# Lost in the Store: Assessing the Confusion Potential of Store Environments

**Marion Garaus**[1] · **Udo Wagner**[2]

**Abstract** Confusion in retailing has attracted increasing attention in the literature. Most of the extant studies concentrate on product-related confusion or the mental state retail shopper confusion. Nevertheless, several recent studies have emphasized the relevance of exploring confusion as an objective property of the store environment. Drawing on this literature stream, the present research conceptualizes the construct store environmental confusion and its six formative dimensions, of which each can be measured through two environmental properties: complexity and conflict. This research seeks to develop, validate, and test a parsimonious index for store environmental confusion. The predictive and nomological validity of the store environmental confusion index is assessed by structural equation modelling and results confirm the hypothesis that store environmental confusion produces undesirable consumer intentions. This research is the first to quantitatively assess the confusion potential of different design factors. The resulting store environmental confusion index can be used to evaluate the confusion potential of various store environments, thereby helping retailers provide customers with a clear and non-confusing store design.

M. Garaus
marion.garaus@modul.ac.at

✉ U. Wagner
udo.wagner@univie.ac.at

1 Department of International Management, modul University Vienna, Am Kahlenberg 1, 1190 Vienna, Austria

2 Department of Marketing, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

🖄 Springer

**JEL-Classification Code**  M31

## 1 Introduction

More than 20 years ago, Pine and Gilmore (1998) introduced the term experience economy to describe consumers' increasing desire for experiential value in retailing and services. Since then, marketers and retailers have identified numerous ways to create extraordinary shopping experiences. However, this strategy might backfire since highly arousing store environments likely result in confusion (Bitner 1992). In recent years, scholars and marketers have recognized the confusion potential of hyper-arousing retail store environments. Forbes magazine notes that most new technologies that seek to revolutionize the in-store shopping experience for consumers "may actually make the customer experience more complex or confusing" (Varon 2016). Similarly, academic research demonstrates that poor signage (Bitner 1992; Mitchell et al. 2005), a misleading visual merchandising strategy (Mitchell and Papavassiliou 1997), poorly crafted in-store experiences (Beverland et al. 2006), electronic price tags (Brüggen et al. 2011), and a confusing store layout (Baker et al. 2002) can all lead to shopper confusion.

The relevance of confusion during shopping situations is well reflected in efforts to develop scales that assess confusion. For example, extant scales in this field consider what consumers feel during confusing shopping situations (Garaus and Wagner 2016) or measure the personality trait "consumer confusion proneness" (Walsh et al. 2007). However, the only research that explicitly concentrates on store-related confusion is the work of Garaus et al. (2015). Using a qualitative survey, the authors categorized various in-store elements that exhibit confusion potential into social, design and ambient factors. Nonetheless, the focus of their work was to explore experienced retail shopper confusion, defined as "as a three-dimensional, temporary mental state consisting of the cognitive effort necessary to deal with confusion (cognition), emotions reflecting the discomfort associated with confusion (emotion), and restricted behavioral intentions (conation)" (Garaus et al. 2015, p. 1004).

The high relevance of confusion in shopping situation manifests not only in scale development efforts, but also by the identification of negative consequences of confusing store environments. Bitner (1992, p. 63) states that "unpleasant environments that are also high in arousal (lots of stimulation, noise, confusion) are particularly avoided." Furthermore, empirical studies demonstrate undesirable consumer responses to confusing shopping environments, including decreases in: shopping value (Garaus et al. 2015), unplanned expenditures, in-store exploration, repeat visit intentions, store patronage intention, and spending time (Garaus and Wagner 2016). Despite the acknowledged importance of environmentally induced confusion (see also Mitchell et al. 2005), elements and properties that determine confusing store environments have received relatively little attention in the extant literature and no efforts have yet been made to develop a valid and reliable measure of confusion evoked by store environments.

In light of the enormous budgets that are spent on store design, a better understanding of the potential negative effects of specific store factors and their properties

would offer retailers the opportunity to build store environments that produce less (or even no) confusion potential. A valid, parsimonious measure of confusion evoked by the store environment would allow retailers to continuously monitor the confusion potential of store environments.

Against this backdrop, the study has three research objectives. First, this study seeks to present a conceptualization of store environmental confusion (SEC), wherein SEC is defined as a second-order, formative construct. Second, this research develops and empirically tests a parsimonious instrument that measures SEC through specific environmental properties. In doing so, it creates a more nuanced view on the confusion potential of store environments. Third, the nomological and predictive validity of SEC is established by embedding and testing this new construct within a nomological research framework by relating it to feelings of confusion, which in turn result in avoidance behavior.

The present research contributes to the extant confusion literature by developing, validating, and testing a parsimonious measure for SEC on manifest properties of the store environment. In line with prior research on store induced confusion, this study shows that SEC results in avoidance behavior.

The paper continues with a literature review, where SEC is differentiated from similar constructs. In doing so, it combines various research streams on confusion and highlights the lack of a measurement instrument for assessing SEC, thereby justifying the need for a new scale. The article then follows established index development guidelines. A general discussion, theoretical and managerial implications, and limitations as well as suggestions for further research conclude this paper.

## 2 Literature Review

Ever since Friedman (1966) first proposed that misleading unit price information will evoke confusion during shopping, the construct of consumer confusion has taken on increasing importance in the consumer behavior literature. Early research on consumer confusion concentrated on the causes of product- and market-related confusion (Mitchell and Papavassiliou 1997; 1999). Research revealed that missing or unclear product instructions (Mitchell and Papavassiliou 1997; Schweizer 2004; Walsh et al. 2007), unclear product pricing (Mitchell & Papavassiliou 1999), an ambiguous brand image (Mitchell and Papavassiliou 1997), and ambiguous front of pack labels (Leek et al. 2015) all cause confusion. Moreover, complexity in general seems to evoke confusion (e.g., complex technology (Drummond 2004) or the complexity of products (Kasper et al. 2010; Leek and Chansawatkit 2006; Leek and Kun 2006)).

Walsh and colleagues (Walsh et al. 2007; Walsh and Mitchell 2010) provided the first formal conceptualization of consumer confusion. These authors conceptualized consumer confusion proneness as personality trait and a product-related construct by relating it to: "how easily/often consumers experience this state of confusion or as consumers' general tolerance for processing similar, too much or ambiguous information, which negatively affects their information processing and decision-making abilities" (Walsh and Mitchell 2010, pp. 839–840). In contrast, Schweizer (2004,

p. 29) concentrates on the confusion potential of shopping situations by defining consumer confusion as an "emotional state that makes it difficult for consumers to select and interpret stimuli." In a subsequent conference paper, Schweizer et al. (2006) developed a 25-item scale reflecting the six consumer confusion properties variety, novelty, complexity, conflict, comfort, and reliability. However, only five items consider SEC factors; the other items relate to product-induced confusion.

More recently, Garaus and Wagner (2016, p. 3461) published a scale measuring the negative feelings that constitute retail shopper confusion. The authors define retail shopper confusion as "a three-dimensional, reflective second-order construct, consisting of the three reflective first-order dimensions", represented by feelings: (1) inefficiency, which captures the degree to which cognitive processing abilities are exceeded; (2) irritation, which represents affective feelings regarding the discomfort associated with retail shopper confusion; and (3) helplessness, which describes the restriction in behavioural intention. Even though the authors point to the confusion potential in store environments, the scale does not allow for assessing the sources of confusion within environments.

The most valuable contribution to understanding SEC has been published by Garaus et al. (2015). The authors offer initial evidence on the confusion potential of the store environment through use of a qualitative survey. Based on these data, they classify 64 confusion triggers. In particular, the researchers define the confusion potential of ambient, design and social factors by referring to four environmental properties (variety, novelty, complexity, and conflict). However, no efforts were undertaken to develop a reliable and valid measurement instrument for assessing SEC.

This review of existing scales to measure confusion (proneness) reveals the literature lacks a scale for assessing SEC. The absence of such a measurement instrument also implies that it remains unclear which specific store factors evoke considerable confusion, why they evoke confusion, and how retailers can assess and avoid such confusion triggers.

