size, namely 28 responses.



Fig. 2. This Box-Plot covers the highest in-store rating of each favorite product.

These two analyses show that we can in fact observe a measurable spillover effect from store design onto product ratings. We can see that the majority of favorite products (28 of 50) are rated most highly in store B, while relatively few respondents (9 of 50) rated their favorite product highest in store A. However, favorite product ratings given in store A are discernibly higher than those given in the other environments. This is a compelling observation, showing that the proposed method allows us to understand specific tendencies within a population, which might be of interest to retailers and design managers. If, for example, the group of respondents who rated their favorite products highest in store A turned out to be trendsetters or very loyal customers, design managers could make a solid argument for paying more attention to the ratings given by this group of respondents in particular, and for paying more attention to design cues in this specific setting. Thus, data collected using this method can describe and differentiate tendencies that might otherwise be obscured by aggregate methods, and thereby form the basis of new design strategies.

Favorite products are now established, and it was indicated how the different store environments impacted the ratings. But in order to answer the question 'Do consumers rate products more highly when they are presented in stores which they find attractive?' we will need to compare product ratings with ratings of the interior spaces themselves.

Procedure for Store Ratings (Part 2)

After rating the products in the stores, but before rating the products in the neutral zone, each respondent was returned to the stores, one after the other. At the entrance of each store, the researcher asked the respondent to look into the store and describe the interior of the store in their own words. This task was intended to focus the respondent's attention on the environment, as opposed to the products. These responses were typed by the researcher on the touch screen. For example, one respondent gave the following description of store A: 'white, cold, modern'; while another respondent used the words 'industrial, factory, bright'. Next, respondents were asked to rate each interior according to his or her personal taste. These ratings were taken on a VAS where 'I don't like at all' appeared on the left extreme, and 'I like very much' on the right. These ratings will be referred to in the following as 'Store Ratings.'

At the end of the interview, respondents were asked to provide their gender and age, and finally asked whether they would ever consider buying clothes from any of the stores in focus. The purpose of the last question was to eliminate participants who might have been in the stores for reasons other than shopping. This information was also typed on the touch screen and stored along with the data for each individual.

Results and Discussion - Store Ratings

Store ratings were converted to numbers by measuring the distance from the middle of the scale (zero) to the mark noted by the respondent. Positive scores (to the right of zero) from 1 to 50 were marked to the right of the midpoint, and negative scores (to the left of zero) from -1 to -50 to marks placed on the left. A high score was interpreted as a high degree of preference or liking for that store. Fig. 3 shows store ratings for all respondents. Store B (mean score 30.1) and store C (mean score 30.8) were rated highest, while store A (mean

score 21.1) was rated lowest (fig. 3). Store C was not only rated highest, but respondents were more closely in agreement in their ratings of store C than store A.



Fig 3. Box-Plot of the expressed store ratings for store A, B, and C.

Interestingly, store ratings fluctuated widely between the different periods of data collection. Notably, shop A's rating ranged from a mean rating of 11.7 in the first period to 30.8 in the second, and 21.5 in the third. The mean ratings of shop B and C were more consistent throughout. Since the only change in the stores from period to period was the product collections and their presentation, we surmise that these variations are the result of changes in product presentation. For example, during the first period shop A contained some discounted products, and the merchandise varied more than during the second period, where a professional merchandiser had just organized the collection, which might have left the store with a cleaner impression. This suggests that it is not advisable to evaluate store design in isolation, since stores atmospheres are always subject to some degree to the collection of products displayed within them. Products should therefore always be a part of any evaluation of a store atmosphere.

Comparing Product Ratings and Store Ratings

To find out whether a highly rated store design has a corresponding effect on product perception, the task remains to correlate favorite product ratings with store ratings (model 2).



Model 2. Favorite products were selected, and the study was delimited to focus on in-store ratings of favorite products and store ratings.

Each respondent reported three store ratings, and three favorite product ratings, one from each of the test environments. Linking store ratings with in-store ratings of favorite products in aggregate, we see a significant correlation ($p = .001$) (Fig. 5). However, if we consider the ratings given inside each store separately, differences occur. For example, comparison between favorite product rating and shop ratings from shop A for the same respondent reveals a positive correlation ($p < .001$) (Fig 6). But the same comparison for shop B shows a statistically insignificant correlation, $p = .206$ (fig. 7), and shop C also shows an insignificant correlation of $p = .708$ (fig. 8).

