Consumer susceptibility to front-of-package (FOP) food labeling: Scale development and validation

Michał Folwarczny [a,*], Valdimar Sigurdsson [b], R.G. Vishnu Menon [b,c], Tobias Otterbring [d]

[a] Discipline of Marketing, J.E. Cairnes School of Business & Economics, University of Galway, Ireland
[b] Department of Business Administration, Reykjavik University, Iceland
[c] School of Communication, Journalism and Marketing, Massey University, New Zealand
[d] Department of Management, University of Agder, Norway

A B S T R A C T

Current research in food science has explored the influence of front-of-package (FOP) labeling systems on consumer decision-making, yielding mixed results. We suggest that these inconsistent findings regarding FOP labeling effectiveness stem from a failure to consider a pivotal individual-level variable: consumer susceptibility to FOP labeling (CSFL). In the present research, we define this focal construct and develop and psychometrically validate a seven-item instrument that captures the construct across six studies ($N = 1134$). The current research may assist in segmenting consumers based on their susceptibility to FOP labeling, thereby facilitating the creation of targeted interventions tailored to this individual difference. Notably, the CSFL scale is positively correlated with consumers' willingness to purchase food items with genuine, third-party FOP labels, but not products lacking labels or products with fictitious FOP labels. This supports the predictive validity of the scale in determining important consumption-related outcomes.

1. Introduction

Although retailers change their assortments regularly, the average grocery store carries more than 30,000 items, making it difficult for consumers to make informed decisions about which products to buy (Baum, 2019). Policymakers, food manufacturers, and other entities in the food industry have introduced front-of-package (FOP) labels to help shoppers navigate this dense jungle of product alternatives and simplify their decision-making process (Kühne, Reijnen, Granja, & Hansen, 2022). For example, food manufacturers sometimes use traffic light labels to inform consumers about the nutritional value of their products (i.e., green indicates high nutritional value and red indicates low nutritional value; Golan, Kuchler, Mitchell, Greene, & Jessup, 2001).

Despite efforts to promote certain product choices, FOP food labeling systems do not always work as intended. For instance, some studies demonstrate that introducing traffic light labeling on seafood neither increases sales of green-labeled products nor decreases sales of red-labeled alternatives (Hallstein & Villas-Boas, 2013). Similarly, adding FOP labels indicating healthier choices leads to a mere 1.5–2% increase in the frequency of purchasing these products compared to their less healthy alternatives (Temple, 2020). On the other hand, research shows

that nutritional FOP warnings are effective in helping consumers perceive products with unfavorable nutritional profiles as unhealthy (Cabrera et al., 2017). Numerous studies support the effectiveness of nutritional FOP warnings in assisting consumers in making healthier food choices (Ares et al., 2018, 2021). Interestingly, fictitious seafood sustainability labels sometimes increase consumers' willingness to pay for labeled products more than genuine third-party certificates (Sigurdsson et al., 2022). In light of these findings, contradictory at times, what determines whether consumers' preferences and decisions sway in response to the introduction of FOP labeling systems?

Previous research has suggested that certain consumers may be more likely than others to rely on FOP labeling systems when purchasing food (Sigurdsson et al., 2020). Fewer than 10% of consumers say they always or often buy seafood with sustainability labels, and 75% say they never buy such products (Clonan, Holdsworth, Swift, Leibovici, & Wilson, 2012), supporting the notion that there are different segments of consumers in terms of their propensity to use FOP labeling systems when shopping. Yet, it remains unknown why some consumers actively adjust their food preferences and decisions when confronted with FOP labeling, while others appear unaffected by such cues. In other words, the existing literature largely neglects the role of individual differences in consumer

susceptibility to FOP labeling systems.

The current paper aims to make three major contributions to address this gap.

First, we define the construct of interest: consumer susceptibility to FOP labeling (CSFL). Second, we develop a scale to capture this focal construct and validate this tool psychometrically. Third, we discuss how the scale can be used in research on the effectiveness of FOP labeling systems, with implications for retail practices among actors working in the food industry.

1.1. Front-of-package (FOP) labels

FOP labeling systems vary in complexity, scope, the information conveyed, and other characteristics. FOP labels are developed by entities such as food manufacturers, retailers, governmental agencies, nonprofit organizations, as well as experts outside the industry, and convey information about a product's nutritional profile (e.g., sugar content), summary information (e.g., American Heart Association's "Heart Check" symbol), and information about food groups (e.g., whole grain product; Wartella et al., 2010). FOP labeling systems use a variety of means to provide information to consumers, including numeric data (e. g., percentage of daily calorie needs met by one serving), color-coded and textual warnings (e.g., high sodium content), and simple logos that summarize product attributes (e.g., the "keyhole" label used in Nordic countries, which is given to relatively healthy products within their categories; Muzzioli, Penzavecchia, Donini, & Pinto, 2022).

Notably, as highlighted by Petersen, Hartmann, and Hirsch (2021), FOP labels occasionally misinterpret the actual nutritional properties of products. Additionally, some FOP labels are fabricated specifically for research studies of product marketing campaigns (Ares et al., 2020; Celhay & Remaud, 2018). For example, researchers or retailers may label products with a greenish tag with the word "environmentally sustainable" despite these products being derived from unknown sources (Sigurdsson et al., 2022). In summary, FOP labeling systems come in various forms and are designed "to provide consumers with accessible and easy-to-understand nutrition information to assist them make more informed food choice decisions and encourage food manufacturers to improve the nutritional quality of their products" (Pettigrew, Jongenelis, Hercberg, & Julia, 2022, p. 136). In this paper, however, we do not focus only on health-related FOP labeling, but examine FOP labeling systems across domains, including those directly unrelated to health. This is important as FOP labels also convey qualities such as a product's environmental impact, including organic status or low greenhouse gas emissions, as well as its origin, whether local or imported (Duckworth et al., 2022). Therefore, this research emphasizes the broad conceptualization of FOP labeling systems across domains including, among others, product origin, sustainability, health impact, and the ethics of sourcing methods.

1.2. FOP labels and consumers' decision-making

The literature to date has identified numerous external factors associated with consumer preferences and decisions, such as food healthiness, convenience, and price (Prescott, Young, O'neill, Yau, & Stevens, 2002). Yet, internal factors (e.g., food neophobia and disgust), which constitute the focus of the current research also play an influential role in shaping consumer preferences (Siegrist, Hartmann, & Keller, 2013). In food-related contexts, the literature differentiates between animal reminder disgust and core disgust (Hartmann & Siegrist, 2018). Reminder disgust refers to reactions to stimuli like animal corpses or bodily damage, exemplified by statements such as, "It would bother me to be in a science class and see a human hand preserved in a jar." Core disgust, on the contrary, pertains to responses to the oral incorporation of offensive items such as monkey meat (Olatunji, Haidt, McKay, & David, 2008). For example, consumers who have higher (vs. lower) scores on the animal reminder disgust and core disgust are less (vs.

more) inclined to enter a lottery to win a ticket to an insectarium with a buffet of insect dishes (Hamerman, 2016).

Although the literature has examined the influence of FOP labeling systems on consumer outcomes such as purchase intentions or perceptions of food healthiness (Ikonen, Sotgiu, Aydinli, & Verlegh, 2020), little scholarly interest has been directed toward individual-level predictors of the propensity to use FOP labeling in food-related decision-making contexts. Notwithstanding this general lack of research, the sparse literature on this topic has established, for example, the moderating role of consumers' dominant information processing system on the use of FOP labels in food-related judgments and decision-making, consistent with the dual process theory of mental processing that distinguishes between fast and automatic versus slow and deliberate information processing (Sanjari, Jahn, & Boztug, 2017). However, this account cannot be solely taken into consideration when explaining individual differences in the extent to which consumers rely on FOP labeling systems in their food-related decision-making because when consumers face difficulties in processing information, they tend to switch from automatic to deliberate processing (Alter, Oppenheimer, Epley, & Eyre, 2007). This, in turn, implies that fast versus slow information processing depends on situational factors rather than on predispositions alone. Thus, the dual process model of mental operations does not explain why some consumers rely more on FOP labels for their food choices than others, as recent studies suggest (cf. Sigurdsson et al., 2020).

1.3. Consumer susceptibility to FOP labeling

Building on previous research indicating that consumers vary in their susceptibility to external factors—such as the opinions of others—in making decisions (Bearden, Netemeyer, & Teel, 1989), we propose that shoppers also vary in their susceptibility to FOP labeling systems. We formally define consumer susceptibility to FOP labeling as "the extent to which consumers rely on FOP labeling in their food choices and information search activities to meet their existing preference patterns." Because susceptibility to FOP labeling is a trait variable, its values should be relatively stable over time, with only minor shifts due to external factors such as the complexity of FOP labeling (cf. Alter et al., 2007). We expect that susceptibility to FOP labeling systems is related to other trait variables that moderate reliance on heuristics rather than deliberate thinking. Hence, individual differences in thinking styles (analytic vs. holistic thinking; Nisbett, Peng, Choi, & Norenzayan, 2001) may be an example of variables related to our focal construct.