## 3 Store Environmental Confusion Measurement Instrument Development

We propose that specific environmental properties of store design factors represent suitable measures for assessing SEC triggers. By way of illustration, the theoretical underpinnings of this research suggest that the two environmental properties complexity and conflict of the store design factor signage reflect the SEC dimension "signage confusion". Formative index procedures uncover the full scope of the SEC construct and identify and validate six SEC dimensions (aisles, customer flow, shelving and storage, signage, space allocation, visual merchandising). Fig. 1 outlines the four stages of the index development. Stages 1 and 2 are concerned with conceptual issues and hence draw primarily upon theoretical considerations. However, the highly fragmented literature on SEC further requires empirical data (i.e., initial survey) to help offer a more precise specification of SEC dimensions
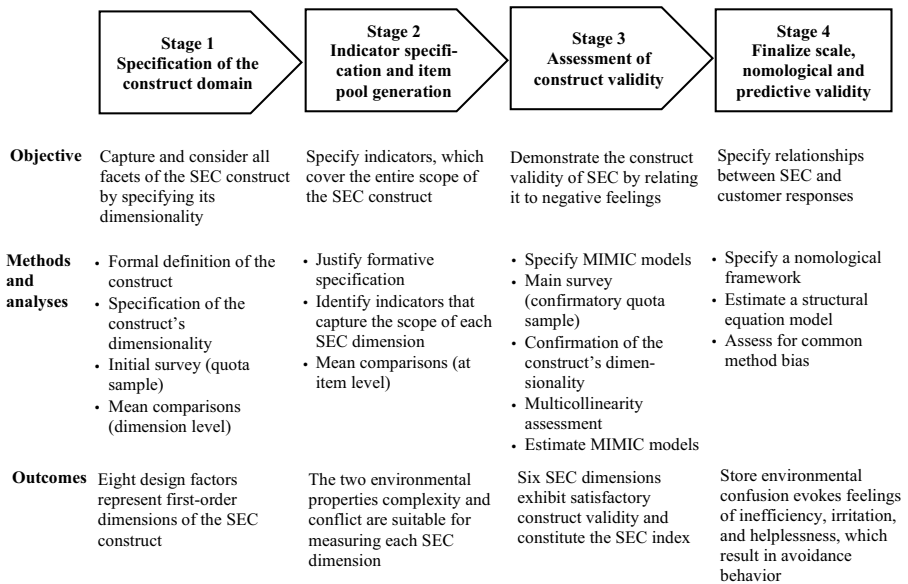
| | **Stage 1** Specification of the construct domain | **Stage 2** Indicator specification and item pool generation | **Stage 3** Assessment of construct validity | **Stage 4** Finalize scale, nomological and predictive validity |
|---|---|---|---|---|
| **Objective** | Capture and consider all facets of the SEC construct by specifying its dimensionality | Specify indicators, which cover the entire scope of the SEC construct | Demonstrate the construct validity of SEC by relating it to negative feelings | Specify relationships between SEC and customer responses |
| **Methods and analyses** | • Formal definition of the construct<br>• Specification of the construct's dimensionality<br>• Initial survey (quota sample)<br>• Mean comparisons (dimension level) | • Justify formative specification<br>• Identify indicators that capture the scope of each SEC dimension<br>• Mean comparisons (at item level) | • Specify MIMIC models<br>• Main survey (confirmatory quota sample)<br>• Confirmation of the construct's dimensionality<br>• Multicollinearity assessment<br>• Estimate MIMIC models | • Specify a nomological framework<br>• Estimate a structural equation model<br>• Assess for common method bias |
| **Outcomes** | Eight design factors represent first-order dimensions of the SEC construct | The two environmental properties complexity and conflict are suitable for measuring each SEC dimension | Six SEC dimensions exhibit satisfactory construct validity and constitute the SEC index | Store environmental confusion evokes feelings of inefficiency, irritation, and helplessness, which result in avoidance behavior |

**Fig. 1** Index Development Stages, Objectives, Methods and Analysis, and Outcomes

and indicators that measure these dimensions. Stages 3 and 4 validate the previous conceptualizations using data collected in a large-scale quota-based survey.

Thus, stage 1 starts with a specification of the construct domain and determines the number of dimensions. Stage 2 identifies the indicators that measure these dimensions. Again, the conceptualization of formative indicators is guided by theory, and empirical data is used to create an initial item pool. Stage 3 assesses the validity of each SEC dimension by estimating eight multiple indicators multiple causes (MIMIC) models and verifies the dimensionality of each SEC dimension. Finally, stage 4 assesses the nomological and predictive validity of the instrument by providing statistically sound and meaningful estimates of measurement and structural parameters.

### 3.1 Stage 1: Specification of the Construct Domain

#### 3.1.1 Theoretical and Conceptual Considerations

**Theoretical Considerations** The specification of a construct domain is important for capturing and considering all of the domain's facets (Diamantopoulos and Winklhofer 2001). Stage 1 in the scale development process thus strove to conceptualize the dimensionality of the SEC construct. Prior qualitative research identified design factors, ambient factors, and social factors as confusion sources in retail environments (Garaus et al. 2015). These store factors draw on Baker's (1987) classification of the store environment. However, alternative classifications exclude social factors, as social interactions often occur as a consequence of store particularities (Bitner 1992; Donovan and Rossiter 1982). For instance, music can influence the desire to

interact with sales staff (Dubé et al. 1995) and store design and store layout can create spatial crowding (Mehta 2013). In line with these considerations, social factors were not considered subsequently in this study.

Moreover, Baker (1987) suggests that customers perceive ambient factors unconsciously. They take such conditions for granted and become aware of them only in the case of their absence or malfunction. Furthermore, research indicates that consumers are not aware of the effect of music on their behavioral intentions (North et al. 1997). Similarly, scent influences both cognition and behavior unconsciously (Holland et al. 2005). The unconscious nature of ambient factors makes the retrospective measurement of its confusion potential impossible. Hence, ambient factors disqualify for inclusion in the SEC index. The concentration on store design factors is also in line with the objective of this research, namely to develop a parsimonious measurement instrument for SEC; therefore, the present study focuses on the physical store environment by considering only design factors for the development for the SEC index.

The previously discussed literature suggests that eight design factors (aisles, architecture, customer flow, shelving and storage, signage, space allocation, technology, and visual merchandising) might constitute SEC (cf. Garaus et al. 2015). We label these factors with the suffix "confusion." For instance, the confusion potential of the design factor "signage" is captured by "signage confusion". As such, each of the eight design factors represent a defining component of the construct SEC. They assess customers' perceptions of the confusion potential of each respective design factor and are not distinct entities from SEC, but rather belong to the same entity. Thus, SEC cannot occur without any confusion potential among the eight design factors. Accordingly, they are conceptualized as SEC dimensions, and thus are conceptually different from exogenous causes of the construct SEC (cf. Lee et al. 2013). Exogenous causes would rather relate to retail management decisions, such as the installation of new technologies or the complete redesign of a store.

In sum, we conceptualize SEC as a formative construct, as each dimension represents an important and unique aspect of the higher-order construct SEC (Bollen and Lennox 1991). The formative conceptualization of the construct further implies that the confusion dimensions do not necessarily need to correlate with each other (Diamantopoulos et al. 2008). Hence, in a particular store, one design factor (e.g., signage) may bear high confusion potential while another design factor (e.g., visual merchandising) may possess limited confusion potential. A store environment is perceived as very confusing when it scores highly on all of these dimensions.

Despite identifying eight design factors that likely represent SEC dimensions, no empirical data are available so far suggesting which design factor, if any at all, indeed possesses confusion potential. Accordingly, it is not clear if all eight design factors are suitable to capture the SEC construct. Due to its newness, no theoretically-based selection criterion exists to select certain design factors to represent SEC dimensions while eliminating others.

**Preliminary Conceptual Considerations Regarding Indicators of SEC Dimensions** In line with the multidimensional perspective of SEC (Garaus et al. 2015), this research proposes that it is not the existence of design factors themselves, but

rather the specific structural environmental properties variety, novelty, complexity, and conflict (Mehrabian and Russell 1974) that bear confusion potential. These environmental properties serve as potential indicators of each environmental confusion dimension. Variety (or information density; Cupchik and Berlyne 1979) refers to stimuli associated with uncertainty, such that a wider range of alternative possibilities results in greater uncertainty. In product-related confusion, the variety of products represents a major trigger for confusion (Walsh et al. 2007). Novelty captures new, unusual, unfamiliar, mystical, changing, or surprising stimuli (Berlyne 1960). Complexity describes the degree of variation within a stimulus pattern (Nasar 1987), such that it does not reflect the *number* of elements in an environment (variety) but rather the *specific patterns* of these elements, their dissimilarity, and the potential for processing the elements as one unit (Berlyne 1960). Herrmann et al. (2013) demonstrate that complex environmental stimuli inhibit information processing. Finally, conflict is the opposite of coherence, and refers to the assignment of two or more stimuli with different meanings (Nasar 1987).

In this regard, it is important to consider the relationship between these environmental properties, arousal, and affective evaluations. Research acknowledges that the relationship between arousal and affective evaluations (i.e., with respect to the store environment in the present context) follows a curvilinear relationship, with most favorable (i.e., least negative) evaluations experienced at moderate levels of arousal. Low arousal levels are experienced as unpleasant and evoke feelings of boredom (Berlyne 1960; Di Muro and Murray 2012), while intense levels of arousal evoke feelings of confusion (Bitner 1992). Consumers perceive nothing positive in confusion and confusion evokes negative feelings only (Garaus and Wagner 2016). Based on this evidence, we postulate that (substantial) SEC requires higher levels of arousal (i.e., levels exceeding the optimal arousal level). Thus, the present research will focus on higher levels of arousal[1] which further permits assuming linearity between confusing design factors and environmental properties.

### 3.1.2 Empirical Validation

An initial data collection (data set I) intended to clarify the role of design factors representing SEC dimensions. Accordingly, a survey attempted to quantitatively assess the confusion potential of the eight design factors (i.e., aisles, architecture, customer flow, shelving and storage, signage, space allocation, technology, and visual merchandising) (Baker 1987; Bitner 1992). Conducting this data collection required specifying indicators that measure these factors (operationalized by the environmental properties variety, novelty, complexity, conflict). The quantitative assessment of the confusion potential of each of these factors sought to empirically verify the theoretically derived environmental confusion dimensions.

The questionnaire included examples for all 4 × 8 combinations (32 items) of the four properties that measure the eight design factors (e.g., for the design factor signage, the question for assessing its variety read, "The environmental property variety, e.g., lots of signage, different sizes, colors or shapes of signage, evokes

---

[1] Sect. 3.3.2. will deal more thoroughly with the concrete threshold specified in this research.

confusion[2]"). As response format the questionnaire offered a five-point Likert scale for each of the 32 items. Twenty respondents pretested the preliminary questionnaire for comprehensibility and logical flow.

Twenty students received course credits for interviewing at least ten respondents each based on a predefined quota. The sample was representative of the national population of the European country of investigation in terms of age, gender, and education (52% females; age: 16% 15–24 years, 16% 25–34 years, 19% 35–44 years, 17% 45–54 years, 12% 55–64 years, 10% 65–74 years, 10% older than 75 years; education: 11% university, 12% high school, 14% vocational school, 33% apprenticeship, 30% compulsory school).