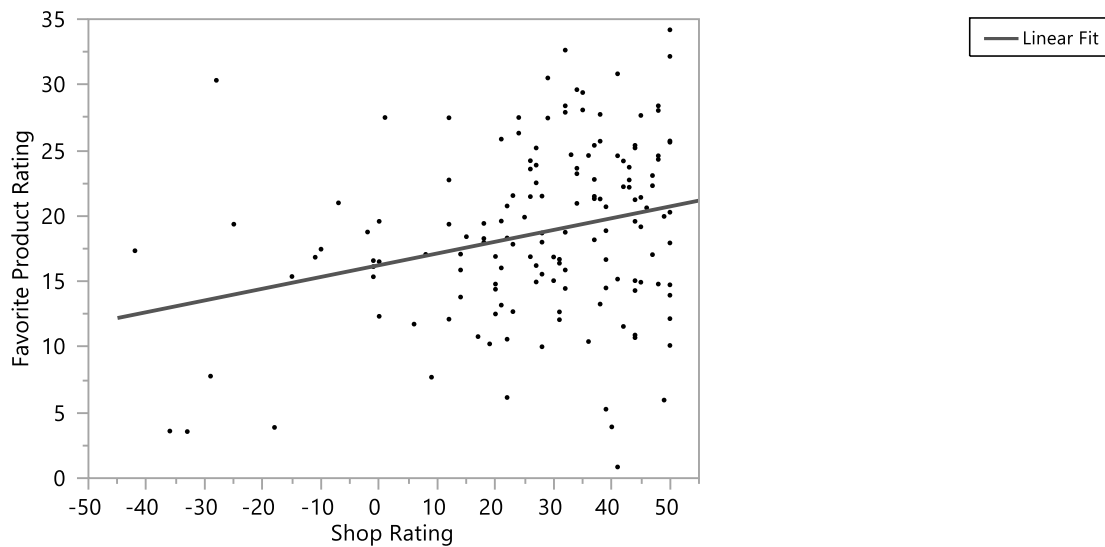


Fig 5. Correlation between Favorite Product Ratings in all stores against Shop Ratings. Linear Fit, Favorite Product Rating = $16,240478 + 0,0895712 \times \text{Shop Rating}$.

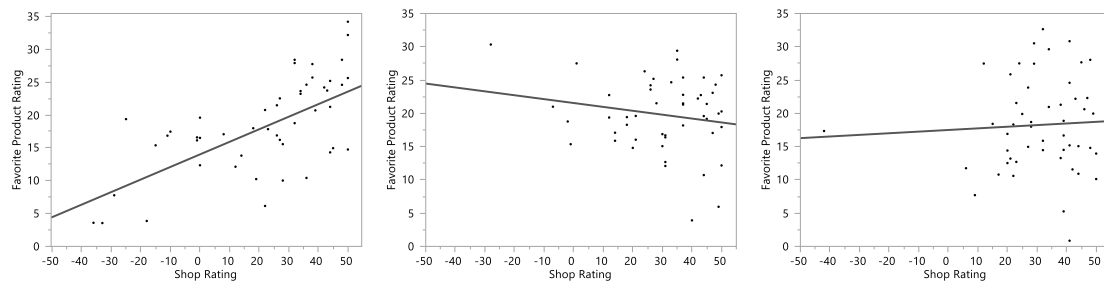


Fig. 6. Left. Significant correlation between Favorite Product Ratings against Shop Ratings in shop A ($p < .001$).

Fig. 7. Middle. Insignificant correlation between Favorite Product Ratings against Shop Ratings in shop B ($p = .206$).

Fig. 8. Right. Insignificant correlation between Favorite Product Ratings against Shop Ratings in shop C ($p = .708$).

In other words, taken in aggregate, a highly-rated store design seems to correspond with product ratings. But isolating the stores makes it clear that some designs influence product ratings more than others. In fact, only in the case of shop A a statistically significant correlation is found between favorite product rating and shop rating. Interestingly, shop A scored lowest among the three interiors in terms of its mean shop rating, (fig. 3), which seems to indicate a discrepancy between consumers' expressed preference for a store environment and that same environment's effect on their responses, at least measured in terms of product preference.