In what follows, we develop a scale to capture our trait-level variable of consumer susceptibility to FOP labeling and examine the psychometric properties of this newly developed instrument. In Study 1a, we develop an initial pool of items, which is then tested for face validity in Study 1 b, leading to the refinement and reformulation of certain items. In Study 2, we truncate the scale and evaluate its psychometric properties. In Studies 3a–b, we test the convergent and divergent validity of the scale and conduct a test-retest reliability analysis. Finally, in Study 4, we examine the criterion validity of the consumer susceptibility to FOP labeling (CSFL) scale to ensure that it predicts willingness to buy products as a function of FOP labeling systems used when presenting food products.

2. General method

In analyzing the psychometric properties of the scale, we chose to use nonparametric item response theory (NIRT) instead of the traditional factor analysis approach, followed by the specification of Cronbach's alpha. There were several reasons for this analytic decision. Item response theory (IRT), either parametric or nonparametric, offers distinct advantages over factor analysis. Unlike factor analysis, which assumes linear relationships and focuses on correlations between items, IRT models the performance of individual items based on their

underlying latent properties (Embretson & Reise, 2013; Sijtsma & Molenaar, 2002). As such, IRT provides more detailed item-level diagnostics, allowing for greater understanding and refinement of individual items, which is not possible with factor analysis (Bortolotti, Tezza, de Andrade, Bornia, & de Sousa Júnior, 2013; Dima, 2018). In addition, IRT is more robust when handling missing data (Hambleton & Swaminathan, 2013).

NIRT also offers several advantages over parametric item response theory (PIRT). NIRT is more robust to violations of parametric assumptions, whereas PIRT models assume that the latent trait is normally distributed and that item response functions (IRFs) follow a parametric form (Sijtsma & Molenaar, 2002). This inherent flexibility makes NIRT particularly potent to the complexity of real-world data and makes it a preferred choice for datasets that cannot adhere to strict assumptions. NIRT models thus make fewer assumptions about the distribution of the latent feature and IRFs, making them more robust to violations of parametric assumptions (Dima, 2018; Embretson & Reise, 2013). Further, NIRT can be used with a broader range of item types, including polytomous items (items with more than two response options), graded response items (items with ordered response options), and rating scale items (De Ayala, 2013). On top of these advantages, NIRT models are more flexible in modeling item response functions than PIRT models, which typically assume that IRFs follow a particular parametric form. For that reason, NIRT models allow for greater flexibility in representing the relationship between the latent trait and the item response (Gibbons & Chakraborti, 2014). Nonparametric models also require smaller sample sizes than their parametric equivalents, making the process of scale development more efficient (Chen et al., 2014).

Perhaps due to these numerous advantages, nonparametric item response theory, including the Mokken scale analysis (Mokken, 1971), has been extensively used in scale development and validation in food-related research in diverse areas such as perception of food additives, food security, feeding difficulty, and overeating (e.g., Bagnasco et al., 2015; Bearth, Cousin, & Siegrist, 2014; Faye, Baschieri, Falkingham, & Muindi, 2011; Folwarczny, Li, Sigurdsson, Tan, & Otterbring, 2021; Pasquale et al., 2023; Watson, 1996).

We began assessing the psychometric properties of the scale by applying a Mokken scale analysis (MSA) to examine its unidimensionality, monotonicity, local independence, and invariant item ordering (Van der Ark, 2007, 2012). Scale unidimensionality means that a set of items can be located on a continuum of a latent variable. The monotonicity criterion means that the probability of affirming a particular response does not decrease as the values of the latent variable increase. Local independence conveys the notion that the items are related only by the latent variable they are intended to measure. Finally, invariant item ordering is characterized by the items or responses being arranged similarly for different levels of the latent variable (Dima, 2018; Sijtsma & Hemker, 1998; Sijtsma & Junker, 1996). Whereas a set of items can be considered a scale if the first three criteria are met, which are essential for a monotone homogeneity model, an invariant item ordering is necessary to meet the criteria for the double monotonicity model (Mokken, 1971; Van der Ark, 2007, 2012).

Unidimensionality was formally assessed by examining the homogeneity coefficients H, ranging from 0 to 1, with 0 indicating no association between items and 1 indicating a perfect association between them, i.e., a strongly unidimensional scale (Van der Ark, 2007). To further investigate whether the scale was unidimensional, we applied the automated item selection procedure (AISP), which positions items in scales with increasing degrees of homogeneity (Hemker, Sijtsma, & Molenaar, 1995). The remaining scale properties, i.e., monotonicity, local independence and invariant item ordering were evaluated using tools from the *mokken* package for R (Van der Ark, 2007).

All analyzes were performed in R version 4.2.2 (R Core Team, 2022). Studies were coded in the free-to-use and publicly available PsyToolkit (Stoet, 2010, 2017). This project has been peer-reviewed and deemed low risk (Massey University Human Ethics Notification number: 4,000,

023,879). All participants accepted informed consent forms before participating in the studies described below. We used the same exclusion criteria across Studies 2–4.

3. Item generation (Study 1 a)

The goal of Study 1 was to develop an initial pool of items for further evaluation. Consistent with recommendations (e.g., Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018), we developed several times more items than the number of items in the desired final scale and asked laypeople from the target population to suggest additional items.

3.1. Participants and procedure

Participants in Study 1a were the paper's four authors, who have extensive experience in research on FOP labeling systems. We additionally recruited 10 US participants from Prolific Academic who declared no dietary restrictions (Mean age = 29.5, SD = 6.9 years, four females, five males, one participant did not identify with either of the two genders, mean annual income = 44,700 USD), with this sample combination (14 participants in total) following best practices in scale development research, considering that it usually results in broadening the coverage of a target construct (Boateng et al., 2018). In estimating the desired sample size, we aimed to double the number of items developed by the researchers. Therefore, 8–10 additional participants were deemed suitable for this task, as this number exceeds common sample size conventions at the initial stage of the scale development process in food-related contexts (e.g., Hemmer, Hitchcock, Lim, Kovacic, & Lee, 2021).

The researchers initially generated 25 items. After becoming familiar with the definition of our focal construct—as described in the introduction—the 10 additional Prolific participants were then asked to "write three short and positively worded sentences below that measure the importance of food labeling/certification to consumers when purchasing food," resulting in an additional 30 items. Subsequently, this combined pool of 55 items was independently evaluated by four raters (i.e., the researchers involved in this project) on a four-point scale for relevance to the construct of interest (1 = *Not at all relevant*; 4 = *Very relevant*).

3.2. Results and discussion

To assess the degree of agreement between raters, we calculated an intraclass correlation coefficient (ICC; Shrout & Fleiss, 1979). Results from a mean-rating, consistency, two-way mixed-effects model indicated good reliability of ratings between the four raters, ICC = 0.86, 95% CI [0.78, 0.91] (Koo & Li, 2016). The raters estimated the mean relevance of these 55 items to the focal construct as slightly below the scale midpoint (M = 2.42). At this stage, items whose relevance to the focal construct was evaluated as median and below were removed from the item pool, leaving 21 items for further evaluation. Subsequently, we combined five items that referred to specific product categories (e.g., "When buying groceries, I seek labeled/certified seafood") into one more generic item that referred to food more broadly defined, thereby compressing these five items into one generic item, resulting in 17 items used for the follow-up evaluation in Study 1b.

4. Scale construction and face validity (Study 1 b)

Study 1 b was conducted to further evaluate the 17 items developed in Study 1a and select the most relevant items to the construct of interest. Participants from the scale's target population also rated the comprehensibility and grammatical correctness of the items, providing input for possible item reformulation and refinement.

4.1. Participants and procedure

In Study 1 b, we recruited 36 US participants from Prolific Academic who declared no dietary restrictions (Mean age = 38.2, *SD* = 14.4, 47.2% females, 52.8% males, mean annual income = 46,600 USD). Similar sample sizes have been used in other face validity studies recruiting non-experts online (e.g., Umanath & Coane, 2020).

Participants were first informed of the purpose of the study and presented with an example of an FOP food label along with our formal definition of consumer susceptibility to FOP labeling. They then rated how relevant the 17 items were to the construct of interest on a four-point scale ranging from 0 (*Not at all relevant*) to 3 (*Very relevant*).

Next, they indicated how easy these items were to understand on a five-point scale ranging from 0 (*Very Difficult*) to 4 (*Very Easy*). Finally, they indicated on a binary scale whether the items were grammatically correct (0 = *No*; 1 = *Yes*) and provided demographic information.