Respondents qualified for the survey if they had been living in this country for at least three months and shopped regularly for groceries. The high number of interviewers combined with quota specifications enabled questionnaire distribution to a wide variety of respondents while simultaneously representing shoppers typical of the country under investigation. This approach, in turn, guarantees that respondents had varied shopping experiences that are essential for identifying the confusion potential of a broad range of shopping environments.

The first page of the questionnaire introduced respondents to the research topic (exploring the confusion potential of the four environmental properties of the store design factors). A short paragraph highlighted that the research focused on confusing store environments, not on other store characteristics. The instruction for all questions read as follows: "Please indicate which of the following properties of '*design factor*' evoke confusion in a shopping situation." Afterwards, for each design factor the following statement appeared: "The following properties of '*design factor*' in a store evokes confusion in a shopping situation." Subsequently, respondents indicated the extent to which they agreed that these design factors in combination with the four properties described the confusion potential of store environments (five-point rating scale). For instance, for the design factor aisles, the item assessing the confusion potential of the environmental property complexity reads: "complexity (e.g., no overview of where to find products)" and for the environmental property conflict the item reads: "conflict (e.g., too narrow or wide aisles, barriers, difficult access to other aisles, bad transition to other aisles)" (please see Table 5).

The questionnaire did not provide any stimuli of specific confusing store environments but showed some small Cliparts from the Microsoft Office about typical shopping situations (showing women doing the grocery shopping and clothes shopping) to make the questionnaire more vivid and appealing. After eliminating obvious response errors, 214 questionnaires qualified for further analysis.

Because participants might have experienced fatigue when answering 32 items, leading to satisficing response behavior (Weijters and Baumgartner 2012), initial analyses checked for response bias. First, we examined the standard deviations of each subject's responses; postulating that the responses would follow a discrete uniform distribution resulted in a benchmark of $\sqrt{2}$ for the standard deviations. Nearly constant evaluations of items (implying less diligent response behavior) would produce smaller standard deviations. As a consequence, elimination of nine question-

---

[2]  The complete list of items used in the questionnaire is available upon request from the authors.

**Table 1** Evaluation of the Confusion Potential of Design Factors through Environmental Properties (based on data set I)

| Design Factors | Properties | | | | Overall Mean | Results of RM ANOVA Contrast $F$-Tests | | | | | | | |
| | Novelty 1 | Variety 2 | Conflict 3 | Complexity 4 | | Groups of Comparison | | | | | | | |
| | | | | | | 1 vs. 3 | | 1 vs. 4 | | 2 vs. 3 | | 2 vs. 4 | |
| | | | | | | $F$ | p-level | $F$ | p-level | $F$ | p-level | $F$ | p-level |
| *Global Mean* | 2.98 | 3.09 | 3.46 | 3.56 | 3.28 | | | | | | | | |
| *Design Factors* | | | | | | | | | | | | | |
| Signage | 2.74 | 3.01 | 3.50 | 3.30 | 3.13 | **39.77** | **<0.01** | **23.72** | **<0.01** | **25.35** | **<0.01** | **8.06** | **0.01** |
| Technology | 3.12 | 2.80 | 3.61 | 3.03 | 3.14 | **20.51** | **<0.01** | 0.84 | 0.36 | **52.37** | **<0.01** | **5.53** | **0.02** |
| Shelving and storage | 2.67 | 2.65 | 3.54 | 3.71 | 3.14 | **61.83** | **<0.01** | **89.31** | **<0.01** | **64.73** | **<0.01** | **85.32** | **<0.01** |
| Architecture | 3.08 | 3.17 | 2.73 | 3.72 | 3.17 | *11.45* | *<0.01* | **42.60** | **<0.01** | *14.54* | *<0.01* | **32.32** | **<0.01** |
| Visual merchandising | 2.54 | 3.12 | 3.75 | 3.61 | 3.26 | **94.40** | **<0.01** | **81.49** | **<0.01** | **35.81** | **<0.01** | **23.63** | **<0.01** |
| Space allocation | 3.35 | 3.04 | 3.55 | 3.41 | 3.34 | **4.31** | **0.04** | 0.42 | 0.52 | **23.50** | **<0.01** | **12.08** | **<0.01** |
| Customer flow | 3.08 | 3.50 | 3.61 | 3.83 | 3.50 | **30.83** | **<0.01** | **75.09** | **<0.01** | 1.26 | 0.26 | **14.98** | **<0.01** |
| Aisles | 3.29 | 3.40 | 3.59 | 3.85 | 3.53 | **6.87** | **0.01** | **27.39** | **<0.01** | 2.61 | 0.11 | **17.44** | **<0.01** |

Entries represent means on five-point scales, evaluated over the sample (1 = "do not agree at all"; 5 = "totally agree"), $n = 205$

Bold figures indicate that mean differences are negative (in accordance with expectations) and statistically significant as indicated by RM contrast test; Italic figures indicate that mean differences are positive (in conflict with expectations) and statistically significant

naires whose responses exhibited standard deviations of less than $\sqrt{2}/2$ occurred; this results in the final sample size of $n = 205$. Next, Mood's (1940) asymptotic multinomial runs test explored the response patterns. Intuitively, this test assesses whether runs occur at random or follow regular patterns. Too few runs would indicate that a respondent evaluated many items in a row the same way, likely from fatigue (Weijters and Baumgartner 2012). In contrast, too many runs could indicate deliberate systematic response behavior. The runs test did not reject the hypothesis that they were random for 76% of the cases. Hence, at least 76% of respondents appeared to conscientiously have evaluated the questionnaires. Finally, conducting an analysis of the correlation coefficients for all items across respondents concluded checking for response biases. Ninety percent of the correlations were less than 0.4, indicating that respondents discriminated diligently among the different items. The left part of Table 1 presents mean evaluations of these 32 items.

The column labelled "Overall Mean" in Table 1 shows the overall means of all eight design factors, where each design factor is measured by novelty, variety, conflict, and complexity. The entries in Table 1 are ordered according to increasing composite evaluations of design factors. The design factor "signage" (3.13) contribute the least to SEC and the design factor "aisles" (3.53) the most. In general, all eight design factors seem to reflect the confusion potential environments quite well. The row labelled "Global Mean" in Table 1 presents means (over design factors) of the environmental properties.

Following the expectation that the scale midpoint value of three reflects neutral responses, potentially confusing design factors should exceed the scale midpoint (cf. Parks and Floyd 1996). One-sample $t$-tests for each design factor against the scale midpoint assessed their confusion potential. These tests revealed all design factors bear considerable confusion potential, with an overall mean being significantly (for a type I error of 5%) higher than 3. This exploratory analysis demonstrates the confusion potential of aisles, architecture, customer flow, shelving and storage, signage, space allocation, technology, and visual merchandising. This empirical validation supports the specification of design factors as SEC dimensions. More formally, at this stage in the scale development process, SEC is conceptualized as an eight-dimensional, formative construct.

### 3.2 Stage 2: Indicator Specification and Initial Item Pool Generation

#### 3.2.1 Theoretical and Conceptual Considerations

Having assessed the dimensionality of the SEC construct, the next step of the scale development process sought to identify indicators used to measure the eight SEC dimensions. This specification formed the basis for the development of an edited item pool.

Prior research emphasizes the relevance of correct indicators. Misspecifying a formative indicator as reflective would not only bias the measurement of the latent construct, but might also affect all other coefficients in a model (Bollen and Diamantopoulos 2017). Moreover, misspecifying formative indicators might lead to indicator exclusion based on traditional reliability statistics, which are not applica-

ble to formative measures and lead to incorrect item elimination criteria (Bollen and Lennox 1991).

Three considerations guided the specification that the indicators of SEC should be formative (cf. Bagozzi and Yi 2012; Diamantopoulos and Winklhofer 2001). First, the indicators should induce variation in the respective SEC dimension. Second, each indicator should capture a specific aspect of a SEC dimension. This postulation implies that—in contrast to reflective measurement specification—the indicators are not interchangeable. Third, these indicators were not required to correlate with each other. Moreover, Diamantopoulos and Winklhofer (2001) emphasize the relevance of identifying indicators that cover the entire scope of the latent variable. This is of particular importance for the present index construction (e.g., when referring to the dimension signage confusion, the indicators assessing this construct need to capture the whole signage confusion potential).

Some literature provides guidance on selecting properties that are suitable for measuring SEC dimensions. Nasar (1987) showed that complexity and conflict can effectively describe the arousing properties of business signs, and high arousal levels have been found to correlate positively with confusion (Bitner 1992; Garaus and Wagner 2016). A recent study confirms that visually complex environments hamper processing fluency (Ketron 2018). However, the theoretical evidence provided by these reports does not sufficiently warrant the inclusion of only complexity and conflict as formative indicators for each SEC dimensions while eliminating novelty and variety. Hence, we follow the advice of Bollen (2011) and employ statistical procedures to investigate the relationships between the four environmental properties for each SEC dimension. This approach will guarantee that the developed SEC index captures only indicators that indeed constitute SEC and likely result in negative consumer responses.

### 3.2.2 Empirical Validation

The empirical validation identified the properties that best capture the confusion potential of each SEC dimension by relying on the same sample as in stage 1 (i.e., data set I). Stage 2, however, concentrated on environmental properties (the row labelled "Global Mean" in Table 1 presents means of novelty, variety, conflict, and complexity) rather than on overall evaluation of design factors. Columns of environmental properties are sorted according to increasing confusion potential: on average, the property novelty (2.98) appears to be of least importance for SEC and the property complexity (3.56) of most importance. The low relevance of novelty is probably due to infrequent changes in store environments. According to the global mean of variety, shoppers did not assess the variety of design factors (e.g., the variety of signage) as particularly confusing. This result is especially noteworthy because variety represents a major antecedent of product-related confusion and indicates that confusing properties have differing importance for products and stores.