Conclusion and General Discussion

The review of the literature revealed a gap in scholarly understanding of whether a so-called spillover effect exists between store design and product preference. Given that interiors affect people in ways that they cannot necessarily articulate, this study's primary objective was to propose a method for determining whether store design preference spill over onto product preference in a measurable way, and to do this without bringing the respondent's conscious attention to the store interior. A field experiment was designed to

establish the extent to which a preference for three differently designed fashion stores affected the rating of fashion products from within those stores. Results from this project have theoretical, methodological, and managerial contributions to make. Each of these will be discussed separately below. Briefly: on the theoretical level, we expand theories dealing with the supportive aspect of retail design, and spillover effects; methodologically, we contribute to the study of the unconscious effects of interior spaces, and underscore the value of considering consumer data on the individual level as opposed to the aggregate; managerially, this project has implications for design managers and retailers, who can begin to understand the importance of how their designs are experienced by actual users, and how they might go about gathering this valuable information.

Theoretical Implications

The present study shows that individual consumers do indeed rate the same products differently in different contexts. From this we can conclude that there is a measurable spillover effect from context to merchandise, and that this effect has a relation to the design of the interior. Considering respondent data in aggregate, a significant correlation between preferred store design and product preference exists. But considering results from the individual test stores shows that store preference supported favorite products with statistical significance in only one of three stores. Thus, a well-matched interior design can have a supportive or enhancing effect on product preference. Results also indicate that the existence of a supporting effect from an interior is not something that the average consumer is aware of. The store that influenced the product ratings most positively, store A, was not the store that consumers rated the highest in aggregate; stores B and C were both rated higher, in fact. In other words, the enhancing effect does not necessarily go hand-in-hand with preferences that consumers express, which suggests that the practice of relying on consumer opinion, or on direct interrogation as a method of inquiry, should be viewed with circumspection.

Methodological Implications

Methodologically, the approach used in this study allowed us to investigate the extent to which atmospheric context and cues affected specific product ratings without directing the respondents' attention to the contexts themselves. Results showed a measurable effect. Repeated individual measurements enabled studying how each product's ratings were affected within various interiors, a methodological approach that made it possible to measure how much a particular store design influenced specific product ratings. This information cannot be gleaned from studying average or aggregated ratings. Focusing on favorite products revealed that most participants rated their favorite product highest in store B, but that the mean product rating in store B was lower than that in both store A and store C. In this way it is clear that our method of data collection and analysis allowed a more nuanced picture to emerge. Such information could enable decision-makers to define more specific consumer segments, allowing them to target these in productive ways.

Aggregated results indicated a positive correlation between store preference and product preference. On the surface, this seems to indicate that customers did in fact rate products higher in shops that they liked. But our individual measurements enabled the analysis of data from each store separately, which revealed that the effect was produced by only one of the three interiors, shop A. This observation underscores the risk of treating the average customer as representative of the market as a whole.

Managerial implications

The research presented here has several implications for retail and design managers. First and foremost is valuable information about how individual consumers experience retail environments, and how to access this information, which can be of obvious benefit in making strategic decisions about store design. For example, is it more profitable for a given brand to create a design that will appeal strongly to a smaller group of highly engaged customers, or is a broader, more lukewarm appeal to a larger consumer segment the better approach? In this regard, shop A was an interesting case: Individuals who preferred shop A also rated their favorite products higher in its interior. In fact, this effect was strong enough to have a significant effect on the average of all three stores. We can safely conclude that shop A's interior did indeed have a positive influence on product preference. Interestingly, shop A scored the lowest store rating of the three stores, which seems to indicate that managers should be very cautious about relying on consumer interviews to gather information about a store design. A qualitative study of the respondents who were most positively affected by shop A might reveal a small group of enthusiastic individuals, who lifted the average disproportionately. If these

specific consumers turned out to be just browsing and not willing to buy, an argument could be made for ignoring them. If, on the other hand, they turned out to be trendsetters, or loyal customers, a solid argument could be made for paying greater attention to their preferences.

Another observation that might be relevant for practitioners is the fact that the same stores were rated differently during the course of the data collection period. Since the store design was not changed during that period, this indicates that stores are perceived differently according to the products on display. In other words, the current collection and the presentation of the merchandise can significantly impact the overall impression of a store design. This is clearly relevant for design managers to bear in mind when evaluating store design concepts. The appearance of a store design cannot be considered in and of itself, as products themselves are an important part of the equation.