4.2. Results and discussion

We calculated the level of agreement between raters similarly to Study 1a and found a good reliability of ratings in terms of relevance, *ICC* = 0.88, 95% CI [0.79, 0.95] (Koo & Li, 2016). Raters estimated the mean relevance of the 17 items to the construct of interest as just above the scale midpoint (*M* = 1.53). Again, we removed the items whose relevance to the focal construct was below the median. In this way, nine items remained for further testing, which was rated as relatively easy to understand (*M* = 3.33) and grammatically correct (*M* = 0.90). We reworded one item from "I believe that companies should always strive to offer labeled food" to "I believe that companies should always offer labeled food." Study 1 b thus yielded a final pool of 9 items for formal psychometric assessment in Study 2.

5. Scale refinement (Study 2)

Study 2 was the first formal psychometric evaluation of the selected nine items. Here we performed a Mokken scale analysis (Mokken, 1971; Van der Ark, 2007, 2012) and reported conventional internal consistency reliability indices.

5.1. Participants and procedure

Although there is no consensus in the literature on an appropriate sample size for scale development studies, it is generally recommended to recruit at least 300 participants at the scale refinement stage, with more participants reducing measurement error, and longer scales generally requiring more participants than shorter ones (Boateng et al., 2018). We therefore recruited 558 US participants from Qualtrics who declared no dietary restrictions. Given recent concerns about data quality on some crowdsourcing platforms (Eyal, David, Andrew, Zak, & Ekaterina, 2021), we used the recommended multistep approach to filter responses (Craft, Tegge, Freitas-Lemos, Tomlinson, & Bickel, 2022). We found that 118 participants failed either the "Captcha" question or an attention check question, described in more detail below. Further manual review of the data revealed that 15 participants gave nonsensical answers when asked to estimate the population of their country, such as "more" or "na." These 133 participants, or 23.8% of the original sample, were excluded for analysis to ensure high data quality. Thus, the final sample consisted of 425 participants (Mean age = 47.1, *SD* = 16.2, 51.8% females, 48.0% males, 0.2% did not identify with either of the two genders, mean annual income = 48,200 USD).

Participants first learned about FOP labeling and were shown an example of such a label used in the food industry. They then indicated their agreement with the nine items selected in Study 1 b on a nine-point scale ranging from 1 (*Disagree strongly*) to 9 (*Agree strongly*). Within this pool of items, which were presented in a random order, we embedded an attention check question ("After reading this question, choose the

number seven using the slider below"). After completing this questionnaire, participants provided their demographic data and free-text responses to another attention check question that asked them about the approximate population size of their country. Finally, they completed a third and final attention check question asking them to indicate the number of bicycles they saw in a "captcha" image.

5.2. Results and discussion

We began the psychometric evaluation of the nine items with a Mokken scale analysis, which revealed that these items formed a homogeneous and strong scale, *H* = 0.54, *SE* = 0.03 (Mokken, 1971; Van der Ark, 2007). We then used the automated item selection procedure algorithm to test whether these items were scalable with increasing homogeneity thresholds. The recommended minimum homogeneity threshold for items included in a scale is 0.30 (Hemker et al., 1995). However, we used a more conservative approach and increased our lower-bound homogeneity threshold to 0.50 (for a similar approach, see, e.g., Folwarczny, Li, et al., 2021). Two items did not meet this lower-bound homogeneity threshold and were removed from subsequent analysis.

We recalculated the homogeneity coefficient for the revised set of seven items and found that these items formed a homogeneous and strong scale, *H* = 0.60, *SE* = 0.03 (Mokken, 1971; Van der Ark, 2007). None of the items violated the monotonicity criteria, local independence, and invariant item order ordering; hence, all criteria for a double monotonicity model were met (Mokken, 1971). The internal consistency reliability of this scale was excellent, α = 0.90, ω = 0.91, 95% CI [0.89, 0.93], mean average correlation between items = 0.57. The scale's means ranged from 6.60 to 7.71 and standard deviations ranged from 1.32 to 2.01. The minimum mean was 1.57, while the maximum mean was 9.00, indicating that participants' scores spanned a broad spectrum of possible responses.

Study 2 showed that seven items met the criteria for a double monotonicity model and formed a strong scale (Mokken, 1971). This scale was further evaluated in Study 3a to ensure that the seven items measure the construct of interest (see Table 1, for the final set of items included in the CSFL scale).

6. Convergent and discriminant validity (Study 3 a)

Study 2 resulted in the development of a seven-item scale aimed at capturing consumers' susceptibility to FOP labeling systems (i.e., the CSFL scale). Although the previous study confirmed good psychometric properties of the scale, it remains unknown whether this newly developed instrument actually measures the construct of interest rather than other related constructs. Therefore, in Studies 3a–b, we compared the results of our scale against those obtained from several other instruments to test the convergent and discriminant validity of the CSFL scale.

Given that the construct to be measured by the CSFL scale should be related to thinking styles, pro-environmental consumption values, and impression management tactics, we used the "GREEN" scale (Haws, Winterich, & Naylor, 2014), the Holistic Cognition Scale (Lux, Grover, &

Table 1
The final set of items in the consumer susceptibility to FOP labeling (CSFL) scale.

CSFL 1:	Food labeling allows for informed choices by consumers
CSFL 2:	I prefer grocery stores with a vast selection of labeled products
CSFL 3:	When buying groceries, I seek labeled products
CSFL 4:	I believe that companies should always offer labeled food
CSFL 5:	I tend to select food items with food labels
CSFL 6:	Food labels allow me to know what I am consuming
CSFL 7:	It is important that my food is labeled

Note. Participants responded on a nine-point scale ranging from 1 (*Disagree strongly*) to 9 (*Agree strongly*).

Teo, 2021), and both the Social Desirability (Stöber, 2001) and Impression Management (Bolino & Turnley, 1999) Scales to test the convergent validity of our instrument. To test the discriminant validity of the scale, we used the Anticipated Food Scarcity Scale, an instrument also developed for food-related contexts but used to measure an unrelated construct (Folwarczny, Li, et al., 2021). In addition, we included a brief measure of the Big Five Personality Traits (Gosling, Rentfrow, & Swann, 2003) as an exploratory measure.

Study 3a was the first wave of the two studies to examine the test-retest reliability of this instrument. To further test the generalizability of our results, we used a different crowdsourcing platform to collect data in Studies 3a–b than in Study 2 (Prolific Academic rather than Qualtrics, as in Study 2).

6.1. Participants and procedure

Consistent with the considerations from Study 2 for estimating sample size, we attempted to recruit at least 300 participants for the two-wave study. However, because of the expected attrition rate of 15–20% based on our previous longitudinal studies with participants recruited through Prolific Academic, we aimed to recruit at least 15% more participants for Study 3a, such that the final sample in the second wave of data collection (Study 3 b) would still be around 300 participants. We recruited 351 US participants from Prolific Academic who declared no dietary restrictions. We used a similar approach to filter responses as previously. Here, 12 participants failed either the "captcha" question or an attention check question. A manual review of the free-text responses resulted in the removal of one additional response, resulting in excluding 13 responses, or 3.7% of the original sample. The final sample consisted of 339 participants (Mean age = 42.3, $SD = 14.0$, 48.2% females, 50.3% males, 1.5% did not identify with either of the two genders, mean annual income = 45,300 USD).

Participants were first given general instructions about the duration of the study and were informed that this study was the first of two waves of data collection, with 25% more financial rewards for taking part in the second wave. They then completed a series of psychometric instruments presented in a random order, as described below, and finally provided their demographic data.

6.1.1. Consumer susceptibility to FOP labeling

Participants completed the seven-item CSFL scale developed and described in Study 2 using the same nine-point response format. We averaged the responses to create a CSFL index ($\alpha = 0.93$, $M = 6.76$, $SD = 1.58$).

6.1.2. Green consumption values

Green consumption values were measured using the "GREEN" scale, with this construct being defined as "the tendency to express the value of environmental protection through one's purchases and consumption behaviors" (Haws et al., 2014, p. 337). Participants indicated their agreement with six items (e.g., "I am concerned about wasting the resources of our planet") on a seven-point scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*). We averaged responses to create a GREEN index ($\alpha = 0.93$, $M = 4.54$, $SD = 1.34$). Because FOP labels often signal product attributes such as sustainability, we expect green consumption values to be positively associated with CSFL.