The global means of conflict (3.46) and complexity (3.56) suggest these two properties possess major confusion potential. One-sample $t$-tests of global means against scale midpoints were significant for both properties. A repeated measure-

ment ANOVA contrast test reinforces the difference in confusion potential between novelty and variety against conflict and complexity ($F = 107.51$, $p < 0.01$).

The lower left part of Table 1 shows mean confusion of each environmental property for each SEC dimension. The corresponding right part of Table 1 reports the results of pairwise (pairing of environmental properties) repeated measurement contrast tests for each SEC dimension (which resulted in 32 planned comparisons). Thus, these contrasts analyze the superiority of the environmental properties conflict and complexity as compared to variety and novelty at the factor level. The vast majority (81%) of these comparisons is in line with the global findings (i.e., the confusion potential of conflict and complexity exceeds that of novelty and variety) and is statistically significant; four differences are not significant (those indicated by regular font in Table 1), and only two comparisons are not consistent with expectations but statistically significant (italic font in Table 1). Taken together, these results offer further empirical evidence for the suitability of the two environmental properties conflict and complexity for measuring SEC dimensions.

Prior research also demonstrates that conflict and complexity are highly relevant properties in predicting consumers' responses to shopping environments (Deng and Poole 2010; Nasar 1987; Orth et al. 2016), while variety and novelty do not appear to reflect the confusion potential of store environments. From an index construction viewpoint, Bollen (2011, p. 362) point out: "removal of a causal indicator might change the nature of the latent variable." Nevertheless, he also mentions (p. 362) that "if the coefficient of the causal indicator is not significant or is the wrong sign, it is possible that you have an invalid causal indicator. Decisions on whether to eliminate indicators must be made taking account of the theoretical appropriateness of the indicator and its empirical performance in the researcher's and the studies of others." The limited theoretical evidence in combination with this very simple analysis prevent us from relying on only one sample before eliminating formative indicators. Hence, advanced statistical procedures with another sample are used to validate the preliminary finding of stage 2 of this index development process, namely that novelty and variety do not reflect valid indicators for the SEC dimensions (see Sect. 3.3.2).

### 3.3 Stage 3: Assessment of Construct Validity

#### 3.3.1 Theoretical and Conceptual Considerations

While the empirical analysis of this section will offer insights into both, stages 2 and 3 of this index development process, theoretical considerations deal with the assessment of construct validity of each SEC dimension (aisles, architecture, customer flow, shelving and storage, signage, space allocation, technology, and visual merchandising).

In contrast to reflectively measured constructs, construct validity assessment of formative constructs is still challenging. Formative constructs are not identified (Bollen 2011) and treating formative indicators as reflective ones lead to biased coefficients, especially when the formative indicators have low intercorrelations (Jarvis et al. 2003). Recent literature suggests to predetermine the weights based

on theoretical reasoning (Cadogan et al. 2013; Lee and Cadogan 2013), i.e. extant theoretical knowledge determines the contribution (i.e. weight) of each indicator.

However, Howell et al. (2007) list a number of shortcomings of such an approach. In essence, they conclude that the predetermination of formative indicator weights is accompanied with a high potential of loss of information, due to the great variety of possible configurations. Based on this reasoning, they argue that subjective weighting of formative measures is only appropriate when a specific dependent phenomenon is of interest, and little meaning and interpretation is attached to the index itself. In support of these notions, Diamantopoulos (2013) endorses that setting fixed weights does not allow for testing the significance of indicators and Diamantopoulos and Temme (2013, p. 162) report that—compared to alternative approaches that rely on pre-defined weights for indicators or composites of those—"only the MIMIC model's fit could be considered acceptable according to conventional criteria."

Complementing this line of thoughts, we note that the estimation of the weight-parameters of each indicator is essential for assessing construct validity, as they capture the contribution of each formative indicator to the construct. Accordingly, indicators associated with non-significant weight-parameters should be considered for elimination as they "cannot represent valid indicators of the construct" (Diamantopoulos et al. 2008, p. 1215). The estimation of multiple indicators and multiple causes (i.e. MIMIC) models is an alternative to predetermining indicator weights. By doing so, the problem of under-identification of the measurement model is circumvented by adding three reflective indicators (Diamantopoulos and Riefler 2008). This approach is the preferred option for evaluating formative measurement models since it does not require the inclusion of additional constructs for model identification and thus adheres to the requirement of model parsimony. In addition, estimates of measurement parameters are more stable than structural ones (MacKenzie et al., 2005). Employing a MIMIC model transforms a formative measurement model into a function that predicts a linear combination of reflective indicators (Bollen 2011).

The major problem some scholars associate with MIMIC models is that the meaning of the latent variable is not theoretically grounded in the formative indicators but rather empirically based in the covariance between the latent variable and its reflective indicators (Treiblmaier et al. 2011). By including reflective items, the error term in MIMIC models does not depend on the conceptual meaning of the formative model but results from the unexplained variance when trying to predict the common factor of the formatively measured construct and its reflective indicators (Lee et al. 2013). Accordingly, the formatively measured latent construct changes when the number and content of the reflective indicators change.

In contrast, Diamantopoulos suggests that theoretical considerations should guide the interpretation of the MIMIC model. Alternatively, one could interpret the focal construct as being measured by its formative indicators, while "impacting several directly observed variables" (Diamantopoulos 2013, p. 33). We appraise both perspectives as legitimate and pay particular effort to solve the shortcoming of MIMIC models as discussed by Cadogan et al. (2013). We propose that MIMIC models that include reflective indicators of a specific mediating variable overcome these limitations.

In the present research, each formatively measured confusion construct (e.g., signage confusion) is per definition related to the mediating mental state retail shopper confusion. Compelling empirical and theoretical results provide evidence that confusing store elements evoke the negative mental state confusion (Garaus et al. 2015; Garaus and Wagner 2016). As such, the negative feelings associated with confusion (inefficiency, irritation, and helplessness), relate per definition to confusing store elements, while no theoretical arguments exist for any other mediating construct. Accordingly, confusing store elements cannot be modelled with any other set of reflective indicators than retail shopper confusion. This theoretical reasoning also hinders the application of Adams et al.'s (2003) approach, namely considering the formative indicators as independent predictors of a variety of outcomes. Conceptually, confusing store elements do not influence outcomes directly, but through the specific mediating mental state retail shopper confusion.

Based on these considerations MIMIC models appear to be the most suitable approach for assessing construct validity for the eight SEC dimensions. In addition, the procedure undertaken follows Bollen (1989) who suggests eliminating non-significant formative indicators as they do not represent valid indicators of the respective construct. Composite scores of the three feelings associated with confusion (inefficiency, irritation, and helplessness; Garaus and Wagner 2016) qualify as reflective indicators that allow for the estimation of MIMIC models. In particular, each SEC dimension likely evokes the feelings inefficiency, irritation, and helplessness.

### 3.3.2 Empirical Validation

A second descriptive study (i.e., face-to-face interviews; data set II) provided fresh data for this confirmatory analysis. To capture many confusing shopping experiences from a great variety of consumers, a pre-defined quota (age, gender, and education) guided respondent selection. Thirty students in a graduate class on consumer behavior received course credit for interviewing at least twenty respondents each who had resided in the country under investigation for at least three months and regularly shopped for groceries. Students were allowed to interview family members, friends, or colleagues but had to strictly observe their individual quotas. Thus, the overall sample characteristics (age, gender, and education) match the quotas for the country under investigation. Respondents read the questionnaire on their own, but the interviewers could provide clarification in case of questions and ambiguities. The data collection lasted three weeks and resulted in 552 response records (after data cleaning, 53% females; age: 16% 15–24 years, 16% 25–34 years, 18% 35–44 years, 17% 45–54 years, 14% 55–64 years, 10% 65–74 years, 9% older than 75 years; education: 9% university, 14% high school, 16% vocational school, 32% apprenticeship, 29% compulsory school).

Individual interviews were structured as follows: (i) A cartoon stimulus exposed respondents to a confusing shopping situation. The cartoon showed a grocery store with many different and conflicting signs, various acoustic stimuli, a confusing store layout, and four obviously confused shoppers. (ii) The questionnaire started with an assessment of the quota criteria (age, gender, education). (iii) Afterwards, respondents evaluated the eight design factors according to their confusion potential

**Table 2** Piecewise Linear Relationship between Aggregated SEC Dimensions and Aggregated Feelings of Retail Shopper Confusion (based on data set II)

| | $x_s$ assumed continuous | | | $x_s$ assumed discrete, set a priori | | |
| | Estimate[a] | | Confidence interval[b] | Estimate[c] | | $p$-level |
| --- | --- | --- | --- | --- | --- | --- |
| Threshold $x_s$ | 2.10 | – | [1.64, 2.56] | 3 | – | – |
| Slope$_1$ $\alpha_{11}$ | –0.38 | – | [–1.21, 0.46] | 0.31 | – | <0.01 |
| Slope$_2$ $\alpha_{21}$ | 0.53 | – | [0.44, 0.63] | 0.58 | – | 0.05 |
| $R^2$ | 0.21 | $F_{4,540} \approx 34.83 < 0.01$ | | 0.20 | $F_{2,541} = 66.30$ | <0.01 |

[a]by means of nonlinear regression analysis
[b]for a type I error of 0.05
[c]by means of ordinary least squares regression analysis

in terms of variety, novelty, complexity, and conflict. Since these properties might have appeared to be too abstract, concrete examples for each environmental property based on the research of Garaus et al. (2015) were included. For instance, "no overview of where to find products" accounts for assessing aisle confusion for the environmental property complexity (see Table 5). (iv) Another series of items asked for consumers' feelings (inefficiency, irritation, and helplessness) in a confusing shopping situation. (v) Finally, several items inquired about respondents' avoidance behavior in a confusing shopping situation (see Appendix, Table 7).