Limitations and further research

The method proposed and applied in this study has advantages and disadvantages. Advantages include the fact that real-life studies of store environments provide a more realistic assessment of their effects than artificial settings (Baker et al. 2002; Ballantine, Parsons, and Comeskey 2015; A. M. Kent & Kirby 2009; van Rompay et al. 2012; Spence et al. 2014). Disadvantages include the complications involved with performing research in actual stores, where control of all variables is difficult. Our research team faced challenges in controlling the location of the test products during the experiment; shoppers could—and did—walk away with test items when the researcher wasn't looking. Results also showed variations in store ratings among the different periods of data collection, which could indicate that the store environment as a whole is affected by changes in collections and displays. While this type of variation can be controlled in a laboratory setting, they are impossible to eliminate completely in a real-life, operating store environment. Furthermore, the possibility cannot be eliminated that respondents were distracted in one way or another during the experiment, and the possibility therefore exists that such distraction may have influenced the ratings to some extent. Having said these things, the state of mind present in an actual shopper in an actual store cannot be reproduced in an artificial setting, where the respondent will always know on some level that he or she is participating in an experiment.

Apart from these inherent complications, the following limitations deserve mention. First, the results are based on a field study with three stores, six products and 50 participants. The differences in the design between the three test stores were big enough that a difference was discernible, but it is conceivable—perhaps even likely—that bigger contrasts and wider stylistic variation would produce even bigger contrasts in the results. Next, none of the test products were part of the current collections; they were instead displayed as single items blended in among the collection itself. This somewhat artificial circumstance could also have had an influence, providing a clue for respondents that something was unnatural about the situation. We believe that possibility to be negligible for the results, however, since the same situation obtained in all test stores. But one possibility for making the experiment more realistic would be to select test products from the current collection in the test store(s). Doing so would also make it possible to record respondents' buying intentions in the different stores, which the present study did not address. It would certainly be relevant to know how many of the respondents actually purchased products from the study, with a view to finding out whether a threshold seems to exist where like equals buy. This data would be of obvious interest to retailers. Finally, the study was based on data from one mall, which introduces a geographic and cultural limitation; it was furthermore limited to respondents who were already in the stores, and therefore provides no information about potential customers.

With this background, we can suggest several directions for further research. First of all, it would be useful to establish a connection between highly rated stores and increased sales. Our study indicates that positively rated store atmospheres do in certain cases enhance product ratings but does not establish whether a highly rated store increases sales. Next, additional research on the particular environmental qualities or atmospheric cues that serve to amplify product evaluation is needed in order to determine whether other factors, apart from liking a store design, are at work. For example, it might prove interesting to employ our method to ascertain whether qualities like friendliness or exclusiveness have an influence, or whether specific interior styles affect preferences for specific products. It might be, for example, that specific kinds of atmospheric cues produce a feeling of well-being in consumers, that in turn creates a positive effect on their responses to products in that store, without giving them any explicit preference or liking for the interior itself. Additional research, including more complex, combinatorial studies, is needed in order to categorize the relevant influencing factors and their degree of influence.

The method applied in this study attempts to isolate unconscious effects that spill over from an environment onto objects within that environment and juxtaposes these results with evaluations of the interior gathered by direct interrogation. Combining this type of study with more detailed explicit responses from consumers might produce a more nuanced account of how specific design cues are interpreted, whether the effect is conscious or unconscious, and to what extent. In this way, the method used here can be applied to other variables in order to determine which, if any, variables are more important for consumers. With further development, the method could become an effective aid in choosing between different design solutions. Rapid developments in virtual reality technology make it easy to imagine applications of this method as a tool for choosing between highly detailed interior drafts. Use of methods like the one proposed here in a virtual setting would allow testing of interiors at a much earlier stage, which could save a great many resources.