6.1.3. Preferences for analytic versus holistic thought

The analytic thinking style, typical of Western cultures, revolves

around focusing attention primarily on an object rather than the context in which a given object appears, as well as on logical, rule-based thinking; the holistic thinking, which is typical of East Asian cultures, is characterized by focusing attention primarily on context rather than on a particular object, and by an aversion to the use of categories and formal logic (Nisbett et al., 2001). We measured preferences for holistic versus analytic thinking styles using the 16-item Holistic Cognition Scale[1] (HCS; Lux et al., 2021). Here, participants indicated their agreement with the items such as "It is impossible to understand the pieces without considering the whole picture" using a seven-point scale ranging from 1 (*Completely disagree*) to 7 (*Completely Agree*). We averaged responses to the four previously established dimensions (Lux et al., 2021) to create an index of attention ($\alpha = 0.35$, $M = 4.95$, $SD = 0.74$), an index of causality ($\alpha = 0.59$, Mean $= 3.72$, $SD = 1.01$), an index of contradiction ($\alpha = 0.47$, $M = 5.10$, $SD = 0.79$), and an index of change ($\alpha = 0.37$, $M = 3.39$, $SD = 0.78$). We reverse-coded the items in the causality and change subscales; therefore, the higher scores in all HCS subscales correspond to more holistic (as opposed to analytic) thinking. Because FOP labels convey information that is generally independent of context (i.e., other attributes of a product, such as its packaging), we expect all HCS subscales to be negatively associated with CSFL (i.e., positively associated with analytic rather than holistic thinking style).

6.1.4. Big Five Personality Traits

The Big Five personality traits (extroversion, agreeableness, conscientiousness, emotional stability, and openness to experience) were measured using the TIPI brief measure (Gosling et al., 2003). Participants indicated the extent to which each of the 10 pairs of traits (e.g., "Extraverted, enthusiastic") applied to them using a seven-point scale ranging from 1 (*disagree strongly*) to 7 (*agree strongly*). We averaged responses to the items corresponding to extroversion ($r = 0.74$, $p < .001$, $M = 3.18$, $SD = 1.71$), agreeableness ($r = 0.36$, $p < .001$, $M = 5.27$, $SD = 1.35$), conscientiousness ($r = 0.59$, $p < .001$, $M = 5.40$, $SD = 1.44$), emotional stability ($r = 0.72$, $p < .001$, $M = 4.76$, $SD = 1.63$), and openness to experience ($r = 0.38$, $p < .001$, $M = 5.10$, $SD = 1.32$). We measured personality traits because they are important for predicting a variety of consumer-related outcomes. However, this analysis should be considered exploratory because no clear pattern of associations was expected.

6.1.5. Anticipated food scarcity

Anticipated food scarcity was captured by the Anticipated Food Scarcity Scale (AFSS) and is formally defined as "the perception of future food resources becoming insufficient in terms of availability and accessibility" (Folwarczny, Li, et al., 2021). Here, participants indicated their agreement with 8 items (e.g., "Food shortages will occur more frequently") on a seven-point scale ranging from 1 (*Disagree Strongly*) to 7 (*Agree Strongly*). Similar to the procedures above, we averaged these responses to create the AFSS index ($\alpha = 0.97$, $M = 4.96$, $SD = 1.52$). We used AFSS because this scale captures food-related outcomes in domains that CSFL was not intended to cover. Therefore, AFSS was used to test the discriminant validity of CSFL, and we predicted a lack of correlation between these two scales.

6.2. Results and discussion

We first performed a psychometric evaluation of the CSFL scale in the same manner as in Study 2. The 7 items formed a homogenous and strong scale, $H = 0.68$, $SE = 0.02$ (Mokken, 1971; Van der Ark, 2007).

---

[1] Because of the very low reliability estimates of the original four factors, we performed a factor analysis with oblimin rotation to extract two factors corresponding to the analytic and holistic thinking styles. However, because these two factors also yielded low reliability coefficients below $\alpha = 0.60$, the results of this scale should be interpreted with appropriate caution.

We did not observe major violations of the criteria of monotonicity, local independence, or invariant item order ordering. Therefore, the criteria for a double monotonicity model were met again (Mokken, 1971). The internal consistency reliability of this scale was excellent, $\alpha = 0.93$, $\omega = 0.94$, 95% CI [0.92, 0.95], mean inter-item correlation $= 0.65$. The scale's means ranged from 6.01 to 7.42, and standard deviations ranged from 1.47 to 2.28.

Next, we calculated Pearson correlation coefficients between the CSFL scale index and the other measures included in this study, which are presented in Table 2.

The CSFL scale index showed negligible and low correlations with the other instruments. Positive, albeit weak, associations between CSFL and green consumption values are warranted because environmentally conscious consumers should generally make more informed choices than their peers who are less interested in such consumption (Haws et al., 2014), and FOP labels often communicate product attributes such as environmental impact and place of origin. Thinking styles, whether analytic or holistic, were weakly associated with CSFL, with no clear pattern of results. Weak and positive correlations between CSFL and all Big Five personality traits are surprising, although the link between CSFL and conscientiousness is grounded in theory, given previous research connecting this trait with an analytic information processing style (e.g., Gadassi, Gati, & Dayan, 2012). Importantly, the lack of a significant association between CSFL and AFSS, the latter measuring food-related outcomes that were not of interest in the development of the CSFL scale (which is also primarily intended for use in the food industry), is evidence of the discriminant validity of CSFL.

7. Convergent and discriminant validity (Study 3 b)

To prevent extensive participant fatigue and thus increase the chances that participants would be attentive to their assigned tasks, we split the data collection into two waves. We also wanted to investigate the test-retest reliability of the CSFL scale.

Therefore, participants who had taken part in Study 3a were invited to take part in Study 3b one week after the completion of the first wave of data collection, with both studies including a roughly equal number of items.

7.1. Participants and procedure

In Study 3b, we recruited 293 participants from Study 3a's sample, meaning that the attrition rate was only 16.5%, and, accordingly, we found a high and thus desired response rate of 83.5%. Eight participants, or 2.7% of the total sample, failed the "captcha" question and were thus excluded from further analysis. The final sample consisted of 285 participants (Mean age = 43.6, $SD = 14.1$, 49.5% females, 48.8% males, 1.8% did not identify with either of the two genders, mean annual income = 45,600 USD).

The procedure in Study 3 b was similar to that of the earlier study with two exceptions. First, participants completed a different set of psychometric scales in addition to the CSFL scale. Second, we did not collect demographic data this time, as these data had already been collected in the previous wave of data collection. Below, we describe the measures implemented in Study 3b.

7.1.1. Consumer susceptibility to FOP labeling
Participants completed the CSFL scale again, and their responses were averaged to create a CSFL index ($\alpha = .94$, $M = 6.86$, $SD = 1.64$).

7.1.2. Social desirability
We measured social desirability, defined as the tendency to give responses that are considered positive and desirable by others, using the SDS-17 scale (Stöber, 2001). Participants indicated whether each of the items (e.g., "I sometimes litter") was descriptive of them (1 = *true*) or not (0 = *false*). Following the scale author's instructions, we did not include the item on illegal drug use (Stöber, 2001). Participants' responses to the remaining 16 items were averaged to create an SDS index ($\alpha = 0.85$, $M = 0.57$, $SD = 0.26$). Considering that consumers sometimes buy food to impress others (Folwarczny, Otterbring, & Ares, 2023) and that FOP labels may signal socially desirable qualities, we expect a positive relationship between CSFL and social desirability.

7.1.3. Impression management
Impression management is a process by which people consciously or unconsciously strategically change the way they are perceived by others (Schlenker, 2012). We used a scale developed by Bolino and Turnley (1999) that captures the five facets of impression management reported in the literature: self-promotion, ingratiation, exemplification, intimidation, and supplication. Participants read 22 items and indicated how often they behaved in the ways described (e.g., "Talk proudly about your experience or education"). They did so on a five-point scale ranging from 1 (*never behave this way*) to 5 (*often behave this way*). We averaged items corresponding to self-promotion ($\alpha = 0.88$, $M = 2.31$, $SD = 0.93$), ingratiation ($\alpha = 0.86$, $M = 2.91$, $SD = 1.00$), exemplification ($\alpha = 0.68$, $M = 2.20$, $SD = 0.81$), intimidation ($\alpha = 0.88$, $M = 1.42$, $SD = 0.61$), and supplication ($\alpha = 0.90$, $M = 1.62$, $SD = 0.73$) to create corresponding index variables. We expected a positive association between impression management and CSFL for the same reason as described above, i.e., consumers may be keen to buy products with FOP labels to impress others.

7.2. Results and discussion

We reevaluated the psychometric properties of the CSFL scale, with the results confirming good psychometric properties of the 7 items, which formed a homogenous and strong scale, $H = 0.74$, $SE = 0.02$, without major violations of the criteria of monotonicity, local

Table 2
Pearson correlation coefficients between consumer susceptibility to FOP labeling (CSFL) and other measures in Study 3a.