**Indicator Validation**   The discussion in Sect. 3.1.1. pointed to a potential curvilinear relationship between confusion-induced arousal and (negative) affective evaluations. In the present context, this type of nonlinearity might apply to the relationship between the store environmental design factors and the evoked feelings[3]. In order to keep the investigations tractable, analysis is carried out at an aggregate level (composites over all environmental design factors, i.e., $x$; and over the three types of feelings, i.e., $y$). Low arousal situations (i.e., boredom) are not covered here; therefore, a piecewise linear relationship with only two regimes is postulated:

$$y_t = \begin{cases} \alpha_{10} + \alpha_{11} \cdot x_t & \text{for } x_t \leq x_s \\ \alpha_{20} + \alpha_{21} \cdot x_t & \text{for } x_t > x_s \end{cases} \quad \text{and} \quad \alpha_{10} + \alpha_{10} \cdot x_s = \alpha_{20} + \alpha_{21} \cdot x_s$$

The threshold $x_s$ might be either set to the midpoint of the scale (i.e., to 3) or assumed to be a continuous quantity and thus estimated econometrically. Table 2 presents the estimated regression parameters. The estimated threshold ($\widehat{x}_s = 2.10$) roughly corresponds to the vertex of the functional relationship between arousal and (negative) affective evaluations as outlined in Sect. 3.1.1; consequently, the corresponding slope parameter $\left(\widehat{\alpha}_{11}^{(1)} = -0.38\right)$ is not significantly different from zero (for a type I error of 0.05). Data collection is based on a discrete number of

---

[3]  We gratefully acknowledge the recommendation of a reviewer who suggested performing this insightful analysis.

**Table 3** Correlation and VIF Values of Formative Indicators (based on data set II)

| | VIF | Aisle complex | Aisle conflict | Flow complex | Flow conflict | Shelf complex | Sign complex | Sign conflict | Space complex | Space conflict | Visual merch. complex | Visual merch. conflict |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aisle complex | 2.09 | 1.00 | – | – | – | – | – | – | – | – | – | – |
| Aisle conflict | 1.49 | 0.52 | 1.00 | – | – | – | – | – | – | – | – | – |
| Flow complex | 2.44 | 0.60 | 0.36 | 1.00 | – | – | – | – | – | – | – | – |
| Flow conflict | 2.02 | 0.44 | 0.34 | 0.65 | 1.00 | – | – | – | – | – | – | – |
| Shelf complex | 1.58 | 0.47 | 0.39 | 0.48 | 0.43 | 1.00 | – | – | – | – | – | – |
| Sign complex | 1.60 | 0.44 | 0.32 | 0.43 | 0.38 | 0.42 | 1.00 | – | – | – | – | – |
| Sign conflict | 1.72 | 0.43 | 0.36 | 0.41 | 0.49 | 0.39 | 0.50 | 1.00 | – | – | – | – |
| Space complex | 1.94 | 0.50 | 0.34 | 0.51 | 0.44 | 0.43 | 0.45 | 0.41 | 1.00 | – | – | – |
| Space conflict | 1.74 | 0.38 | 0.32 | 0.47 | 0.46 | 0.34 | 0.34 | 0.42 | 0.59 | 1.00 | – | – |
| Visual merch. complex | 2.43 | 0.51 | 0.42 | 0.51 | 0.39 | 0.44 | 0.41 | 0.38 | 0.45 | 0.40 | 1.00 | – |
| Visual merch. conflict | 2.33 | 0.43 | 0.40 | 0.48 | 0.44 | 0.43 | 0.39 | 0.44 | 0.40 | 0.41 | 0.72 | 1.00 |

**Fig. 2** MIMIC Model of Signage Confusion
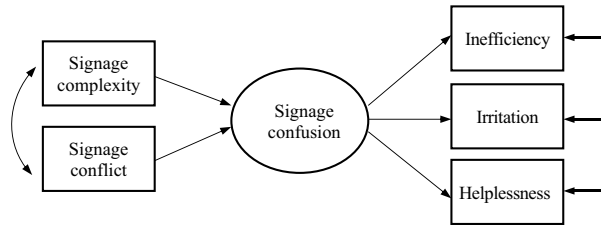


response categories, which suggests setting $x_s$ a priori (to 3, the midpoint of the scale in the present case). Metaphorically, this threshold corresponds to the range of arousal in which negative evaluations start to decrease substantially, i.e., leaving the interval of optimal arousal in the arousal-negative evaluation relationship. This pattern is reflected by the slope parameter $\left(\widehat{\alpha}_{21}^{(2)} = 0.58\right)$, which is significantly different from the slope parameter of the first regime $\left(\widehat{\alpha}_{11}^{(2)} = 0.31\right)$. Clearly, these results confirm the postulated piecewise linear relationship. Furthermore, the findings reinforce the previous postulate of not including environmental properties variety and novelty because of their limited confusion potential (as determined for the first data set, with global means not significantly different from 3; cf. Table 1).

**MIMIC Models**   Following the suggestion of Diamantopoulos and Siguaw (2006), measurement assessment of each formative construct started with a multicollinearity check. Table 3 presents the correlations between indicators and variance inflation factors (VIF). Consistently, all VIF statistics did not exceed the threshold of 3 (Hair et al. 2006). Accordingly, multicollinearity is not of concern for the present data.

The analysis proceeded with the estimation of eight MIMIC models in Lisrel version 8.51 (Jöreskog and Sörbom 1997). Composite scores of the three feelings

**Table 4** Fit Statistics of MIMIC Models (SEC Dimensions) (based on data set II)

|  | $\chi^2 / df$ | RMSEA | SRMR | GFI | NNFI | CFI | $R^2$ |
|---|---|---|---|---|---|---|---|
| *Threshold* | ≤5 | ≤0.08 | ≤0.05 | ≥0.90 | >0.90 | >0.90 | >0.02 |
| *SEC Dimensions* | | | | | | | |
| Aisles[a] | 2.09 | 0.05 | 0.02 | 1.00 | 1.00 | 1.00 | 0.31 |
| Architecture | 11.14 | 0.14 | 0.06 | 0.98 | 0.76 | 0.90 | 0.30 |
| Customer Flow[a] | 2.72 | 0.06 | 0.02 | 1.00 | 0.99 | 1.00 | 0.30 |
| Shelving and Storage[a] | 1.22 | 0.02 | 0.02 | 1.00 | 1.00 | 1.00 | 0.16 |
| Signage[a] | 4.69 | 0.08 | 0.04 | 1.00 | 0.96 | 0.98 | 0.25 |
| Space Allocation[a] | 3.24 | 0.06 | 0.03 | 1.00 | 0.98 | 0.99 | 0.26 |
| Technology | 12.99 | 0.15 | 0.06 | 1.00 | 0.84 | 0.94 | 0.26 |
| Visual Merchandising[a] | 1.64 | 0.03 | 0.01 | 1.00 | 1.00 | 1.00 | 0.32 |

*df* degrees of freedom, *RMSEA* root mean square error of approximation, *SRMR* standardized root mean residual, *GFI* goodness-of-fit index, *NNFI* non-normed fit index, *CFI* comparative fit index
[a]Store environmental confusion dimensions marked with an superscripted a are retained for the SEC index
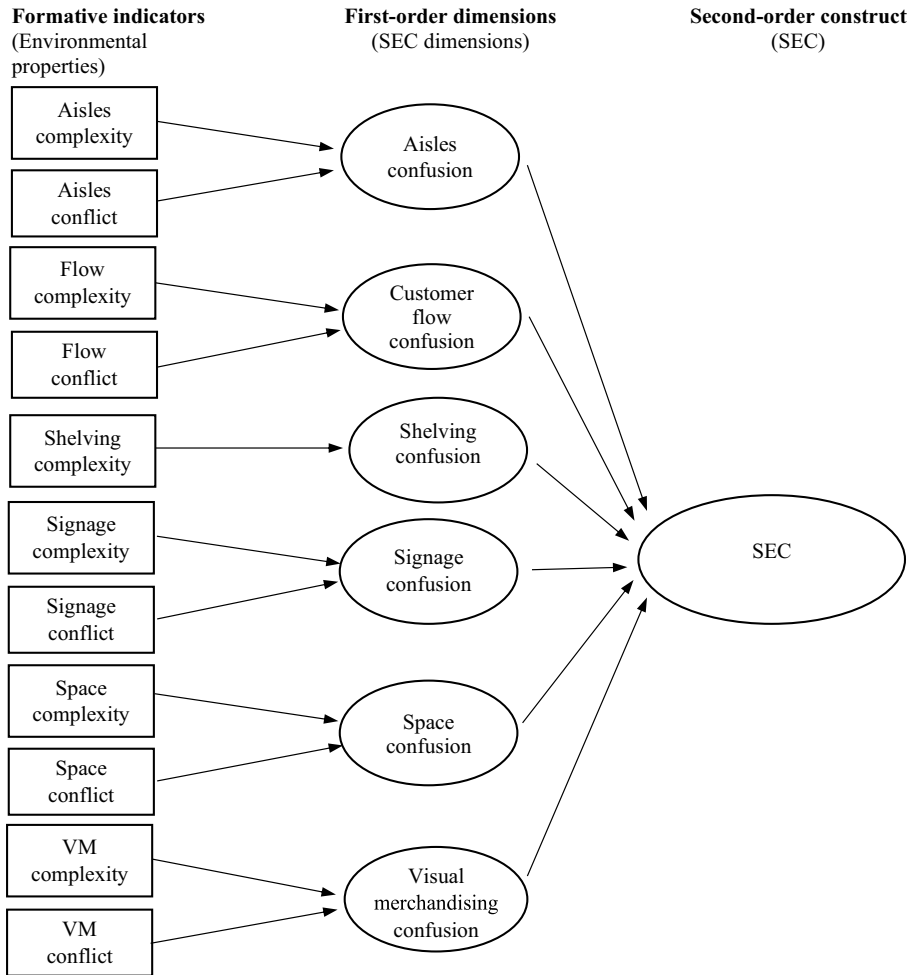
**Fig. 3**  Measurement Model of SEC

(inefficiency, irritation, and helplessness) were added as reflective indicators to each
SEC dimension for model identification (see Fig. 2 for a MIMIC model of the
environmental confusion dimension signage).