References

- Areni, Charles S., and David Kim. 1994. "The Influence of In-Store Lighting on Consumers' Examination of Merchandise in a Wine Store." *International Journal of Research in Marketing* 11(2): 117–25.
- Arnheim, Rudolf. 1974. *Art and Visual Perception: A Psychology of the Creative Eye*. 1974. New. University of California Press.
- Baker, J, A Parasurama, D. Grewal, and G. B. Voss. 2002. "The Influence of Multiple Store Environment Cues on Perceived Merchandise Value and Patronage Intentions." *Journal of Marketing*: 120–41.
- Ballantine, Paul W., Richard Jack, and Andrew G. Parsons. 2010. "Atmospheric Cues and Their Effect on the Hedonic Retail Experience." *International Journal of Retail & Distribution Management* 38(8): 641–53.
- Ballantine, Paul W, Andrew Parsons, and Katrina Comeskey. 2015. "A Conceptual Model of the Holistic Effects of Atmospheric Cues in Fashion Retailing." *International Journal of Retail & Distribution Management* 43(6): 503–17.
- Bellizzi, Joseph a., and Robert E. Hite. 1992. "Environmental Color, Consumer Feelings, and Purchase Likelihood." *Psychology and Marketing* 9(5): 347–63.
- Bellizzi, Joseph A, Ayn E Crowley, and Ronald W Hasty. 1983. "The Effects of Color in Store Design." *Journal of Retailing* 59(1): 21.
- Chebat, Jean-Charles et al. 2012. "The Effect of Ambient Scent on Consumers' Perception, Emotions and Behaviour: A Critical Review." *Journal of Business Research* 30(4).
- Clark, Andy. 2006. "Where Brain, Body and World Collide." *Journal of the American Academy of Arts and Sciences* 127(2): 257–80.
- Custers, P., Y. de Kort, W. IJsselsteijn, and M. de Kruiff. 2010. "Lighting in Retail Environments: Atmosphere Perception in the Real World." *Lighting Research and Technology* 42(3): 331–43.
- Damasio, Antonio R. 1995. *Relations industrielles Descartes' Error: Emotion, Reason, and the Human Brain*.
- Donovan, Robert J., John R. Rossiter, Gilian Marcoolyn, and Andrew Nesdale. 1994. "Store Atmosphere and Purchasing Behavior." *Journal of Retailing* 70(3): 283–94.
- Donovan, Robert J, and John R Rossiter. 1982. "Store Atmosphere: An Environmental Psychology Approach." *Journal of Retailing* 58(1): 34.
- Frijda, Nico H. 1989. "Aesthetic Emotions and Reality." *American Psychologist*: 1546–47.
- Frontczak, Monika, and Pawel Wargocki. 2011. "Literature Survey on How Different Factors Influence Human Comfort in Indoor Environments." *Building and Environment* 46(4): 922–37.
- Gehl, Jan. 2011. *Life between Buildings: Using Public Space (Original Version: Livet Mellem Husene, 1970)*. Washington: Islands Press.
- Groepel-Klein, Andrea. 2005. "Arousal and Consumer In-Store Behavior." *Brain Research Bulletin* 67(5): 428–37.
- Hagtvedt, Henrik, and Vanessa M. Patrick. 2008. "Art Infusion: The Influence of Visual Art on the Perception