Variable	1	2	3	4	5	6	7	8	9	10	11
1. CSFL											
2. GREEN	.24***										
3. HCS: Attention	.15**	.16**									
4. HCS: Causality	−.17**	.05	−.09								
5. HCS: Contradiction	.12*	.16**	.36***	−.09							
6. HCS: Change	−.18**	−.12*	−.20***	.29***	−.08						
7. TIPI: Extroversion	.12*	.17**	.10	−.05	.01	−.15**					
8. TIPI: Agreeableness	.25***	.24***	.17**	−.09	.14*	−.24***	.18***				
9. TIPI: Conscientiousness	.21***	.14**	.02	−.08	.02	−.18***	.21***	.40***			
10. TIPI: Emotional stability	.18***	.08	.01	−.12*	−.08	−.16**	.32***	.34***	.43***		
11. TIPI: Openness to experience	.14*	.16**	.09	−.06	.16**	.00	.32***	.17**	.11	.23***	
12. AFSS	.05	.21***	.11*	.12*	.10	−.04	−.02	−.01	−.11*	−.12*	.06

Note. This table presents Pearson correlation coefficients. ***$p < .001$; **$p < .01$; *$p < .05$. "HCS" stands for holistic cognition scale. "TIPI," or the Ten-Item Personality Inventory (Gosling et al., 2003) is a brief measure of personality.

independence, and invariant item order ordering (Mokken, 1971; Van der Ark, 2007). The internal consistency reliability of this scale was excellent, $\alpha = 0.94$, $\omega = 0.95$, 95% CI [0.94, 0.96], mean inter-item correlation = 0.70. The scale's means ranged from 6.12 to 7.50, and standard deviations ranged from 1.38 to 2.28.

Next, we calculated the correlations between the scores of waves 1 and 2 to examine the test-retest reliability of the scale. We found a strong positive correlation across both waves, $r = 0.77$, 95% CI [0.73, 0.82], $p < .001$. The confidence intervals around the Pearson correlation coefficient indicate a substantial correlation between the scores from these two data collections (Cohen, 1988, p. 80). Thus, the CBLE scale is stable over time.

Finally, we calculated Pearson correlation coefficients between the CSFL scale and the other instruments (see Table 3).

Similar to the previous study, we found only weak and negligible correlations between the CSFL scale and scores on the other scales. Positive associations between the CSFL and social desirability and one of the facets of impression management suggest that consumers may be interested in labeled products to create a desirable image in the eyes of others, consistent with recent related research (Folwarczny et al., 2023; Gasiorowska, Folwarczny, Tan, & Otterbring, 2023). Considering that all of these correlations were weak, Study 3 b supports the notion that CSFL is associated with the desire to impress others, although it is a qualitatively different construct.

8. Criterion validity (Study 4)

Studies 2–3 b confirmed good psychometric properties of the CSFL scale. The relatively weak correlations between the instrument and other scales analyzed across Studies 3a–b suggest that the scale captures a new construct that is qualitatively different from numerous potentially related constructs. However, the predictive validity of the CSFL scale is yet to be examined. Therefore, in Study 4, we tested whether the CSFL scale predicts willingness to buy (WTB) products with FOP labeling systems. Because previous studies have shown that consumers generally respond similarly to fictitious and genuine labels (Sigurdsson et al., 2020, 2022, 2023), we specifically sought to investigate whether the scale can predict whether consumers with high (vs. low) susceptibility to FOP labeling are more willing to buy products with genuine third-party certificates rather than products with fictitious labels or products without any certification.

We chose fish fillets as the product type to test the predictive power of our newly developed CSFL scale for several reasons. First, previous research has shown that certain FOP labels on fish products have minimal, if any, influence on consumer preferences (Hallstein & Villas-Boas, 2013; Sigurdsson et al., 2022, 2023). Second, consumers often indicate that they value other seafood attributes, such as country of origin, more than FOP labeling when buying fish products (Sigurdsson et al., 2020). Third, whereas fish is not the predominant protein source among US

participants, three-quarters reported consuming fish in the past 30 days (Jahns et al., 2014). Therefore, measuring the predictive validity of the CSFL scale in the context of fish products provides a conservative test of the scale's criterion validity, as FOP labeling is neither the highest rated attribute nor an effective strategy for increasing willingness to pay or daily frequency of consumption when it comes to fish. Habituation effects can therefore likely be ruled out as confounding factors for our potential results (Wathieu, 2004).

8.1. Participants and procedure

Estimating the desired sample size for research designs with repeated measures and interactions, where linear mixed models are the preferred analytic approach, is complex. Previous studies using similar research designs in food-related contexts have achieved satisfactory power to detect similar interaction effects as those examined in Study 4 with approximately 100–120 participants (Folwarczny, Christensen, Li, Sigurdsson, & Otterbring, 2021; Folwarczny, Otterbring, Sigurdsson, & Gasiorowska, 2022). Using a conservative approach, we aimed to increase this sample size by 50%. Consequently, we recruited 175 US participants from Prolific Academic who declared no dietary restrictions, two of whom (1% of the total sample) failed the "captcha" question and were excluded from further analysis. The final sample thus consisted of 173 participants (Mean age = 41.8, $SD = 13.5$, 46.2% females, 53.2% males, 0.6% did not identify with either of the two genders, mean annual income = 51,900 USD).

We used a within-subjects design in which participants were asked to complete a series of tasks. The first task was to indicate WTB nine fish fillets on a tray resembling the fillets offered in supermarkets ("How likely would you be to buy each of the 9 fish fillets if they were available in a grocery store that you usually visit."). Participants responded on a 201-point scale ranging from −100 (*Very unlikely*) to 100 (*Very likely*). The fillets were presented in random order and differed in their FOP labeling (i.e., some fillets had genuine certificates, some had fictitious labels, and others were unlabeled). Two fillets had the American Heart Association certificate (the genuine, third-party certificate) in the upper-right corner. Two fillets had a B-Corp mark in the upper-right corner, also a genuine, third-party certificate. Two fillets bore a fictitious label in the upper-right corner designed for this study (this label showed a blue fish with the words "CSAP certified" underneath). Finally, the remaining three fillets did not bear any FOP label.

In addition, participants completed the CSFL scale. We created a CSFL index by averaging responses to the items included in the scale ($\alpha = .94$, $M = 6.69$, $SD = 1.67$). To avoid possible confounds from the task in which participants reported their WTB fish fillets, we separated these two tasks (i.e., evaluating fish fillets and filling out the SCFL scale) by adding a filler task in which participants reported their demographic data, completed an unrelated, 17-item questionnaire capturing dominance and prestige orientations (sample item: "Others do what I ask of them for fear of consequences") when seeking status (Körner, Heydasch, & Schütz, 2022), and answered the attention check questions, similar to the previously described studies.

8.2. Results and discussion

Because our data were nested and thus autocorrelations occurred across measurements, we used linear mixed models to perform the analysis with the *lme4* and *lmerTest* packages for R (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). We added random intercepts for participants and product types (i.e., measurements). As predictors (fixed effects), we added the CSFL index and a label type (a genuine certificate, a fictitious label, and no label, with the "no label" category as the reference category in all analyses). The model also included the interaction term between the two predictors, with WTB fish fillets as the focal dependent variable.

Within this model, we found no main effect of the CSFL index, $b =$

Table 3
Pearson correlation coefficients between consumer susceptibility to FOP labeling (CSFL) and other measures in Study 3 b.

Variable	1	2	3	4	5	6
1. CSFL						
2. SDS	.22***					
3. IM: Self-promotion	.09	−.03				
4. IM: Ingratiation	.13*	−.03	.46***			
5. IM: Exemplification	−.03	−.08	.46***	.55***		
6. IM: Intimidation	−.05	−.17**	.42***	.24***	.28***	
7. IM: Supplication	−.07	−.34***	.25***	.30***	.34***	.44***

Note. This table presents Pearson correlation coefficients. ***$p < .001$; **$p < .01$; *$p < .05$.
"IM" stands for impression management.

0.95, $t(1548) = 0.55$, $p = .581$, $\beta = 0.03$, 95% CI [-0.07, 0.12]. The main effect of the fictitious label, $b = -11.01$, $t(1548) = -0.32$, $p = .747$, $\beta = -0.11$, 95% CI [-1.16, 0.95], and the main effect of the genuine label, $b = -10.06$, $t(1548) = -0.35$, $p = .724$, $\beta = 0.14$, 95% CI [-0.74, 1.02], were also non-significant. Similarly, we found no significant interaction between the CSFL index and the fictitious label, $b = 0.69$, $t(1548) = 0.38$, $p = .702$, $\beta = 0.02$, 95% CI [-0.08, 0.12]. However, we found a marginally significant interaction between the CSFL index and the genuine label, $b = 2.72$, $t(1548) = 1.80$, $p = .072$, $\beta = 0.08$, 95% CI [0.00, 0.16].