All MIMIC models show moderate or large effect sizes. With regard to the factor
loadings, the indicator conflict of the SEC dimension shelving and storage and the
indicator conflict of the SEC dimension architecture exhibit non-significant loadings.
The architecture construct also shows minor overall fit values. Unsatisfactory fit
values also apply to the SEC dimension technology (see Table 4). Therefore, these
two dimensions were not considered further in the analysis. Satisfactory fit values
of the dimension shelving and storage qualified this dimension for retention in
further analysis (see Fig. 3). Hence, following the suggestion of Diamantopoulos
and Winklhofer (2001), the non-significant indicator is omitted and only the indicator

**Table 5**   SEC Index[a]

Instruction before each SEC factor:
Please indicate which of the following properties of "*SEC dimension*" evoke confusion in a shopping situation

| SEC dimensions | Item |
| --- | --- |
| *Architecture[a]* | *Complexity (e.g., hardly manageable store; lack of orientation and clarity)* |
| | *Conflict (e.g., inappropriate architecture for type of business, such as exclusive, costly façade for a discount store)* |
| **Aisles** | **Complexity** (e.g., no overview of where to find products) |
| | **Conflict** (e.g., too narrow or wide aisles, barriers, difficult access to other aisles, bad transition to other aisles) |
| **Customer Flow** | **Complexity** (e.g., labyrinthine structure, unclear customer flow, bad orientation) |
| | **Conflict** (e.g., no customer paths to specific products, conflicting or ambiguous customer flow) |
| **Shelving and Storage** | **Complexity** (e.g., high shelves, inaccessible shelves, difficult-to-access products) |
| | *Conflict[a] (e.g., shelves as barriers, shelves as inappropriate for offered products such as exclusive wine in cheap-looking shelves)* |
| **Signage** | **Complexity** (e.g., complex signage content, lots of information on signage) |
| | **Conflict** (e.g., wrong signage, conflicting contents, multiple signage, similar signage) |
| **Space Allocation** | **Complexity** (e.g., unstructured arrangement within departments) |
| | **Conflict** (e.g., incoherent arrangement of departments, mismatched subdivisions) |
| *Technology[a]* | *Complexity (e.g., complicated use of technical devices such as scales)* |
| | *Conflict (e.g., price labels do not match check-out prices, lack of space caused by too many technical devices)* |
| **Visual Merchandising** | **Complexity** (e.g., price labels do not match products, barely understandable, unstructured visual merchandising) |
| | **Conflict** (e.g., conflicting price labels, wrong signage on shelves, wrong promotional signage) |

[a]Items printed in italics were included in the data collection for both studies but are excluded from the final index of SEC

complexity constitutes shelve confusion. The remaining indicators exhibit significant loadings. In sum, six design factors (aisles, customer flow, shelving and storage, signage, space allocation, and visual merchandising) and 11 indicators constitute the SEC index (see Table 5).

### 3.4 Stage 4: Assessment of Nomological and Predictive Validity

#### 3.4.1 Theoretical and Conceptual Considerations

The final step in the index development process assesses nomological validity, a process that specifies the relationships between the focal and related constructs and confirms the multidimensional nature of the focal construct (MacKenzie et al. 2011). In line with previous research (e.g., Garaus and Wagner 2016), SEC is thought to result in avoidance behavior (i.e., low unplanned expenditures, low store exploration, low revisit intentions, low store patronage intentions, and low spending time). Drawing on extant literature (Garaus and Wagner 2016; Garaus et al. 2015) this link between SEC and avoidance behavior is not direct, but rather indirect via the mediating state retail shopper confusion. This expectation results in including retail shopper confusion, operationalized through the three feeling states inefficiency, irritation, and helplessness, as a mediating construct in the structural equation model.

#### 3.4.2 Empirical Validation

**Nomological Validation**    Nomological validity of the SEC construct also relied on dataset II used in stage 3. Sect. 3.3.2 already detailed that step (v) of the interview prompted respondents to imagine a confusing shopping situation and asked them to answer items that sought to operationalize avoidance behavior (see Appendix).

Structural equation modeling tested the nomological validity of the SEC index. In line with prior research (cf. Landis et al. 2000), the use of composite scores for the three feelings inefficiency, irritation, and helplessness that are associated with confusion (mediating construct) and the six SEC dimensions allowed us to estimate a parsimonious model[4].

The structural model possesses an excellent fit, such that all fit statistics exceed the recommended threshold levels ($\chi^2 / df = 2.72$, $RMSEA = 0.05$, $SRMR = 0.05$, $GFI = 0.95$, $NNFI = 0.93$, $CFI = 0.95$). In addition, all exogenous constructs exhibited satisfactory power over their endogenous constructs. All $R^2$ values are between 0.16 and 0.46. Consistent with these highly satisfactory fit statistics, all but two path coefficients were statistically significant for a type I error of 0.05. Specifically, the complexity and conflict of aisles ($\gamma_1 = 0.24$, $p < 0.01$), customer flow ($\gamma_2 = 0.17$, $p < 0.01$), signage ($\gamma_4 = 0.16$, $p < 0.01$), and visual merchandising ($\gamma_6 = 0.27$, $p < 0.01$) increased negative feelings associated with confusion, and this increase in turn resulted in avoidance behavior. Shelving and storage, and space allocation did not influence the mental state retail shopper confusion. Hence, for the current sample the confusion potential of these two SEC dimensions is low. Since both SEC dimensions exhibit satisfying fit statistics on a construct level (i.e., MIMIC models, see Table 4), they are retained in the final SEC index (see Table 5).

---

[4] This model was assumed to be linear. We checked this assumption against several alternatives, i.e., a quadratic model, a semilog model, and a threshold model. None of these alternatives outperformed the linear model in terms of fit or statistical properties.

**Table 6** Structural Model Estimates of the SEC Nomological Framework (based on data set II)

| SEC dimensions | | | |
|---|---|---|---|
| *Paths to feelings associated with confusion* | *Estimate ($\gamma$)* | *p-value* | *Effect size* |
| Aisles | 0.24 | <0.01 | 0.46 |
| Customer Flow | 0.17 | <0.01 | |
| Shelving and Storage | −0.07 | 0.10 | |
| Signage | 0.16 | <0.01 | |
| Space Allocation | 0.06 | 0.13 | |
| Visual Merchandising | 0.27 | <0.01 | |
| **SEC consequences** | | | |
| *Paths to avoidance behavior* | *Estimate ($\beta$)* | – | *Effect size* |
| Unplanned expenditures | −0.40 | <0.01 | 0.16 |
| In-store search | −0.58 | <0.01 | 0.33 |
| Repeat purchase intention | −0.68 | <0.01 | 0.46 |
| Store patronage intentions | −0.53 | <0.01 | 0.28 |
| Spending time | −0.52 | <0.01 | 0.27 |

Standardized parameters are shown

The estimates with respect to the hypothesized consequences of SEC mediated through negative feelings associated with confusion were all significant and in the proposed directions. That is, SEC decreases unplanned expenditures ($\beta_1 = -0.40$, $p < 0.01$) and in-store search ($\beta_2 = -0.58$, $p < 0.01$). The negative relationship between negative feelings associated with confusion and repeat purchase intention ($\beta_3 = -0.68$, $p < 0.01$) was the strongest path in the model. In addition, negative feelings were negatively related to store patronage intentions ($\beta_4 = -0.53$, $p < 0.01$) and time spent in the store ($\beta_5 = -0.52$, $p < 0.01$) (see Table 6).

**Checking for Common Method Bias** Reliance on the same scale formats and measuring variables at the same point in time might result in systematic response behavior or artifactual covariance (Podsakoff et al. 2003). Two approaches examined whether common method bias had indeed occurred. First, Harman's one-factor test explored whether a single factor can account for all the variance in the data. The chi-square difference test between the structural equation model and a one-factor solution confirmed the superiority of the structural model ($\Delta\chi^2_{(df=28)} = 977.27$, $p < 0.01$) and to some extent eliminates common method bias concerns. Second, a partial correlation procedure included a measure of a potential source of method variance as covariate in the analysis: A marker variable controlled for common method bias (Lindell and Whitney 2001; Podsakoff et al. 2003). In particular, one item asked respondents whether they prefer plastic or paper bags when shopping. Adjustment of the zero-order correlations among the constructs by partialling out this marker variable did not find any changes in the signs or the significance levels of the factor loadings. Thus, common method bias did not appear to affect the findings. Overall, the results demonstrate a high degree of predictive and nomological validity for the SEC construct.

## 4 Discussion, Implications, Limitations and Future Research

### 4.1 Discussion

Three objectives inspired the present study. The first objective delineates a conceptualization of SEC. Second, the research sought to develop a measurement instrument for assessing SEC. The third objective provides an assessment of the nomological and predictive validity of the construct SEC. Introducing this construct follows the call to explore the confusion potential of store environments. In particular, this study expanded Garaus et al.'s (2015) research by verifying the confusion potential of six store design elements (aisles, customer flow, shelving and storage, signage, space allocation, and visual merchandising) through environmental properties complexity and conflict. In doing so, this study seeks to move the literature beyond a one-facet classification of the store environment, to explain why and how shoppers perceive store environmental stimuli as confusing.