- and Evaluation of Consumer Products." *Empirical Studies of the Arts* XLV: 379–389.
- Haug, A., and M. B. Münster. 2015. "Design Variables and Constraints in Fashion Store Design Processes Article Information : " *International Journal of Retail & Distribution Management* Vol. 43(Iss 9): 831–48.
- Herrmann, Andreas, Manja Zidansek, David E. Sprott, and Eric R. Spangenberg. 2013. "The Power of Simplicity: Processing Fluency and the Effects of Olfactory Cues on Retail Sales." *Journal of Retailing* 89(1): 30–43.
- Heskett, John. 2005. *Design: A Very Short Introduction*. Oxford University Press.
- Kent, A. M., and A. E. Kirby. 2009. "The Design of the Store Environment and Its Implications for Retail Image." *The International Review of Retail, Distribution and Consumer Research* 19(4): 457–68.
- Kent, Tony, and Dominic Stone. 2007. "The Body Shop and the Role of Design in Retail Branding." *International Journal of Retail & Distribution Management* 35(7): 531–43.
- Kirby, A. E., and A. M. Kent. 2010. "Architecture as Brand: Store Design and Brand Identity." *Journal of Product & Brand Management* 19(6): 432–39.
- Knoferle, Klemens M., Eric R. Spangenberg, Andreas Herrmann, and Jan R. Landwehr. 2012. "It the Mix: The Interactive Effect of Music Tempo and Mode on in-Store Sales." *Marketing Letters* 23(1): 325–37.
- Kotler, Philip. 1973. "Atmospherics as a Marketing Tool." *Journal of Retailing* 49(4): 48–65.
- Krackhardt, David. 1992. "The Strength of Strong Ties: The Importance of Philos in Organizations." *Networks and Organizations: Structure, Form, and Action* 216: 216–39.
- Kubovy, Michael. 2000. *8 Encyclopedia of psychology Visual and Design Arts*. New York: Oxford University Press.
- Lawson, Bryan. 2001. *Language of Space*. Architectural Press, Elsevier.
- Leder, Helmut, Benno Belke, Andries Oeberst, and Dorothee Augustin. 2004. "A Model of Aesthetic Appreciation and Aesthetic Judgments." *British journal of psychology* 95(Pt 4): 489–508.
- Lynch, Jr., John G. 1982. "On the External Validity of Experiments in Consumer Research." *Journal of Consumer Research* 9: 225–40.
- Martindale, Colin, and Kathleen Moore. 1988. "Priming, Prototypicality, and Preference." *Journal of Experimental Psychology: Human Perception and Performance* 14(4): 661–70.
- Mattila, Anna S., and Jochen Wirtz. 2001. "Congruency of Scent and Music as a Driver of In-Store Evaluations and Behavior." *Journal of Retailing* 77(2): 273–89.
- Milliman, Re. 1982. "Using Background Music to Affect the Behavior of Supermarket Shoppers." *The journal of Marketing* 46(3): 86–91.
- Morin, Sylvie, Laurette Dubé, and Jean-Charles Chebat. 2007. "The Role of Pleasant Music in Servicescapes: A Test of the Dual Model of Environmental Perception." *Journal of Retailing* 83(1): 115–30.
- Newman, Andrew J., and Gordon R. Foxall. 2003. "In-store Customer Behaviour in the Fashion Sector: Some Emerging Methodological and Theoretical Directions." *International Journal of Retail & Distribution Management* 31(11): 591–600.
- North, Adrian C., Lorraine P. Sheridan, and Charles S. Areni. 2016. "Music Congruity Effects on Product Memory, Perception, and Choice." *Journal of Retailing* 92(1): 83–95.
- Petermans, Ann, Anthony Kent, and Koenraad Van Cleempoel. 2014. "Photo-Elicitation: Using Photographs to Read Retail Interiors through Consumers' Eyes." *Journal of Business Research* 67(11): 2243–49.
- Quartier, Katelijjn, Jan Vanrie, and Koenraad Van Cleempoel. 2014. "As Real as It Gets: What Role Does Lighting Have on Consumer's Perception of Atmosphere, Emotions and Behaviour?" *Journal of Environmental Psychology* 39.
- Ramachandran, V. S., and William Hirstein. 1999. "The Science of Art A Neurological Theory of Aesthetic Experience." *Journal of Consciousness Studies* 6(6/7): 15–51.
- Reber, Rolf, Norbert Schwarz, and Piotr Winkielman. 2004. "Processing Fluency and Aesthetic Pleasure: Is

Beauty in the Perceiver's Processing Experience?" *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc* 8(4): 364–82.

- van Rompay, T. J. L., K. Tanja-Dijkstra, J. W. M. Verhoeven, and a. F. van Es. 2012. "On Store Design and Consumer Motivation: Spatial Control and Arousal in the Retail Context." *Environment and Behavior* 44(6): 800–820.
- Spangenberg, Eric R., Bianca Grohmann, and David E. Sprott. 2005. "It's Beginning to Smell (and Sound) a Lot like Christmas: The Interactive Effects of Ambient Scent and Music in a Retail Setting." *Journal of Business Research* 58(11 SPEC. ISS.): 1583–89.
- Spence, Charles, Nancy M. Puccinelli, Dhruv Grewal, and Anne L. Roggeveen. 2014. "Store Atmospheric: A Multisensory Perspective." *Psychology and Marketing* 31(7).
- Thurstone, L.L. 1928. "Attitudes Can Be Measured." *American Journal of Sociology* 33(4): 529–54.
- Turley, L.W., and Jean-Charles Chebat. 2002. "Linking Retail Strategy, Atmospheric Design and Shopping Behaviour." *Journal of Marketing Management* 18(1–2): 125–44.
- Turley, L.W., and Ronald E. Milliman. 2000. "Atmospheric Effects on Shopping Behavior." *Journal of Business Research* 49(1990): 193–211.
- Tversky, Barbara. 2008. "Embodied and Situated Cognition." *Artificial Life* 11(1–2): 5–11.
- Wright, Benjamin D. 1997. "Measurement." *Educational Measurement: Issues and Practice*: 33–45.
- Yalch, Richard F, and Eric R Spangenberg. 2000. "The Effects of Music in a Retail Setting on Real and Perceived Shopping Times." *Journal of Business Research* 49(2): 139–47.
- Zhao, Mingjie, Yang-Seon Kim, and Jelena Srebric. 2015. "Occupant Perceptions and a Health Outcome in Retail Stores." *Building and Environment* 93: 385–94.