To understand the nature of this interaction, we performed the simple slopes analysis. The slope of the CSFL index was statistically non-significant for both, unlabeled fillets, $b = 0.95$, $t(514) = 0.54$, $p = .587$, $\beta = 0.03$, 95% CI [-0.07, 0.13], and fillets with ficititous labels, $b = 1.64$, $t(341) = 0.88$, $p = .380$, $\beta = 0.05$, 95% CI [-0.06, 0.16]. However, the slope of the CSFL index was significant for fish fillets with genuine labels, $b = 3.67$, $t(687) = 2.37$, $p = .018$, $\beta = 0.10$, 95% CI [0.02, 0.18].

In sum, the results of Study 4 show that the CSFL scale predicts WTB fish fillets with genuine, third-party certificates but not fish fillets with fictitious labels or fillets without any FOP labels, thus attesting to the predictive validity of our novel instrument.

Fig. 1 illustrates these key findings.

9. General discussion

The results from previous studies suggest that different segments of consumers are differentially prone to rely on FOP labeling when purchasing food (Clonan et al., 2012; Sigurdsson et al., 2020). However, no prior scale has been developed to capture consumers' susceptibility to FOP labeling systems. In this paper, we created such a scale. The CSFL scale captures a unidimensional construct and has good psychometric properties, which we confirmed in three studies. The CSFL scale also predicts WTB fish fillets with genuine, third-party certificates but not unlabeled fillets and those with fictitious FOP labels, thereby underscoring its ability to predict what it is assumed to predict. Moreover, given its narrow set of items, the scale is an appropriate instrument for academic research and retail practice.

Whereas some FOP labeling systems are viable tools to achieve desired consumer outcomes (e.g., Ares et al., 2018, 2021; Sigurdsson et al., 2022), the effectiveness of such systems has been called into question (e.g., Hallstein & Villas-Boas, 2013; Temple, 2020). This discrepancy could be due to the omission of one potential moderator: consumers' susceptibility to FOP labeling systems. Thus, applying the newly developed scale to study consumer behavior could lead to more accurate policymaking regarding FOP labeling campaigns aimed at promoting outcomes such as healthier food choices. Indeed, previous research has found that messages tailored to individual differences (as opposed to one-size-fits-all approaches) considerably increase sales of targeted products (Matz, Kosinski, Nave, & Stillwell, 2017). Therefore, identifying consumers who are less susceptible to FOP labeling systems may encourage the development of alternative interventions that better target this segment.

The newly developed CSFL scale shows only weak correlations with other instruments, such as the GREEN scale, designed to capture consumers' propensity to purchase sustainable products (Haws et al., 2014). This finding suggests that the construct examined herein is not merely a ramification of a broader category of pro-environmental tendencies. Indeed, FOP labels vary considerably in terms of the information they provide, and these are not limited to environmentally friendly attributes (Golan et al., 2001; Hallstein & Villas-Boas, 2013; Kühne et al., 2022). Furthermore, the CSFL scale does not appear to be related to thinking styles when treated as analytic versus holistic, as we did not observe a clear pattern of correlations between the CSFL and thinking styles.

We found weak, albeit significant, relationships between the CSFL scale and social desirability as well as one of the impression management strategies—ingratiation (Bolino & Turnley, 1999; Stöber, 2001). Hence, it is plausible that the tendency to purchase products with FOP labeling is an impression management strategy used by a certain segment of consumers, which is consistent with previous studies (Gasiorowska et al., 2023). Importantly, CSFL was not associated with AFSS, which measures another food-related outcome (Folwarczny et al., 2023). This finding provides evidence behind the scale's discriminant validity.

Study 4 showed that the CSFL was positively associated with WTB fish fillets with genuine FOP. We used two third-party certificates here: B-Corp and American Heart Association labeling. However, we found no such association between the CSFL and WTB fillets that did not have FOP labels. There was also no such association for fish fillets with fictitious labels. This result suggests that the newly developed CSFL scale may not only predict consumer interest in FOP labels but also imply that this scale can discriminate between different labels.

9.1. Limitations and future research

Similar to all academic research, the current work is not without limitations. First, surprisingly, we found positive correlations between all Big Five personality traits and the CSFL scale. However, we used a very brief measure of the personality traits (Gosling et al., 2003), thus calling for further research using more extensive personality inventories. As indicated by the results in Tables 2 and 3, most of the correlation coefficients we observed were weak or negligible, although we found moderately strong and theoretically justified correlations between the CSFL scale and, for example, green consumption values (Study 3a) and social desirability (Study 3b), according to current conventions of what constitutes small, moderate, and large effects (e.g., Funder & Ozer, 2019; Gignac & Szodorai, 2016). The instances of weak correlations could be due to the fact that CSFL is a new construct and therefore has not been well explored in the literature. Another possibility is that the scope of the instruments we selected to test the convergent and discriminant validity of the scale is somewhat limited for this task. Future studies with additional literature-based instruments are essential to assess the convergent and discriminant validity of the CSFL scale more fully, considering that our empirical work merely provides an initial indication of these validity aspects.

Second, although the results from Study 4 are in line with our theorizing, we only measured WTB in one category of grocery products: fish fillets. It is critical to investigate whether our results can be generalized to other product categories, as the CSFL scale is intended to be used in various situations. Additionally and relatedly, the current research, particularly Study 4, measured expressed tendencies rather than real, observable behaviors. Without additional studies that
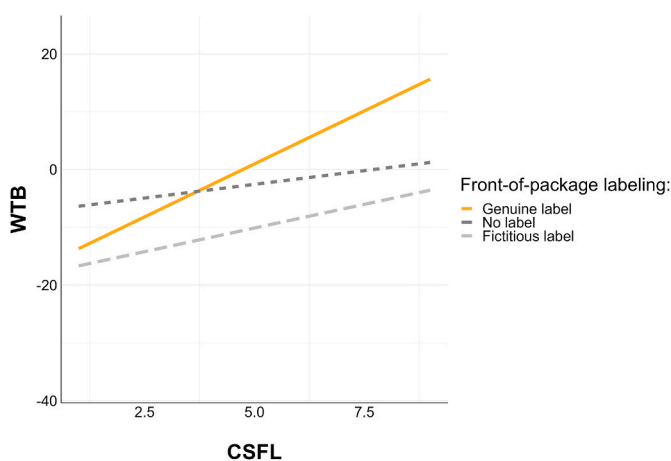

Fig. 1. Willingness-to-buy (WTB) by front-of-package labeling type.
Note. WTB (y-axis) denotes willingness to buy. CSFL (x-axis) is an index of consumer susceptibility to front-of-package food labeling.

preferentially apply the newly developed instrument in settings that are natural to consumers, the criterion validity of the scale has not yet been established (Baumeister, Vohs, & Funder, 2007; Doliński, 2018; Doliński et al., 2017; Otterbring, Viglia, Grazzini, & Das, 2023).

Third, we adopted a recommended multistep approach to ensure data quality across studies (Craft et al., 2022). As a result, in Study 2, we had to exclude nearly a quarter of the sample originally recruited through Qualtrics. In our remaining studies, where Prolific Academic participants were generally very attentive to instructions, exclusions were substantially smaller, supporting research that has demonstrated high quality from this crowdsourcing platform (Eyal et al., 2021). However, it remains essential to test the psychometric properties of the CSFL scale on different samples to rule out interference from possible sampling bias. A related concern about our sample selection is that our US participants likely do not consume fish on a daily basis, as over 80% of Americans do not meet recommended seafood consumption guidelines (Jahns et al., 2014). Consequently, subsequent studies in this area should measure the frequency of fish consumption among participants and strive to diversify their samples. This would allow verification that the effects observed in Study 4 are consistent across populations with different fish consumption habits, which is particularly important given the habitual nature of food purchases combined with cross-cultural differences in food consumption habits (Gatley, Caraher, & Lang, 2014; Guerrero et al., 2009; Machín et al., 2020). In the current research, we did not determine whether our participants were the main grocery shoppers in their households. Consumers who are regular grocery shoppers might respond differently to FOP labels than those who shop infrequently, possibly because of their familiarity and experience with such information. To account for this potential confounding factor, future research should inquire about participants' familiarity with FOP labels and the frequency of food purchases. At a minimum, such studies should determine whether a participant is the primary grocery shopper in their household.