Following established index development guidelines, two large-scale quota-sample based data collections allowed us to construct this parsimonious index. The results deliver strong empirical support for the confusion potential of store environments. Furthermore, the results of the assessment of the nomological validity of the index provide insights into the relationships between SEC, feelings associated with confusion, and behavioral responses. In particular, data confirm that SEC evokes feelings of inefficiency, irritation, and helplessness that mediate the relationship between SEC and avoidance behavior. Therefore, the results confirm extant research that points to the negative consequences of confusing store environments (Garaus et al. 2015; Garaus and Wagner 2016). However, in contrast to these extant studies, we demonstrate the multidimensional nature of the SEC construct and relate store-induced confusion to negative feelings and avoidance behavior.

The present research is the first to explore the varying confusion potential of different design factors. The six design factors aisles, customer flow, shelving and storage, signage, space allocation and visual merchandising (see Table 5 and Fig. 3) bear a high confusion potential. Although all these six SEC dimensions exhibit satisfactory fit statistics when assessing their construct reliability, only four (aisles, customer flow, signage and visual merchandising) increase the negative feelings that constitute confusion in the nomological network in this particular sample. Aisle and visual merchandising bear a higher confusion potential as compared to signage and customer flow. Accordingly, a lack of overview where to find products (complexity) and too narrow or wide aisles, barriers, difficulties to access other aisles and a bad transition to other aisles (conflict) evoke negative feelings of confusion. Likewise, price labels that do not match products, and a barely understandable unstructured visual merchandising strategy (complexity) as well as conflicting price labels, wrong signage on shelves and wrong promotional signage (conflict) represent the environmental properties evoking confusion for the visual merchandising dimension. A complex signage content or lots of information on signage reflect the environmental property complexity for the signage dimension. The environmental property conflict is manifested in wrong signage, conflicting contents, multiple signage and similar signage. Finally, referring to the dimension customer flow, a labyrinthine

structure, an unclear customer flow and a bad orientation are examples of the environmental property complexity. Examples for the environmental property conflict include the absence of customer paths to specific products and a conflicting or ambiguous customer flow.

The properties conflict and complexity are suitable to represent the SEC dimensions. Conflict (i.e., mismatch) leads to enhanced cognitive processing (Mattila and Wirtz 2001) and confusion (Beverland et al. 2006). In addition, complex design factors require greater cognitive processing than simple ones (Herrmann et al. 2013), inducing SEC. These findings further emphasize the need to consider SEC as a unique construct that differs considerably from product-induced confusion (in contrast, the properties variety and novelty are of particular relevance for products).

Finally, this article contributes to the extant literature by applying a new approach for testing for response bias likely caused by respondents' fatigue. Study 1 analyzed response patterns using Mood's (1940) asymptotic multinomial runs test. This test allows for assessment of whether response runs occur at random or follow regular patterns. Even though most studies endeavor to use parsimonious scales to avoid response fatigue, the development of an initial item pool in scale development procedures often requires comprehensive questionnaires. Research emphasizes the risk of response bias when using extensive measurement instruments (Weijters and Baumgartner 2012), especially because capturing the entire construct in the early stage of scale development procedures necessitates large initial item pools (Netemeyer et al. 2003). The present study offers an approach for testing for response fatigue, thereby creating the opportunity to eliminate participants who did not diligently respond. The elimination of biased responses further guarantees the development of a reliable and valid measurement instrument.

### 4.2 Implications

The findings of this research possess several practical implications. So far, the only opportunity to reduce confusion during shopping situations falls into the realm of manufacturing (e.g., by using less similar packaging). This research is the first that offers retailers the opportunity to reduce the confusion potential of store environments by employing a measurement instrument. Two large-scale studies identified that complex and conflicting aisles, customer flows, signage, and visual merchandising bear the highest confusion potential. Managers might apply this information to revise their aisles and customer flow plans to reduce confusion in their store. For example, complexity in aisles implies a lack of organization, and organization in turn can help customers find products or identify blind alleys. Conflict-ridden aisles might be too narrow or too wide, difficult to access, or may not be aisles at all (e.g., star layout). Complexity conveyed by customer flow likely stems from a labyrinthine structure, unclear or poor traffic patterns, or a confusing shop-in-shop design. An ambiguous, unclear, or unidentifiable flow pattern suggests the conflict element. Hence, retailers are encouraged to pay particular attention to these design factors. Moreover, store designers should consider information about the confusing environmental properties complexity and conflict.

However, the confusion potential likely differs among stores. On an individual store level, the SEC index developed here offers retailers the opportunity to identify confusion sources within a particular store. Assessing SEC proposes design factors that managers can easily manipulate to reduce the confusion potential of the entire store environment and create clear and non-confusing store environments. These insights are of high relevance for retailers.

### 4.3 Limitations

Despite the important implications of this study, the findings must be interpreted with several limitations in mind. There remains an ongoing discussion on the validity of MIMIC models (Diamantopoulos, 2013; Diamantopoulos and Temme 2013; Lee et al. 2013; Lee and Cadogan 2013; Cadogan et al. 2013). Many theoretical considerations guided our decision to rely on MIMIC models in this index development process. However, we agree that the applicability of MIMIC models in formative index construction offers room for discussion. Even though we feel that we overcame one major criticism of MIMIC models (namely variance explained by different sets of reflective indicators), we acknowledge that MIMIC models are associated with other problems and that further procedures for testing formative measurement models would enrich the extant literature.

Although much effort was placed into eliminating potential common method bias, both statistical approaches used in the present study cannot completely exclude the possibility of any common method variance. Additional data collection is required to offer further evidence on the robustness of the results of the nomological framework of SEC.

### 4.4 Further Research

These results cannot be generalized to different cultures without further research. Asian customers might perceive narrow aisles as less confusing because, for example, they have different acceptance levels of spatial distance. Hence, in Asia aisle confusion might not be as important as in European countries. Testing the whole SEC framework in various cultural contexts would represent a valuable avenue for future research.

Exploring different retail sectors concerning their inherent confusion potential might be worthwhile. The developed SEC index can be employed to identify industries that exhibit more pronounced confusion potential. Additionally, confusion causes might include the redesign of a whole store or the employment of new technologies. Extending these considerations, analysis of interactions between drivers of confusion offers further research options. In this context, it could be worth comparing SEC sources among different industries (e.g., grocery stores vs. clothing stores).

It is also reasonable to assume that SEC is not only of high relevance to retail industries but also to service settings. Hence, future studies might expand the SEC construct to a service context.

Another fruitful area of future research is the interplay between product-induced and store-induced confusion. Since retailers cannot reduce the confusion potential of products (i.e., similarity of products), it would be particularly interesting to investigate whether a pleasant and clear store design can reduce the overall confusion level during a shopping experience.

# Appendix

**Table 7** Measurement Scales of Feelings constituting Confusion and Avoidance Behavior

| **Feelings constituting confusion** (Garaus and Wagner 2016) (based on data set II) |
| --- |
| *Inefficiency (Cronbach's α = 0.80, AVE = 0.51, CR = 0.81)* |
| Efficient[a] |
| Careful[a] |
| Productive[a] |
| High performing[a] |
| *Irritation (Cronbach's α = 0.84, AVE = 0.64, CR = 0.84)* |
| Annoyed |
| Irritated |
| Unnerved |
| *Helplessness (Cronbach's α = 0.86, AVE = 0.51, CR = 0.86)* |
| Helpless |
| Lost |
| Awkward |
| Baffled |
| Weak |
| Overstrained |
| **Avoidance Behavior** (Donovan et al. 1994; Wakefield and Baker 1998) |
| *Unplanned expenditures* |
| This is a kind of store where I would spend more money than expected |
| *Store exploration* |
| I would explore this store more thoroughly |
| I would avoid exploring the store in more detail.[a] |
| *Revisit intention (Cronbach's α = 0.75)* |
| The likelihood that I would shop in this store in future is high |
| I would avoid returning to this store.[a] |
| I intend to visit this store again |
| *Store patronage intentions* |
| I would enjoy shopping in this store |
| *Spending time* |
| I would like to stay in this store as long as possible |
| I would spend more time in this store than intended |

*AVE* average variance extracted, *CR* construct validity

[a]Reversed items. Respondents were asked to indicate on a five point Likert-scale the extent they agreed to experience the particular feeling during a confusing shopping experience/to react in a confusing shopping situation.

# References

Adams, Peter, Michael D. Hurd, Daniel McFadden, Angela Merrill, and Tiago Ribeiro. 2003. Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics* 112(1):3–56.

Bagozzi, Richard P., and Youjae Yi. 2012. Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science* 40(1):8–34.

Baker, Julie. 1987. The role of the environment in marketing services: The consumer perspective. In *The services challenge: Integrating for competitive advantage*, ed. John A. Czepiel, 79–84. Chicago: American Marketing Association.

Baker, Julie, Ananthanarayanan Parasuraman, Dhruv Grewal, and Glenn B. Voss. 2002. The influence of multiple store environment cues on perceived merchandise value and patronage intentions. *Journal of Marketing* 66(2):120–141.

Berlyne, Daniel E. 1960. *Conflict, arousal, and curiosity*. New York: McGraw-Hill.

Beverland, Michael, Elison A.C. Lim, Michael Morrison, and Milé Terziovski. 2006. In-store music and consumer–brand relationships: Relational transformation following experiences of (mis)fit. *Journal of Business Research* 59(9):982–989.

Bitner, Mary J. 1992. Servicescapes: The impact of physical surroundings on customers and employees. *Journal of Marketing* 56(2):57–71.