10. Conclusion

The use of front-of-package (FOP) labeling systems produces mixed results in academic and non-academic settings. In some cases, such labeling schemes lead to desired behavioral changes whereas in others, these interventions lead to null results. In this work, we propose that consumers vary in terms of how susceptible they are to FOP food labeling systems. We, therefore, develop a 7-item instrument, which we refer to as the consumer susceptibility to FOP labeling (CSFL) scale. This tool was designed to measure the extent to which consumers rely on FOP labeling in their food choices and information search activities to meet their existing preference patterns. We confirmed good psychometric properties of the CSFL scale and conducted initial criterion validity tests, finding that the scale predicts WTB fish fillets with genuine, third-party FOP labels. We found no such effects for fish fillets with fictitious labels and no associations between the CSFL and WTB unlabeled fish fillets. Therefore, the scale may capture consumers' susceptibility to FOP labeling differently depending on the type of label (authentic vs. inauthentic). The relatively narrow set of items in the CSFL scale makes it ideal for academic and nonacademic research in food retailing.

Author contributions

MF: conceptualization, research design, data curation, data analysis and visualization, writing–first draft; VM: research design, writing–review and editing, funding acquisition; VS: research design, writing–review and editing, funding acquisition. TO: research design, writing–review and editing.

Funding

This research was funded by the Icelandic Research Fund, Project

Ethical statement

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Ares, G., Antúnez, L., Cabrera, M., & Thow, A. M. (2021). Analysis of the policy process for the implementation of nutritional warning labels in Uruguay. *Public Health Nutrition, 24*(17), 5927–5940. https://doi.org/10.1017/S1368980021002469

Ares, G., Antúnez, L., Otterbring, T., Curutchet, M. R., Galicia, L., Moratorio, X., et al. (2020). Sick, salient and full of salt, sugar and fat: Understanding the impact of nutritional warnings on consumers' associations through the salience bias. *Food Quality and Preference, 86*, Article 103991. https://doi.org/10.1016/j.foodqual.2020.103991

Ares, G., Aschemann-Witzel, J., Curutchet, M. R., Antúnez, L., Machín, L., Vidal, L., et al. (2018). Product reformulation in the context of nutritional warning labels: Exploration of consumer preferences towards food concepts in three food categories. *Food Research International, 107*, 669–674. https://doi.org/10.1016/j.foodres.2018.03.021

Bagnasco, A., Watson, R., Zanini, M., Rosa, F., Rocco, G., & Sasso, L. (2015). Preliminary testing using mokken scaling of an Italian translation of the Edinburgh Feeding Evaluation in Dementia (EdFED-I) scale. *Applied Nursing Research, 28*(4), 391–396. https://doi.org/10.1016/j.apnr.2015.02.003

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baum, M. (2019). *Grocery productivity numbers to watch.* https://www.fmi.org/blog/view/fmi-blog/2019/12/02/grocery-productivitynumbers-to-watch.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Bearden, W. O., Netemeyer, R. G., & Teel, J. E. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research, 15*(4), 473–481. https://doi.org/10.1086/209186

Bearth, A., Cousin, M.-E., & Siegrist, M. (2014). The consumer's perception of artificial food additives: Influences on acceptance, risk and benefit perceptions. *Food Quality and Preference, 38*, 14–23. https://doi.org/10.1016/j.foodqual.2014.05.008

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Bolino, M. C., & Turnley, W. H. (1999). Measuring impression management in organizations: A scale development based on the jones and pittman taxonomy. *Organizational Research Methods, 2*(2), 187–206. https://doi.org/10.1177/109442819922005

Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., & de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. *Quality and Quantity, 47*, 2341–2360. https://doi.org/10.1007/s11135-012-9684-5

Cabrera, M., Machín, L., Arrúa, A., Antúnez, L., Curutchet, M. R., Giménez, A., et al. (2017). Nutrition warnings as front-of-pack labels: Influence of design features on healthfulness perception and attentional capture. *Public Health Nutrition, 20*(18), 3360–3371. https://doi.org/10.1017/S136898001700249X

Celhay, F., & Remaud, H. (2018). What does your wine label mean to consumers? A semiotic investigation of bordeaux wine visual codes. *Food Quality and Preference, 65*, 129–145. https://doi.org/10.1016/j.foodqual.2017.10.020

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research, 23*, 485–493. https://doi.org/10.1007/s11136-013-0487-5

Clonan, A., Holdsworth, M., Swift, J. A., Leibovici, D., & Wilson, P. (2012). The dilemma of healthy eating and environmental sustainability: The case of fish. *Public Health Nutrition, 15*(2), 277–284. https://doi.org/10.1017/S1368980011000930

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic press.

Craft, W. H., Tegge, A. N., Freitas-Lemos, R., Tomlinson, D. C., & Bickel, W. K. (2022). Are poor quality data just random responses?: A crowdsourced study of delay discounting in alcohol use disorder. *Experimental and Clinical Psychopharmacology, 30* (4), 409–414. https://doi.org/10.1037/pha0000549

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

Dima, A. L. (2018). Scale validation in applied health resesarch: Tutorial for a 6-step R-based psychometrics protocol. *Health Psychology and Behavioral Medicine, 6*(1), 136–161. https://doi.org/10.1080/21642850.2018.1472602

Doliński, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin, 13*(2), 1–14.

Doliński, D., Grzyb, T., Folwarczny, M., Grzybała, P., Krzyszycha, K., Martynowska, K., et al. (2017). Would you deliver an electric shock in 2015? Obedience in the experimental paradigm developed by stanley milgram in the 50 years following the original studies. *Social Psychological and Personality Science, 8*(8), 927–933. https://doi.org/10.1177/1948550617693060

Duckworth, J. J., Randle, M., McGale, L. S., Jones, A., Doherty, B., Halford, J. C., et al. (2022). Do front-of-pack 'green labels' increase sustainable food choice and willingness-to-pay in UK consumers? *Journal of Cleaner Production, 371*, Article 133466. https://doi.org/10.1016/j.jclepro.2022.133466

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*, 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Faye, O., Baschieri, A., Falkingham, J., & Muindi, K. (2011). Hunger and food insecurity in Nairobi's slums: An assessment using IRT models. *Journal of Urban Health, 88*, 235–255. https://doi.org/10.1007/s11524-010-9521-x

Folwarczny, M., Christensen, J. D., Li, N. P., Sigurdsson, V., & Otterbring, T. (2021). Crisis communication, anticipated food scarcity, and food preferences: Preregistered evidence of the insurance hypothesis. *Food Quality and Preference, 91*, Article 104213. https://doi.org/10.1016/j.foodqual.2021.104213

Folwarczny, M., Li, N. P., Sigurdsson, V., Tan, L. K., & Otterbring, T. (2021). Development and psychometric evaluation of the anticipated food scarcity scale (AFSS). *Appetite, 166*, Article 105474. https://doi.org/10.1016/j.appet.2021.105474

Folwarczny, M., Otterbring, T., & Ares, G. (2023). Sustainable food choices as an impression management strategy. *Current Opinion in Food Science, 49*, Article 100969. https://doi.org/10.1016/j.cofs.2022.100969

Folwarczny, M., Otterbring, T., Sigurdsson, V., & Gasiorowska, A. (2022). Seasonal cues to food scarcity and calorie cravings: Winter cues elicit preferences for energy-dense foods. *Food Quality and Preference, 96*, Article 104379. https://doi.org/10.1016/j.foodqual.2021.104379

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gadassi, R., Gati, I., & Dayan, A. (2012). The adaptability of career decision-making profiles. *Journal of Counseling Psychology, 59*(4), 612–622. https://doi.org/10.1037/a0029155

Gasiorowska, A., Folwarczny, M., Tan, L., & Otterbring, T. (2023). Delicate dining with a date and burger binging with buddies: Impression management across social settings and consumers' preferences for masculine or feminine foods. *Frontiers in Nutrition, 10*, Article 1127409. https://doi.org/10.3389/fnut.2023.1127409

Gatley, A., Caraher, M., & Lang, T. (2014). A qualitative, cross cultural examination of attitudes and behaviour in relation to cooking habits in France and Britain. *Appetite, 75*, 71–81. https://doi.org/10.1016/j.appet.2013.12.014

Gibbons, J. D., & Chakraborti, S. (2014). *Nonparametric statistical inference: Revised and expanded*. CRC press.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Golan, E., Kuchler, F., Mitchell, L., Greene, C., & Jessup, A. (2001). Economics of food labeling. *Journal of Consumer Policy, 24*(2), 117–184. https://doi.org/10.1023/A:1012272504846

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Guerrero, L., Guàrdia, M. D., Xicola, J., Verbeke, W., Vanhonacker, F., Zakowska-Biemans, S., et al. (2009). Consumer-driven definition of traditional food products and innovation in traditional foods. a qualitative cross-cultural study. *Appetite, 52* (2), 345–354. https://doi.org/10.1016/j.appet.2008.11.008

Hallstein, E., & Villas-Boas, S. B. (2013). Can household consumers save the wild fish? Lessons from a sustainable seafood advisory. *Journal of Environmental Economics and Management, 66*(1), 52–71. https://doi.org/10.1016/j.jeem.2013.01.003

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.

Hamerman, E. J. (2016). Cooking and disgust sensitivity influence preference for attending insect-based food events. *Appetite, 96*, 319–326. https://doi.org/10.1016/j.appet.2015.09.029

Hartmann, C., & Siegrist, M. (2018). Development and validation of the food disgust scale. *Food Quality and Preference, 63*, 38–50. https://doi.org/10.1016/j.foodqual.2017.07.013

Haws, K. L., Winterich, K. P., & Naylor, R. W. (2014). Seeing the world through GREEN-tinted glasses: Green consumption values and responses to environmentally friendly products. *Journal of Consumer Psychology, 24*(3), 336–354. https://doi.org/10.1016/j.jcps.2013.11.002

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement, 19*(4), 337–352. https://doi.org/10.1177/014662169501900404

Hemmer, A., Hitchcock, K., Lim, Y. S., Kovacic, M. B., & Lee, S.-Y. (2021). Development of food literacy assessment tool targeting adults with low income. *Journal of Nutrition Education and Behavior, 53*(11), 966–976. https://doi.org/10.1016/j.jneb.2021.05.007

Ikonen, I., Sotgiu, F., Aydinli, A., & Verlegh, P. W. (2020). Consumer effects of front-of-package nutrition labeling: An interdisciplinary meta-analysis. *Journal of the Academy of Marketing Science, 48*, 360–383. https://doi.org/10.1007/s11747-019-00663-9

Jahns, L., Raatz, S. K., Johnson, L. K., Kranz, S., Silverstein, J. T., & Picklo Sr, M. J. (2014). Intake of seafood in the US varies by age, income, and education level but not by race-ethnicity. *Nutrients, 6*(12), 6060–6075. https://doi.org/10.3390/nu6126060

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Körner, R., Heydasch, T., & Schütz, A. (2022). Dominance and prestige as self-concept facets. *Journal of Personality Assessment*, 1–20. https://doi.org/10.1080/00223891.2022.2137028

Kühne, S. J., Reijnen, E., Granja, G., & Hansen, R. S. (2022). Labels affect food choices, but in what ways? *Nutrients, 14*(15), 3204. https://doi.org/10.3390/nu14153204

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lux, A. A., Grover, S. L., & Teo, S. T. T. (2021). Development and validation of the holistic cognition scale. *Frontiers in Psychology, 12*, Article 551623. https://doi.org/10.3389/fpsyg.2021.551623

Machín, L., Curutchet, M. R., Gugliucci, V., Vitola, A., Otterbring, T., de Alcantara, M., et al. (2020). The habitual nature of food purchases at the supermarket: Implications for policy making. *Appetite, 155*, Article 104844. https://doi.org/10.1016/j.appet.2020.104844

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences, 114*(48), 12714–12719. https://doi.org/10.1073/pnas.171096611

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*. De Gruyter Mouton. https://doi.org/10.1515/9783110813203

Muzzioli, L., Penzavecchia, C., Donini, L. M., & Pinto, A. (2022). Are front-of-pack labels a health policy tool? *Nutrients, 14*(4), 771. https://doi.org/10.3390/nu14040771

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291–310. https://doi.org/10.1037/0033-295X.108.2.291

Olatunji, B. O., Haidt, J., McKay, D., & David, B. (2008). Core, animal reminder, and contamination disgust: Three kinds of disgust with distinct personality, behavioral, physiological, and clinical correlates. *Journal of Research in Personality, 42*(5), 1243–1259. https://doi.org/10.1016/j.jrp.2008.03.009

Otterbring, T., Viglia, G., Grazzini, L., & Das, G. (2023). Guest editorial: Favoring fieldwork makes marketing more meaningful. *European Journal of Marketing, 57*(7), 1793–1803. https://doi.org/10.1108/EJM-07-2023-987

Pasquale, E. K., Manzano, M. A., Strong, D. R., Eichen, D. M., Tanofsky-Kraff, M., & Boutelle, K. N. (2023). Psychometric properties of the Eating in the Absence of Hunger Questionnaire in treatment-seeking adults with overweight and obesity. *Appetite, 180*, Article 106376. https://doi.org/10.1016/j.appet.2022.106376

Petersen, T., Hartmann, M., & Hirsch, S. (2021). Which meat (substitute) to buy? Is front of package information reliable to identify the healthier and more natural choice? *Food Quality and Preference, 94*, Article 104298. https://doi.org/10.1016/j.foodqual.2021.104298

Pettigrew, S., Jongenelis, M. I., Hercberg, S., & Julia, C. (2022). Front-of-pack nutrition labels: An equitable public health intervention. *European Journal of Clinical Nutrition, 77*, 135–137. https://doi.org/10.1038/s41430-022-01205-3

Prescott, J., Young, O., O'neill, L., Yau, N., & Stevens, R. (2002). Motives for food choice: A comparison of consumers from Japan, taiwan, Malaysia and New Zealand. *Food Quality and Preference, 13*(7–8), 489–495. https://doi.org/10.1016/S0950-3293(02)00010-1

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sanjari, S. S., Jahn, S., & Boztug, Y. (2017). Dual-process theory and consumer response to front-of-package nutrition label formats. *Nutrition Reviews, 75*(11), 871–882. https://doi.org/10.1093/nutrit/nux043

Schlenker, B. R. (2012). Self-presentation. In M. R. Leary, & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 542–570). The Guilford Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Siegrist, M., Hartmann, C., & Keller, C. (2013). Antecedents of food neophobia and its association with eating behavior and food choices. *Food Quality and Preference, 30* (2), 293–298. https://doi.org/10.1016/j.foodqual.2013.06.013

Sigurdsson, V., Larsen, N. M., Alemu, M. H., Gallogly, J. K., Menon, R. V., & Fagerstrøm, A. (2020). Assisting sustainable food consumption: The effects of quality signals stemming from consumers and stores in online and physical grocery retailing.

Journal of Business Research, 112, 458–471. https://doi.org/10.1016/j. jbusres.2019.11.029

Sigurdsson, V., Larsen, N. M., Folwarczny, M., Fagerstrøm, A., Menon, R. V., & Sigurdardottir, F. T. (2023). The importance of relative customer-based label equity when signaling sustainability and health with certifications and tags. *Journal of Business Research, 154*, Article 113338. https://doi.org/10.1016/j. jbusres.2022.113338

Sigurdsson, V., Larsen, N. M., Pálsdóttir, R. G., Folwarczny, M., Menon, R. V., & Fagerstrøm, A. (2022). Increasing the effectiveness of ecological food signaling: Comparing sustainability tags with eco-labels. *Journal of Business Research, 139*, 1099–1110. https://doi.org/10.1016/j.jbusres.2021.10.052

Sijtsma, K., & Hemker, B. T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika, 63*, 183–200. https://doi.org/10.1007/BF02294774

Sijtsma, K., & Junker, B. W. (1996). A survey of theory and methods of invariant item ordering. *British Journal of Mathematical and Statistical Psychology, 49*(1), 79–105. https://doi.org/10.1111/j.2044-8317.1996.tb01076.x

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response Theory* (Vol. 5). Sage.

Stöber, J. (2001). The social desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*(3), 222–232.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using linux. *Behavior Research Methods, 42*(4), 1096–1104. https://doi. org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Temple, N. J. (2020). Front-of-package food labels: A narrative review. *Appetite, 144*, Article 104485. https://doi.org/10.1016/j.appet.2019.104485

Umanath, S., & Coane, J. H. (2020). Face validity of remembering and knowing: Empirical consensus and disagreement between participants and researchers. *Perspectives on Psychological Science, 15*(6), 1400–1422. https://doi.org/10.1177/1745691620917672

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software, 20* (11), 1–19. https://doi.org/10.18637/jss.v020.i11

Van der Ark, L. A. (2012). New developments in mokken scale analysis in R. *Journal of Statistical Software, 48*(5), 1–27. https://doi.org/10.18637/jss.v048.i05

Wartella, E. A., Lichtenstein, A. H., Boon, C. S., et al. (2010). *Examination of front-of-package nutrition rating systems and symbols: Phase I report*. National Academies Press.

Wathieu, L. (2004). Consumer habituation. *Management Science, 50*(5), 587–596. https://doi.org/10.1287/mnsc.1030.0185

Watson, R. (1996). The Mokken scaling procedure (MSP) applied to the measurement of feeding difficulty in elderly people with dementia. *International Journal of Nursing Studies, 33*(4), 385–393. https://doi.org/10.1016/0020-7489(95)00058-5