Bollen, Kenneth A. 1989. A new incremental fit index for general structural equation models. *Sociological Methods & Research* 17(3):303–316.

Bollen, Kenneth A. 2011. Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly* 35(2):359–372.

Bollen, Kenneth, and Richard Lennox. 1991. Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin* 110(2):305–314.

Bollen, Kenneth A., and Adamantios Diamantopoulos. 2017. In defense of causal-formative indicators: A minority report. *Psychological Methods* 22(3):581–596.

Brüggen, Elisabeth C., Bram Foubert, and Dwayne D. Gremler. 2011. Extreme makeover: Short- and long-term effects of a remodeled servicescape. *Journal of Marketing* 75(5):71–87.

Cadogan, John W., Nick Lee, and Laura Chamberlain. 2013. Formative variables are unreal variables: Why the formative MIMIC model is invalid. *AMS Review* 3(1):38–49.

Cupchik, Gerald C., and Daniel E. Berlyne. 1979. The perception of collative properties in visual stimuli. *Scandinavian Journal of Psychology* 20(1):93–104.

Di Muro, Fabrizio, and Kyle B. Murray. 2012. An arousal regulation explanation of mood effects on consumer choice. *Journal of Consumer Research* 39(3):574–584.

Deng, Liqiong, and Marshall Scott Poole. 2010. Affect in web interfaces: A study of the impacts of web page visual complexity and order. *MIS Quarterly* 34(4):711–730.

Diamantopoulos, Adamantios. 2013. MIMIC models and formative measurement: Some thoughts on Lee, Cadogan & Chamberlain. *AMS Review* 3(1):30–37.

Diamantopoulos, Adamantios, and Petra Riefler. 2008. Formative Indikatoren: Einige Anmerkungen zu ihrer Art, Validität und Multikollinearität. *Zeitschrift für Betriebswirtschaft* 78(11):1183–1196.

Diamantopoulos, Adamantios, and Judy A. Siguaw. 2006. Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management* 17(4):263–282.

Diamantopoulos, Adamantios, and Dirk Temme. 2013. MIMIC models, formative indicators and the joys of research. *AMS Review* 3(3):160–170.

Diamantopoulos, Adamantios, and Heidi M. Winklhofer. 2001. Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research* 38(2):269–277.

Diamantopoulos, Adamantios, Petra Riefler, and Katharina P. Roth. 2008. Advancing formative measurement models. *Journal of Business Research* 61(12):1203–1218.

Donovan, Robert J., and John R. Rossiter. 1982. Store atmosphere: An environmental psychology approach. *Journal of Retailing* 58(1):34–57.

Donovan, Robert J., John R. Rossiter, Gilian Marcoolyn, and Andrew Nesdale. 1994. Store atmosphere and purchasing behavior. *Journal of Retailing* 70(3):283–294.

Drummond, Graeme. 2004. Consumer confusion: Reduction strategies in higher education. *International Journal of Educational Management* 18(4):317–323.

Dubé, Laurette, Jean-Charles Chebat, and Sylvie Morin. 1995. The effects of background music on consumers' desire to affiliate in buyer-seller interactions. *Psychology and Marketing* 12(4):305–319.

Friedman, Monroe P. 1966. Consumer confusion in the selection of supermarket products. *Journal of Applied Psychology* 50(6):529–534.

Garaus, Marion, and Udo Wagner. 2016. Retail shopper confusion: Conceptualization, scale development, and consequences. *Journal of Business Research* 69(9):3459–3467.

Garaus, Marion, Udo Wagner, and Claudia Kummer. 2015. Cognitive fit, retail shopper confusion, and shopping value: Empirical investigation. *Journal of Business Research* 68(5):1003–1011.

Hair, Joseph F., William C. Black, Barry J. Babin, Rolph E. Anderson, and Ronald L. Tatham. 2006. *Multivariate data analysis*, 6th edn., Upper Saddle River: Pearson Prentice Hall.

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.

Holland, Rob W., Merel Hendriks, and Henk Aarts. 2005. Smells like clean spirit. Nonconscious effects of scent on cognition and behavior. *Psychological Science* 16(9):689–693.

Howell, Roy D., Einar Breivik, and James B. Wilcox. 2007. Reconsidering formative measurement. *Psychological Methods* 12(2):205–218.

Jarvis, Cheryl B., Scott B. MacKenzie, and Philip M. Podsakoff. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research* 30(2):199–218.

Jöreskog, Karl G., and Sörbom Dag. 1997. *User's reference guide*, 2nd edn., Chicago: Scientific Software International.

Kasper, Hans, Josée Bloemer, and Paul H. Driessen. 2010. Coping with confusion: The case of the Dutch mobile phone market. *Managing Service Quality* 20(2):140–160.

Ketron, Seth. 2018. Perceived product sizes in visually complex environments. *Journal of Retailing* 94(2):154–166.

Landis, Ronald S., Daniel J. Beal, and Paul E. Tesluk. 2000. A comparison of approaches to forming composite measures in structural equation models. *Organizational Research Methods* 3(2):186–207.

Lee, Nick, and John W. Cadogan. 2013. Problems with formative and higher-order reflective variables. *Journal of Business Research* 66(2):242–247.

Lee, Nick, John W. Cadogan, and Laura Chamberlain. 2013. The MIMIC model and formative variables: Problems and solutions. *AMS Review* 3(1):3–17.

Leek, Sheena, and Suchart Chansawatkit. 2006. Consumer confusion in the Thai mobile phone market. *Journal of Consumer Behaviour* 5(6):518–532.

Leek, Sheena, and Dai Kun. 2006. Consumer confusion in the Chinese personal computer market. *Journal of Product & Brand Management* 15(3):184–193.

Leek, Sheena, Isabelle Szmigin, and Emily Baker. 2015. Consumer confusion and front of pack (FoP) nutritional labels. *Journal of Customer Behaviour* 14(1):49–61.

Lindell, Michael K., and David J. Whitney. 2001. Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology* 86(1):114–121.

MacKenzie, Scott B., Philip M. Podsakoff, and Cheryl B. Jarvis. 2005. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology* 90(4):710–729.

MacKenzie, Scott B., Philip M. Podsakoff, and Nathan P. Podsakoff. 2011. Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly* 35(2):293–334.

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):273–289.

Mehrabian, Albert, and James A. Russell. 1974. *An approach to environmental psychology*. Cambridge: M.I.T. Press.

Mehta, Ritu. 2013. Understanding perceived retail crowding: A critical review and research agenda. *Journal of Retailing and Consumer Services* 20(6):642–649.

Mitchell, Vincent-Wayne, and Vassilios Papavassiliou. 1997. Exploring consumer confusion in the watch market. *Marketing Intelligence & Planning* 15(4):164–172.

Mitchell, Vincent-Wayne, and Vassilios Papavassiliou. 1999. Marketing causes and implications of consumer confusion. *Journal of Product & Brand Management* 8(4):319–339.

Mitchell, Vincent-Wayne, Gianfranco Walsh, and Yo Yamin. 2005. Towards a conceptual model of consumer confusion. *Advances in Consumer Research* 32:143–150.

Mood, Alexander M. 1940. The distribution theory of runs. *Annals of Mathematical Statistics* 11(4):367–392.

Nasar, Jack. 1987. The effect of sign complexity and coherence on the perceived quality of retail scenes. *Journal of the American Planning Association* 53(4):499–509.

Netemeyer, Richard G., William O. Bearden, and Subhash Sharma. 2003. *Scaling procedures: Issues and applications*. Thousand Oaks: SAGE.

North, Adrian C., David J. Hargreaves, and Jennifer McKendrick. 1997. In-store music affects product choice. *Nature* 390(6656):132.

Orth, Ulrich R., Jochen Wirtz, and Amelia McKinney. 2016. Shopping experiences in visually complex environments: A self-regulation account. *Journal of Service Management* 27(2):194–217.

Parks, Malcolm R., and Kory Floyd. 1996. Making Friends in Cyberspace. *Journal of Computer-Mediated Communication* 1(4):80–97.

Pine, Joseph B., and James H. Gilmore. 1998. Welcome to the experience economy. *Harvard Business Review* 76:97–105.

Podsakoff, Philip M., Scott B. MacKenzie, Lee Jeong-Yeon, and Nathan P. Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88(5):879–903.

Schweizer, Markus. 2004. *Consumer Confusion im Handel – Ein umweltpsychologisches Erklärungsmodell*. Wiesbaden: Deutscher Universitäts-Verlag.

Schweizer, Markus, Alexander J. Kotouc, and Tillmann Wagner. 2006. Scale development for consumer confusion. *Advances in Consumer Research* 33:184–190.

Treiblmaier, Horst, Peter M. Bentler, and Patrick Mair. 2011. Formative constructs implemented via common factors. *Structural Equation Modeling* 18(1):1–17.

Varon, Lily. 2016. Retailers are prioritizing the wrong digital technologies in brick and mortar stores. https://www.forbes.com/sites/forrester/2016/09/08/retailers-are-prioritizing-the-wrong-digital-technologies-in-brick-and-mortar-stores/#d78322b75bd. Accessed 6 June 2019.

Wakefield, Kirk L., and Julie Baker. 1998. Excitement at the mall: Determinants and effects on shopping response. *Journal of Retailing* 74(4):515–539.

Walsh, Gianfranco, and Vincent-Wayne Mitchell. 2010. The effect of consumer confusion proneness on word of mouth, trust, and customer satisfaction. *European Journal of Marketing* 44(6):838–859.

Walsh, Gianfranco, Thorsten Hennig-Thurau, and Vincent-Wayne Mitchell. 2007. Consumer confusion proneness: Scale development, validation and application. *Journal of Marketing Management* 23(7/8):697–721.

Weijters, Bert, and Hans Baumgartner. 2012. Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research* 49(5):737–747.